

# Schätzung und Anwendung von Discrete-Choice-Modellen im Revenue Management



Freie wissenschaftliche Arbeit  
zur Erlangung des akademischen Grades  
„Master of Science“  
Studiengang: Wirtschaftsingenieurwesen

**an der**  
**Wirtschaftswissenschaftlichen Fakultät**  
**der Universität Augsburg**

– Lehrstuhl für Analytics & Optimization –

Eingereicht bei:	Prof. Dr. Robert Klein
Betreuer:	Prof. Dr. Robert Klein
Vorgelegt von:	David Fleckenstein
Adresse:	Josef-Priller-Straße 25, 3.110 86159 Augsburg
Matrikel-Nr.:	1365501
E-Mail:	david.martin.fleckenstein@student.uni-augsburg.de

Augsburg, 12. Juli 2019

# Inhaltsverzeichnis

<b>Abbildungsverzeichnis.....</b>	<b>IV</b>
<b>Tabellenverzeichnis .....</b>	<b>V</b>
<b>Abkürzungsverzeichnis .....</b>	<b>VI</b>
<b>Symbolverzeichnis .....</b>	<b>VI</b>
<b>1 Einführung .....</b>	<b>1</b>
<b>2 Anwendung der Discrete-Choice-Modelle.....</b>	<b>2</b>
2.1 Revenue Management.....	3
2.2 RM in der Airline-Industrie .....	5
2.3 Optimierungsprobleme im Airline-RM.....	6
2.3.1 Netzwerk-RM-Problem .....	7
2.3.2 Choice-based-Netzwerk-RM-Problem .....	8
2.3.3 Assortment-Optimierungsproblem .....	10
2.3.4 Einführung des Rechenbeispiels .....	10
<b>3 Grundlagen der Discrete-Choice-Modelle.....</b>	<b>13</b>
3.1 Spezifikation des Offersets .....	13
3.2 Substitutionsverhalten .....	14
3.3 Independent Demand .....	15
3.4 Random-Utility-Theory.....	18
3.4.1 Nutzenfunktion .....	18
3.4.2 Irrelevanz des absoluten Nutzens und der Skalierung .....	20
3.5 Spezifikation der Discrete-Choice-Modelle.....	22
3.5.1 Parametrische Modelle .....	22
3.5.2 Nicht-parametrische Modelle .....	23
3.6 Schätzung der Modellparameter .....	24
3.6.1 Maximum-Likelihood-Methode .....	24
3.6.2 Numerische Verfahren .....	26
3.6.3 Expectation-Maximization-Verfahren .....	28
3.6.4 Schätzverfahren für nicht-parametrische Modelle.....	31

<b>4</b>	<b>MNL-Modell</b>	<b>32</b>
4.1	Annahmen und Modellstruktur	32
4.2	Spezifikation	33
4.3	Numerische Maximierung der Log-Likelihood-Funktion	37
4.4	Exemplarische Spezifikation und Schätzung des MNL-Modells	38
<b>5</b>	<b>LC-MNL-Modell</b>	<b>41</b>
5.1	Annahmen und Modellstruktur	42
5.2	Spezifikation	43
5.3	EM-Verfahren von Train (2009)	44
5.4	Exemplarische Spezifikation und Schätzung des LC-MNL-Modells	49
<b>6</b>	<b>Rank-based-Modell</b>	<b>55</b>
6.1	Annahmen und Modellstruktur	56
6.2	Spezifikation	58
6.3	Spaltengenerierungsverfahren von van Ryzin und Vulcano (2015)	59
6.4	Exemplarische Spezifikation und Schätzung des Rank-based-Modells	68
<b>7</b>	<b>Markov-Chain-Modell</b>	<b>72</b>
7.1	Annahmen und Modellstruktur	72
7.2	EM-Verfahren von Simsek und Topaloglu (2018)	75
7.3	Exemplarische Schätzung des Markov-Chain-Modells	84
<b>8</b>	<b>Diskussion und Ausblick</b>	<b>91</b>
	<b>Literaturverzeichnis</b>	<b>93</b>
<b>9</b>	<b>Anhang</b>	<b>96</b>
	<b>Schriftliche Versicherung</b>	<b>100</b>

## Abbildungsverzeichnis

Abbildung 1: Richtung der Anpassung der Parameterschätzung in Abhängigkeit des Gradienten (in Anlehnung an Train (2009, S. 188)).....	27
Abbildung 2: Betrag der Anpassung der Parameterschätzung in Abhängigkeit der Hessematrix (in Anlehnung an Train (2009, S. 188)).....	27
Abbildung 3: Verlauf der MNL-Kaufwahrscheinlichkeit $P_1(S, V_1)$ in Abhängigkeit der deterministischen Nutzenkomponente $V_1$ (in Anlehnung an Train (2009, S. 38)) .....	33
Abbildung 4: Diagramm der MNL-Log-Likelihood-Funktion $LL(\beta_1, \beta_2)$ .....	39
Abbildung 5: Ablauf des EM-Verfahrens von Train (2009) .....	48
Abbildung 6: Diagramm der Log-Likelihood-Funktion $LL(\beta_{11}, \beta_{21})$ mit $d = 1, \dots, 5$ .....	50
Abbildung 7: Diagramm der Log-Likelihood-Funktion $LL(\beta_{12}, \beta_{22})$ mit $d = 6, \dots, 10$ .....	51
Abbildung 8: Diagramm der erwarteten Log-Likelihood-Funktion $ELL(\beta_{11}, \beta_{21}   \beta^0, s^0)$ in der ersten Iteration $i = 0$ .....	53
Abbildung 9: Diagramm der erwarteten Log-Likelihood-Funktion $ELL(\beta_{12}, \beta_{22}   \beta^0, s^0)$ in der ersten Iteration $i = 0$ .....	53
Abbildung 10: Attributgewichte der Segmente $l = 1, 2$ in den Iterationen $i = 0, \dots, 15$ .....	54
Abbildung 11: Anteile der Segmente $l = 1, 2$ in den Iterationen $i = 0, \dots, 15$ .....	54
Abbildung 12: Zielfunktionswert $LL(\beta^i, s^i)$ in den Iterationen $i = 0, \dots, 15$ .....	55
Abbildung 13: Ablauf des Spaltengenerierungsverfahrens von van Ryzin und Vulcano (2015).....	68
Abbildung 14: Diagramm des RMP (6.31) - (6.33) in der ersten Iteration $i = 1$ .....	70
Abbildung 15: Ablauf des EM-Verfahrens von Simsek und Topaloglu (2018).....	83

## Tabellenverzeichnis

Tabelle 1: Transaktionsdaten des fortlaufenden Anwendungsbeispiels .....	12
Tabelle 2: Daten der Produkte des fortlaufenden Anwendungsbeispiels .....	12
Tabelle 3: Daten der Kundensegmente des fortlaufenden Anwendungsbeispiels .....	13
Tabelle 4: Geschätzte Präferenzgewichte $v_j$ der Produkte $j = 0, \dots, 4$ ohne ASK .....	39
Tabelle 5: Kaufwahrscheinlichkeiten $P_{j_d}(S_d)$ der Produkte $j_d$ in den Transaktionen $d = 1, \dots, 10$ .....	40
Tabelle 6: Geschätzte Präferenzgewichte $v_j$ der Produkte $j = 0, \dots, 4$ mit ASK .....	40
Tabelle 7: Präferenzgewichte $v_{lj}$ der Segmente $l = 1, 2$ in der ersten Iteration $i = 0$ .....	51
Tabelle 8: $h(\beta_l   j_d, S_d, \beta^0, s^0)$ für die Segmente $l = 1, 2$ in der ersten Iteration $i = 0$ .....	52
Tabelle 9: Kompatible Präferenzlisten für $d = 1, \dots, 10$ und $\sigma^{1, ID}, \sigma^1$ sowie $\sigma^2$ .....	69
Tabelle 10: Optimale Werte der Variablen $\bar{y}_d$ und $\bar{\mu}_d$ für $d = 1, \dots, 10$ . .....	70
Tabelle 11: Bedingte Erwartungswerte der unbekannten Daten des Typs $\Lambda_j^2$ für Transaktion $d = 2$ und Produkte $j = 0, \dots, 4$ .....	86
Tabelle 12: Bedingte Erwartungswerte der unbekannten Daten des Typs $P_{jk}^2$ für Transaktion $d = 2$ und Produkte $j = 1, 2$ sowie $k = 0, \dots, 4$ .....	87
Tabelle 13: Aktualisierte Parameterschätzung $\rho^2$ .....	89
Tabelle 14: Parameterschätzung $(\lambda^{50}, \rho^{50})$ in Iteration $i = 50$ .....	90
Tabelle 15: Daten zur Schätzung eines LC-MNL-Modells mittels des EM-Verfahrens von Train (2009) auf Basis der Trainingsdaten der Beispielinstantz für die Iterationen $i = 0, \dots, 15$ .....	96
Tabelle 16: Daten zur Schätzung eines Markov-Chain-Modells mittels des EM-Verfahrens von Simsek und Topaloglu (2018) auf Basis der Trainingsdaten der Beispielinstantz für Iteration $i = 1$ .....	99

## Abkürzungsverzeichnis

AIK	Akaike-Informationskriterium
ASK	Alternativenspezifische Konstante
BFGS	Broyden-Fletcher-Goldfarb-Shanno-...
DAR	Displacement-Adjusted-Revenue
EM	Expectation-Maximization-...
FSNC	Full-Service-Network-Carrier
GDS	Globales Distributionssystem
IIA	Independence from irrelevant alternatives
KKT	Karush-Kuhn-Tucker-...
LCC	Low-Cost-Carrier
LC	Latent-Class-...
MNL	Multinomiales Logit-...
NR	Newton-Raphson-...
OLTP	Online Transactional Processing
RM	Revenue Management
RMP	Restricted-Master-Problem
RUM	Random-Utility-Modell
RUT	Random-Utility-Theory

## Symbolverzeichnis

### Allgemeine Symbole

$\beta$	Parameterset eines allgemeinen Discrete-Choice-Modells
$\beta^*$	Optimales Parameterset eines allgemeinen Discrete-Choice-Modells bzgl. des Maximum-Likelihood-Schätzproblems
$\beta^i$	Aktuelles Parameterset eines allgemeinen Discrete-Choice-Modells in Iteration $i$ eines Schätzverfahrens
$\Delta V$	Matrix der Differenzen der deterministischen Nutzenkomponenten
$\Delta V_{jk}$	Differenz der deterministischen Nutzenkomponenten zweier Produkte $j$ und $k$
$\Delta \varepsilon$	Matrix der Differenzen der stochastischen Nutzenkomponenten

$\Delta \varepsilon_{jk}$	Differenz der stochastischen Nutzenkomponenten zweier Produkte $j$ und $k$
$\boldsymbol{\varepsilon}$	Vektor der stochastischen Nutzenkomponenten
$\varepsilon_j$	Stochastische Nutzenkomponente von Produkt $j$
$\lambda$	Ankunftswahrscheinlichkeit
$\sigma$	Standardabweichung
$A$	Anzahl der Produktattribute
$a$	Produktattribut
$D$	Anzahl der Transaktionen
$d$	Transaktion
$E(\cdot)$	(Bedingter) Erwartungswert
$ELL(\cdot)$	Erwartete Log-Likelihood-Funktion
$F(\cdot)$	Verteilungsfunktion einer Wahrscheinlichkeitsverteilung
$f(\cdot)$	Dichtefunktion einer Wahrscheinlichkeitsverteilung
$h(\cdot)$	Bedingte Dichtefunktion der unbekannten Daten eines EM-Verfahrens
$I(\cdot)$	Indikatorfunktion
$I$	Anzahl von Iterationen eines Verfahrens
$i$	Iteration eines Verfahrens
$j, k$	Produkt
$j_d$	Produkt, das in Transaktion $d$ gekauft wurde
$\mathcal{J}$	Menge $\{0, \dots, n\}$ aller Produkte inklusive der Nichtkaufalternative
$K$	Reellwertige Konstante
$L$	Anzahl der Kundensegmente
$L(\cdot)$	Likelihood-Funktion
$LL(\cdot)$	Log-Likelihood-Funktion
$l$	Kundensegment
$l_d$	Segment, dem der für Transaktion $d$ verantwortliche Kunde angehört
$\mathcal{L}$	Menge $\{1, \dots, L\}$ aller Kundensegmente $l$
$n$	Anzahl der Produkte exklusive der Nichtkaufalternative
$P(\cdot)$	(Bedingte) Wahrscheinlichkeit
$P_j(S)$	Wahrscheinlichkeit dafür, dass Produkt $j$ gekauft wird, wenn Offer- set $S$ angeboten wird
$r_j$	Preis des Produkts $j$

$S$	Offerset
$S_d$	Offerset, welches in Transaktion $d$ angeboten wurde
$\mathbf{s}$	Vektor der Segmentanteile an der Gesamtpopulation der Nachfrager
$s_l$	Anteil des Segments $l$ an der Gesamtpopulation der Nachfrager
$U_j$	Subjektiver Nutzen von Produkt $j$ für einen Kunden
$V_j$ bzw. $V_j(\mathbf{x}_j)$	Deterministische Nutzenkomponente von Produkt $j$
$Var(\cdot)$ bzw. $\sigma^2$	Varianz
$\mathbf{x}_j$	Vektor der Attributwerte von Produkt $j$
$x_{aj}$	Wert des Attributs $a$ von Produkt $j$
$\mathbf{z}$	Vektor der unbekannten Daten eines EM-Verfahrens

## Symbole mit kapitelspezifischer Bedeutung

### Kapitel 2

$\Delta_j V_{t+1}(\mathbf{c}_t)$	Opportunitätskosten bei Verkauf von Produkt $j$ bei den Restkapazitäten $\mathbf{c}_t$ in Periode $t$
$\mathbf{A} = [a_{hj}]$	Matrix der Kapazitätsbedarfskoeffizienten
$\mathbf{A}_j$	Spaltenvektor der Matrix der Kapazitätsbedarfskoeffizienten $\mathbf{A}$
$a_{hj}$	Kapazitätsbedarf des Produkts $j$ auf Flug $h$
$\mathbf{c}_t$	Vektor der Restkapazitäten aller Flüge $h$ in Periode $t$
$c_{th}$	Restkapazität des Fluges $h$ in Periode $t$
$h$	Flug
$\mathcal{H}$	Menge $\{1, \dots, m\}$ aller Flüge
$\mathcal{J}(\mathbf{c}_t)$	Menge der Produkte $j$ , für die bei Restkapazität $\mathbf{c}_t$ noch ausreichend Kapazität vorhanden ist
$m$	Anzahl der Flüge
$P_j$	Wahrscheinlichkeit dafür, dass in beliebiger Periode Anfrage nach Produkt $j$ eintrifft
$\bar{r}_j$	Displacement-Adjusted-Revenue von Produkt $j$
$t$	Mikroperiode des Buchungshorizonts
$T$	Ende des Buchungshorizonts
$V_t(\mathbf{c}_t)$	Erwarteter Gesamterlös in Periode $t$ bei Restkapazität $\mathbf{c}_t$

### Kapitel 3

$\Delta$	Parameter zur Festlegung der Schrittweite des NR-Verfahrens
----------	---



$G_i$  Gradient der Log-Likelihood-Funktion in Iteration  $i$  des NR-Verfahrens

$H_i$  Hessematrix der Log-Likelihood-Funktion in Iteration  $i$  des NR-Verfahrens

#### Kapitel 4

$\beta$  Vektor der Attributgewichte in einer Nutzenfunktion

$\beta'$  Vektor der Attributgewichte in einer unskalierten Nutzenfunktion

$\beta_a$  Gewichtung des Attributs  $a$  in einer Nutzenfunktion

$\varepsilon'_j$  Stochastische Nutzenkomponente von Produkt  $j$  einer unskalierten Nutzenfunktion

$U'_j$  Unskalierter, subjektiver Nutzen von Produkt  $j$  für einen Kunden

$v_j$  Präferenzgewicht des Produkts  $j$

#### Kapitel 5

$\beta$  Matrix der Attributgewichte in den Nutzenfunktionen aller Segmente

$\beta_l$  Vektor der Attributgewichte in der Nutzenfunktion des Segments  $l$

$\beta_{al}$  Gewichtung des Attributs  $a$  in der Nutzenfunktion des Segments  $l$

$\gamma$  Anzahl der Elemente jeder Teilmenge von Transaktionen zur Schätzung der Startparameter des EM-Verfahrens von Train (2009)

$C_l$  Considerationset des Segments  $l$

$P_j^{\beta_l^{i,MNL}}(S)$  Kaufwahrscheinlichkeit von Produkt  $j$  aus Offerset  $S$  gemäß einem MNL-Modell mit den Attributgewichten  $\beta_l^i$

$V_{lj}$  bzw.  $V_{lj}(x_j)$  Deterministische Nutzenkomponente von Produkt  $j$  in der Nutzenfunktion des Segments  $l$

#### Kapitel 6

$\alpha_{dl}$  Kompatibilitätskoeffizient von Transaktion  $d$  und Segment  $l$

$\mu$  Vektor der Dualvariablen des Typs  $\mu_d$  des Maximum-Likelihood-Schätzproblems für ein Rank-based Modell

$\mu_d$  Dualvariable des Maximum-Likelihood-Schätzproblems für ein Rank-based Modell

$\pi$  Dualvariable des Maximum-Likelihood-Schätzproblems für ein Rank-based Modell

$(\bar{\pi}, \bar{\mu})$  Optimale Lösung des dualen RMP eines Maximum-Likelihood-Schätzproblems für ein Rank-based Modell

$(\pi^*, \mu^*)$  Optimale Lösung des dualen Maximum-Likelihood-Schätzproblems für ein Rank-based Modell

$\sigma$  Teilmenge  $\{\sigma_1, \dots, \sigma_L\}$  der Menge aller möglicher Präferenzlisten

$\sigma$	Präferenzliste
$\sigma_l$	Präferenzliste des Segments $l$
$\sigma(j)$	Position des Produkts $j$ in Präferenzliste $\sigma$
$\phi(\mathbf{s}, \mathbf{y}, \pi, \boldsymbol{\mu})$	Lagrange-Funktion des Maximum-Likelihood-Schätzproblems für ein Rank-based Modell
$\phi^\sigma(\mathbf{s}, \mathbf{y}, \pi, \boldsymbol{\mu})$	Lagrange-Funktion des RMP eines Maximum-Likelihood-Schätzproblems für ein Rank-based Modell
$H_0$	Nullhypothese des Likelihood-Ratio-Tests
$L^{H_0}$ bzw. $LL^{H_0}$	Maximale (Log-)Likelihood des Maximum-Likelihood-Schätzproblems für ein Rank-based Modell mit den durch $H_0$ implizierten Nebenbedingungen
$LL^\sigma(\mathbf{s})$	Zielfunktion des RMP eines Maximum-Likelihood-Schätzproblems für ein Rank-based Modell mit dem Parameterset $\mathbf{s}$
$\mathcal{M}_j(S)$	Menge der Segmente $l$ , deren Präferenzliste $\sigma_l$ mit dem Kauf des Produkts $j$ aus Offerset $S$ kompatibel ist
$\mathcal{M}_j^\sigma(S)$	Menge der Segmente $l$ , deren in der Teilmenge $\sigma$ aller Präferenzlisten enthaltene Präferenzliste $\sigma_l \in \sigma$ mit dem Kauf des Produkts $j$ aus Offerset $S$ kompatibel ist
$P_j^s(S)$	Kaufwahrscheinlichkeit von Produkt $j$ aus Offerset $S$ gemäß einem Rank-based-Modell mit den Segmentanteilen $\mathbf{s}$
$SP(\mathbf{x}, \mathbf{w})$	Zielfunktionswert des Subproblems des Maximum-Likelihood-Schätzproblems für ein Rank-based Modell
$(\bar{\mathbf{s}}, \bar{\mathbf{y}})$	Optimale Lösung des primalen RMP eines Maximum-Likelihood-Schätzproblems für ein Rank-based Modell
$(\mathbf{s}^*, \mathbf{y}^*)$	Optimale Lösung des primalen Maximum-Likelihood-Schätzproblems für ein Rank-based Modell
$\mathbf{w}$	Vektor der Kompatibilitätsvariablen des Subproblems des Maximum-Likelihood-Schätzproblems für ein Rank-based Modell
$w_d$	Kompatibilitätsvariable von Transaktion $d$ und einem Segment $l$ , das durch die aktuelle Lösung des Subproblems des Maximum-Likelihood-Schätzproblems für ein Rank-based Modell repräsentiert wird
$\mathbf{x}$	Vektor der Ranking-Variablen des Subproblems des Maximum-Likelihood-Schätzproblems für ein Rank-based Modell
$x_{jk}$	Ranking-Variable des Subproblems des Maximum-Likelihood-Schätzproblems für ein Rank-based Modell, die angibt, ob Produkt $j$ vor Produkt $k$ platziert ist
$(\mathbf{x}^*, \mathbf{w}^*)$	Optimale Lösung des Subproblems des Maximum-Likelihood-Schätzproblems für ein Rank-based Modell
$\mathbf{y}$	Vektor der kumulierten Anteile kompatibler Segmente für alle Transaktionen

$y_d$	Kumulierter Anteil aller mit Transaktion $d$ kompatibler Segmente $l$
<b>Kapitel 7</b>	
$\epsilon$	Beliebig kleine, positive Konstante
$\Lambda^d$	Vektor, der angibt welches Produkt der für Transaktion $d$ verantwortliche Kunde global bevorzugt
$\Lambda_j^d$	Parameter, der angibt, ob der für Transaktion $d$ verantwortliche Kunde Produkt $j$ global bevorzugt
$\lambda$	Vektor der Ankunfts wahrscheinlichkeiten
$\lambda_j$	Ankunfts wahrscheinlichkeit eines Kunden, der Produkt $j$ global bevorzugt
$\mathbf{P}^d$	Matrix der Anzahl der Zustandsübergänge im Entscheidungsprozess des für Transaktion $d$ verantwortlichen Kunden
$p_{jk}^d$	Anzahl der Zustandsübergänge von Zustand $X_t = j$ in Zustand $X_{t+1} = k$ im Entscheidungsprozess des für Transaktion $d$ verantwortlichen Kunden
$\rho$	Matrix der Übergangswahrscheinlichkeiten
$\rho_{jk}$	Übergangswahrscheinlichkeit von Zustand $X_t = j$ in Zustand $X_{t+1} = k$
$\Theta^i(S)$	Vektor der erwarteten Anzahl von Besuchen aller Produkte in Verfahrenssiteration $i$ , falls Offersset $S$ angeboten wird
$\Theta_j^i(S)$	Erwartete Anzahl von Besuchen des Produkts $j$ im Verlauf des Entscheidungsprozesses in Verfahrenssiteration $i$ , falls Offersset $S$ angeboten wird
$\Psi_{j_d}^i(j, S_d)$	Wahrscheinlichkeit in Verfahrenssiteration $i$ dafür, dass Produkt $j_d$ aus Offersset $S_d$ gewählt wird, wenn der für Transaktion $d$ verantwortliche Kunde Produkt $j$ global bevorzugt
$T$	Zeitpunkt, zu dem ein absorbierender Zustand einer Markov-Kette erreicht wird
$t$	Zeitpunkt, zu dem eine Beobachtung einer Markov-Kette stattfindet
$X_t$	Zustand einer Markov-Kette in Zeitpunkt $t$

# 1 Einführung

Die Gestaltung der Verkaufsprozesse ist eines der Kerngebiete jeder unternehmerischen Aktivität und hat einen großen Einfluss auf die Profitabilität eines Unternehmens (vgl. Talluri und van Ryzin (2004a, S. 4)). Wenn der Markt, in dem ein Unternehmen operiert, bestimmte Charakteristika aufweist, können die Methoden des Revenue Management (RM) eingesetzt werden, um die Qualität der Managemententscheidungen im Bereich der Distribution zu verbessern. Mithilfe dieser Methoden und leistungsfähiger IT-Systeme können die Entscheidungsprobleme sehr viel exakter, detaillierter und in größerem Maßstab modelliert und gelöst werden, als dies mit manuellen Verfahren möglich wäre. Die Güte der Entscheidungen hängt jedoch nicht allein von der Exaktheit der Modellierung und der Effizienz der Optimierungsverfahren ab, sondern auch von der Qualität der zur Verfügung stehenden Datengrundlage (vgl. Berbeglia et al. (2018, S. 2)). Im Kontext des RM gilt dies insbesondere für Daten, die das Kaufverhalten der Nachfrager beschreiben. Diese werden ihrerseits auf Basis vergangener Transaktionen mithilfe von Discrete-Choice-Modellen ermittelt, die sich in Bezug auf die Modellierungsgüte unterscheiden und so Daten von unterschiedlicher Qualität bereitstellen. Sowohl in der Forschung als auch in der unternehmerischen Praxis wurde die Beschreibung des Kundenwahlverhaltens bis kurz nach der Jahrtausendwende weitgehend vernachlässigt, da der Fokus auf der Weiterentwicklung der RM-Methoden lag (vgl. Berbeglia et al. (2018, S. 2) und Strauss et al. (2018, S. 375)). In der Folgezeit wurden zunächst einfache, wenig exakte Discrete-Choice-Modelle in die Methoden des RM integriert, um die zur Generierung der Inputdaten und Lösung der eigentlichen RM-Probleme benötigte Rechenzeit gering zu halten (vgl. Berbeglia et al. (2018, S. 3)). Ursächlich für diesen Zusammenhang ist, dass mit der Abbildungsgenauigkeit eines Discrete-Choice-Modells auch die Komplexität einerseits des Parameterschätzproblems und andererseits der nachgelagerten Optimierungsprobleme des RM steigt und diese oft nicht in vertretbarer Zeit lösbar sind (vgl. Aouad et al. (2018, S.1)). Für die Anwendbarkeit eines integrierten RM-Verfahrens, welches das Kundenwahlverhalten berücksichtigt, ist dies jedoch eine Grundvoraussetzung, da die Entscheidungsfindung häufig in Echtzeit erfolgen muss (vgl. Strauss et al. (2018, S. 380)).

Gegenstand aktueller Forschung ist die Entwicklung und Integration komplexer Discrete-Choice-Modelle, mit dem Ziel, das Kundenwahlverhalten exakter abzubilden und effiziente Verfahren für die Lösung der damit verbundenen Optimierungsprobleme zu entwickeln (vgl. Blanchet et al. (2016, S. 886)). Das beschriebene Spannungsfeld zwischen der Exaktheit

eines Discrete-Choice-Modells und der Lösbarkeit der Schätz- und RM-Probleme bleibt dennoch bestehen. Die Auswahl des Discrete-Choice-Modells bestimmt somit maßgeblich die Komplexität des integrierten RM-Problems und die erzielbare Modellierungsgüte, d. h. die Qualität der Managemententscheidungen, die aus der Lösung des Problems abgeleitet werden (vgl. Aouad et al. (2018, S.1)). Das Ziel dieser Arbeit ist es, durch die vergleichende Betrachtung vier verschiedener Discrete-Choice-Modelle und deren gemeinsamer theoretischer Grundlagen, die Hintergründe dieser Komplexitäts- und Genauigkeitsunterschiede aufzuzeigen. Ein weiterer Fokus liegt auf der Darstellung des Spezifikationsprozesses der Modelle sowie eines Schätzverfahrens für deren Parameter. Diese erfolgt sowohl theoretisch als auch anhand eines fortlaufenden Anwendungsbeispiels, wobei als Anwendungsfeld exemplarisch das Airline-RM betrachtet wird. Der Leser soll so in die Lage versetzt werden, abhängig von dem konkreten RM-Problem, in welches das Kundenwahlverhalten integriert werden soll, zunächst eine fundierte Modellauswahlentscheidung treffen sowie anschließend die Spezifikation und die Parameterschätzung durchführen zu können.

Das zweite Kapitel dieser Arbeit fungiert als Einführung in das RM im Allgemeinen und das Airline-RM im Besonderen. Im Zuge dessen wird verdeutlicht, in welcher Form Discrete-Choice-Modelle in die Optimierungsprobleme des RM integriert werden können. Im dritten Kapitel werden die theoretischen Grundlagen der Discrete-Choice-Modelle vermittelt, wobei zuerst auf die verschiedenen Formen des Wahlverhaltens und die Folgen einer ungenauen Modellierung dieses Verhaltens eingegangen wird. Anschließend wird gezeigt, wie die Discrete-Choice-Modelle hergeleitet und klassifiziert werden können und welche Strategien allgemein bei der Spezifikation bzw. Parameterschätzung eingesetzt werden können. In den Kapiteln vier bis sieben wird jeweils ein Discrete-Choice-Modell vorgestellt, indem die Modellstruktur, der Spezifikationsprozess und ein mögliches Schätzverfahren erläutert werden.

## **2 Anwendung der Discrete-Choice-Modelle**

Die Modellierung von Kundenwahlverhalten spielt in vielen Wirtschafts- und Unternehmensbereichen eine wichtige Rolle. In dieser Arbeit soll der Fokus auf der Anwendung im Bereich des RM liegen. Im folgenden Kapitel erfolgt zu Beginn eine Charakterisierung der allgemeinen Planungs- und Steuerungsaufgaben des RM und der damit verbundenen Optimierungsprobleme, in welche die Ergebnisse der Kundenwahlmodellierung als zentrale

Parameter eingehen. Anschließend wird beschrieben in welcher spezifischen Form diese Probleme in der Airline-Industrie auftreten. In diesem Kontext bewegt sich auch das fortlaufende didaktische Beispiel dieser Arbeit, dessen Einführung den Abschluss dieses Kapitels bildet.

## 2.1 Revenue Management

Unter dem Begriff des RM werden alle strategischen und operativen Aufgaben zusammengefasst, die sich an der Schnittstelle des Unternehmens mit den Absatzmärkten ergeben (vgl. Talluri und van Ryzin (2004a, S. 2)). Als übergeordnetes Ziel wird dabei die Maximierung des Gewinns verfolgt, wobei dieses häufig äquivalent zur Erlösmaximierung ist, wie später noch deutlich werden wird. Die dafür zur Verfügung stehenden Handlungsmöglichkeiten zielen ausschließlich auf die Steuerung der Nachfrage. Unterstützt durch umfangreiche IT-Systeme werden verschiedene Eigenschaften der Nachfrage ermittelt, z. B. deren zeitliche Verteilung oder die Höhe der damit verbundenen Zahlungsbereitschaft. Für deren Steuerung werden dann Entscheidungen in drei Handlungsfeldern getroffen (vgl. Talluri und van Ryzin (2004a, S. 3)). Erstens wird über die Produktstruktur entschieden, konkret etwa über den Einsatz von Preisdifferenzierung, mithilfe derer eine Kernleistung mit verschiedenen Restriktionen kombiniert wird, sodass unterschiedliche Produkte entstehen. Diese werden getrennt bepreist und richten sich an bestimmte Kundensegmente mit dem Ziel, den Preis möglichst nahe an die Zahlungsbereitschaft der Kunden zu setzen. Dies führt zum zweiten Handlungsfeld, in dem über die Preissetzung, also die anfänglichen Preise und deren Anpassung über den Zeitverlauf hinweg, entschieden wird. Drittens kann das Unternehmen den Verkauf der vorhandenen Produktmenge oder -kapazität beeinflussen, indem diese anteilig den verschiedenen Nachfragesegmenten oder Verkaufskanälen zugeteilt wird. Dies geschieht im Wesentlichen dadurch, dass Kundenanfragen abgelehnt oder akzeptiert werden und das den Kunden angebotene Produktsortiment variiert wird.

Die strategische oder operative Natur der beschriebenen Entscheidungen, wie auch der Handlungsspielraum in den einzelnen Entscheidungsfeldern unterscheiden sich je nach betrachtetem Unternehmen (vgl. Talluri und van Ryzin (2004a, S. 3)). So hat ein Online-Versandhändler beispielsweise im Bereich der Preissetzung viele Freiheiten und kann diese Entscheidungen kurzfristig, also auf operativer Ebene, treffen. Ein Kaufhausbetreiber dagegen muss strategisch über die Preise entscheiden und dabei möglicherweise Preisbindungen berücksichtigen. Operativ besteht für ein solches Unternehmen ein geringerer Handlungsspielraum, da jede kurzfristige Preisanpassung mit einem erheblichen Zeit- und

Kostenaufwand einhergeht.

Wie hoch das vorhandene Optimierungspotential durch die Anwendung von RM tatsächlich ist, hängt maßgeblich von der Unternehmens- und Marktstruktur ab. Es lassen sich daher Bedingungen formulieren, die auf ein für RM geeignetes Unternehmens- und Marktumfeld hindeuten:<sup>1</sup>

- Heterogene, unsichere und volatile Nachfrage: Eine heterogene Nachfrage ist gekennzeichnet durch variable Produktpräferenzen und daraus resultierende Zahlungsbereitschaften der Kunden. Dies ermöglicht beispielsweise die oben beschriebene Preisdifferenzierung. In Zusammenhang mit der Unsicherheit bzgl. der Höhe und der zeitlichen Verteilung der Nachfrage bieten sich Chancen durch die Anwendung der Kapazitätssteuerung. So kann eine Erlössteigerung etwa dadurch erreicht werden, dass Buchungsanfragen von Kunden mit niedriger Zahlungsbereitschaft abgelehnt werden, um Kapazität für später eintreffende, höherwertige Nachfrage vorzuhalten (vgl. Klein und Steinhardt (2008, S. 14-16)).
- Unflexible Produktion und hohe Fixkosten: Inflexibilität entsteht bei der Kapazitätsbereitstellung z. B. aufgrund von langen Vorlaufzeiten, Skaleneffekten oder Kapazitätsrestriktionen. Ein Unternehmen mit flexibler Produktion bzw. Kapazitätsbereitstellung kann auf Nachfrageschwankungen nahezu problemlos mit der Anpassung der Angebotsmenge reagieren (vgl. Talluri und van Ryzin (2004a, S. 14)). Ist dies nicht der Fall, besteht einerseits die Gefahr einer Unterdeckung der Nachfrage und damit entgangener Erlöse (vgl. van Ryzin und Vulcano (2015, S. 281)). Andererseits verfallen bereitgestellte, aber nicht verkaufte Kapazitätseinheiten häufig zu einem bestimmten Zeitpunkt und generieren keine Erlöse. Angesichts der – relativ zu den Grenzkosten beim Verkauf der Kapazität – meist hohen Fixkosten für deren Bereitstellung, kann das RM, etwa durch eine geeignete Preissetzung, zusätzliche niederwertige Nachfrage generieren und so die Erlöse steigern (vgl. Klein und Steinhardt (2008, S. 13f)).
- Digitale Infrastruktur: Viele der operativen RM-Entscheidungen müssen in Echtzeit und auf der Grundlage valider Daten getroffen werden. Daher ist das Vorhandensein von OLTP-Systemen, die Daten über vergangene Transaktionen speichern, verarbeiten und für automatisierte Entscheidungsprozesse zur Verfügung stellen, essentiell

---

<sup>1</sup> In der Literatur werden über die hier aufgeführten Bedingungen hinaus noch weitere Voraussetzungen für die Anwendung von RM-Instrumenten beschrieben (vgl. z. B. Klein und Steinhardt (2008, Kap. 1.2.2) oder Talluri und van Ryzin (2004a, Kap. 1.3.3)).

für die erfolgreiche Anwendung der RM-Methoden (vgl. Talluri und van Ryzin (2004a, S. 15)).

## 2.2 RM in der Airline-Industrie

Die Disziplin des RM entwickelte sich Ende der 1970er Jahre in der US-Amerikanischen Airlinebranche als Reaktion der Full-Service-Network-Carrier (FSNC) auf den Markteintritt erster Low-Cost-Carrier (LCC). Auslöser dieser Entwicklung war die Deregulierung des inländischen Marktes, die den Fluggesellschaften erstmals eine eigenständige Flugplan- und Preisgestaltung erlaubte (vgl. Talluri und van Ryzin (2004a, S. 6f)). Bis heute ist die Airline-Industrie eines der Hauptanwendungsgebiete des RM. Um die Besonderheiten dieses Anwendungsfelds zu analysieren, soll zunächst die Angebotssituation näher betrachtet werden.

Die Angebotsgestaltung von Fluggesellschaften erfolgt im Rahmen eines Prozesses, der mehrere Monate vor der eigentlichen Durchführung des Fluges beginnt und in dessen Verlauf die Anpassung des Angebots an beobachtete Schwankungen der Nachfrage zunehmend schwieriger wird. Spätestens nach dem sogenannten Fleet Assignment, der Zuweisung von Flugzeugtypen zu den einzelnen Flügen des zuvor erstellten Flugplans, ist eine Änderung der Kapazitätsbereitstellung meist nicht mehr wirtschaftlich darstellbar (vgl. Gopalan und Talluri (1998, S. 160)). Die Kosten für diese Bereitstellung der Kapazität, also Crew-, Treibstoff-, Wartungs- und sonstige Betriebskosten wie Leasing- oder Versicherungsgebühren, sind daher als Fixkosten zu betrachten. Zusammen mit den Gemeinkosten der gesamten Airline sind diese somit weitaus höher als die variablen Kosten pro Passagier, zu denen u. a. geringfügig höhere Treibstoffkosten, Sicherheitsgebühren und Steuern gehören (vgl. Belobaba (2009c, S. 118f), Bundesministerium der Finanzen (2011) und BDF (2019)). Anhand dieser Beschreibung wird deutlich, dass die Bedingung einer unflexiblen Produktion sowie hoher Fixkosten und die im RM angenommene Äquivalenz von Gewinn- und Erlösmaximierung in der Airline-Industrie gegeben sind (vgl. Strauss et al. (2018, S. 375)).

Der Vertrieb der Flugtickets erfolgt traditionell über die Globalen Distributionssysteme (GDS), die an interne Reservierungssysteme der Airlines gekoppelt sind. Die Produkte, die über die GDS angeboten werden, stellen eine Kombination aus einer Flugverbindung und einer Buchungsklasse dar (vgl. Strauss et al. (2018, S. 376)). Jedes dieser Produkte wird zu einem bestimmten, fixen Preis und in einer bestimmten Menge angeboten. Da jede Buchungsklasse mit Restriktionen verknüpft ist (z. B. Mindestaufenthalt, Vorausbuchungsfrist oder Umtauschbeschränkung), setzen die Fluggesellschaften auf diese Weise das Prinzip der



Preisdifferenzierung um (vgl. Belobaba et al. (2009, S. 448f)). Ein wesentlicher Nachteil dieses Vertriebskanals ist die Tatsache, dass dieser nicht genügend Flexibilität für einige Instrumente des RM, insbesondere im Bereich der Preissetzung bietet. Dies ist neben der Preisdifferenzierung einer der Gründe dafür, dass die FSNC hauptsächlich Verfahren aus dem Bereich der Kapazitätssteuerung einsetzen (vgl. Klein und Steinhardt (2008, S 29) und Talluri und van Ryzin (2004a, S. 176)).

Der Blick auf die Nachfragestruktur in der Airline-Industrie zeigt ebenfalls, dass die Kapazitätssteuerung als RM-Instrument sehr gut geeignet ist. So existieren nur wenige, vergleichsweise klar abgrenzbare Kundensegmente: die Geschäftsreisenden und Privatreisende sowie deren Teilsegmente. Je nach Segmentzugehörigkeit unterscheiden sich zeitliche Flexibilität und Zahlungsbereitschaft der Kunden (vgl. Belobaba (2009a, S. 63)). Die hochwertigen Buchungsanfragen der wenig flexiblen Geschäftsreisenden treffen eher kurz vor dem Abflugdatum und damit nach denen der weniger zahlungskräftigen Privatreisenden ein. Die Methoden des RM zielen in diesem Zusammenhang darauf ab, den Buchungsverlauf so zu steuern, dass am Ende des Buchungshorizonts noch genügend Kapazität für die hochwertige Nachfrage zur Verfügung steht (vgl. Belobaba (2009b, S. 88)).

Eine weitere Besonderheit des Airline-RM ist, dass die von den FSNC angebotenen Flugtickets häufig Umsteigeverbindungen mit mehrere Einzelflügen beinhalten. Dadurch entsteht die Notwendigkeit einer integrierten Steuerung eines Netzwerks von Ressourcen, denn zusätzlich zur Allokation auf Buchungsklassen muss über die Allokation der Einzelflugkapazitäten auf die verschiedenen Umsteigeverbindungen entschieden werden (vgl. Talluri und van Ryzin (2004a, S. 81)). Ziel ist auch dabei, die Differenzen hinsichtlich des Werts der Nachfrage auszunutzen, um den hochwertigen Anteil in jedem Fall zu bedienen. Mit einer Kapazitätseinheit eines Zubringerfluges etwa, die als Abschnitt einer Umsteigeverbindung verkauft wird, kann ein deutlich höherer Erlös als durch deren Verkauf als Direktflug erzielt werden (vgl. Belobaba (2009b, S. 101)).

## 2.3 Optimierungsprobleme im Airline-RM

Die Optimierungsprobleme des RM können, anhand ihrer Zugehörigkeit zu einem der in Abschnitt 2.1 beschriebenen Handlungsfelder, dem preisbasierten oder dem mengenbasierten RM zugeordnet werden. Die Abbildung des Kundenwahlverhaltens kann generell in Modelle beider Bereiche integriert werden (vgl. Talluri und van Ryzin (2004a, S. 301)). Im Rahmen dieser Arbeit soll jedoch hauptsächlich das Netzwerk-RM-Problem aus dem Gebiet der Kapazitätssteuerung betrachtet werden, da dieses in der Airline-Industrie, wie bereits

beschrieben, eine bedeutende Rolle spielt. In dem nun folgenden Abschnitt wird zunächst das generische Netzwerk-RM-Problem ohne Kundenwahlverhalten betrachtet und anschließend zum sogenannten Choice-based-Netzwerk-RM-Problem erweitert. Danach wird das Assortment-Optimierungsproblem als Teilproblem des Choice-based-Netzwerk-RM-Problems vorgestellt. Abschließend wird das fortlaufende Anwendungsbeispiel dieser Arbeit eingeführt.

### 2.3.1 Netzwerk-RM-Problem

Das Grundproblem der Kapazitätssteuerung beinhaltet die zeitliche und mengenmäßige Allokation der gegebenen Kapazität mehrerer Ressourcen auf die heterogene Nachfrage, die im Verlauf des Buchungsprozesses eintrifft (vgl. Strauss et al. (2018, S. 376) für die folgenden Überlegungen). Der damit korrespondierende Zeitraum wird als Buchungshorizont bezeichnet und in Mikroperioden  $t = 1, \dots, T$  zerlegt. Der Prozess startet mit der Periode  $t = 1$  und endet kurz vor dem Abflug mit der Periode  $t = T$ , nach der die verbleibende Restkapazität verfällt. Jedes Produkt  $j \in \mathcal{J} = \{1, \dots, n\}$  besteht aus einer Flugverbindung, die in einer bestimmten Buchungsklasse zum Preis  $r_j$  verkauft wird. Die Restkapazität ist im Verlauf des Buchungsprozesses für jede Periode  $t$  durch den Vektor  $\mathbf{c}_t$  gegeben. Jedes Element  $c_{th}$  dieser Vektoren entspricht der Sitzplatzanzahl auf einem Flug  $h \in \mathcal{H} = \{1, \dots, m\}$ . Der Ressourcenbedarf jedes Produkts wird durch die Matrix  $\mathbf{A} = [a_{hj}] \in \{0,1\}^{m \times n}$  codiert, wobei  $a_{hj} = 1$  gesetzt wird, wenn Produkt  $j$  den Flug  $h$  enthält.<sup>2</sup> Jeder Spaltenvektor  $\mathbf{A}_j$  dieser Matrix gibt folglich an, aus welchen Einzelflügen die Flugverbindung von Produkt  $j$  besteht. In der klassischen Kapazitätssteuerung ohne Kundenwahlverhalten wird davon ausgegangen, dass jeder eintreffende Kunde je nach Segmentzugehörigkeit nur bereit ist genau ein bestimmtes Produkt zu erwerben. Ist das gewünschte Produkt verfügbar, findet der Kauf in jedem Fall statt, ansonsten verlässt der Kunde den Markt. Da die Auswahl der parallel angebotenen Produkte irrelevant für das Zustandekommen des Kaufs ist, muss die Airline in jeder Mikroperiode lediglich eine binäre Entscheidung über die Annahme oder Ablehnung des Kundenwunsches treffen (vgl. Klein und Steinhardt (2008, S 98)). Somit kann die Wahrscheinlichkeit dafür, dass ein Verkauf von Produkt  $j$  möglich ist, mit  $P_j$  angegeben werden – dem Anteil der Kunden an der Population, die Produkt  $j$  bevorzugen.<sup>3</sup> Die

---

<sup>2</sup> Diese Annahme ist nur im Airline-RM gültig. In anderen Anwendungsfällen sind auch Ressourcenbedarfe  $a_{hj} > 1$  möglich.

<sup>3</sup> Es wird o. b. d. A. angenommen, dass die Wahrscheinlichkeit für das Eintreffen einer bestimmten Anfrage über den gesamten Buchungshorizont hinweg konstant bleibt (vgl. Talluri und van Ryzin (2004b, S. 17)).

Wahrscheinlichkeit dafür, dass die Airline in einer Mikroperiode  $t$  keine Anfrage erreicht, beträgt folglich  $P_0 = 1 - \sum_{j \in \mathcal{J}} P_j$ , wobei angenommen wird, dass in jeder Periode maximal ein Kunde eintrifft. Diese Annahme kann ohne Weiteres garantiert werden, indem die Dauer der Perioden ausreichend kurz gewählt wird.

Das Kapazitätssteuerungsproblem kann zum einen als stochastisch charakterisiert werden, denn es ist zu Beginn des Buchungshorizonts nicht bekannt in welcher Anzahl und Reihenfolge die Buchungsanfragen eintreffen werden. Zum anderen ist das Problem dynamisch, da der Entscheidungsprozess sequentiell abläuft und die Annahme- oder Ablehnungsentscheidung in einer Mikroperiode die Restkapazität und damit den Entscheidungsraum in den darauffolgenden Perioden beeinflusst (vgl. Klein und Steinhardt (2008, S 94f)). Zur Berechnung des optimalen erwarteten Gesamterlöses  $V_t(\mathbf{c}_t)$  und der damit verbundenen Politik kann folgende Bellman'sche Funktionalgleichung für alle Perioden  $t = 1, \dots, T$  und alle möglichen Restkapazitätsvektoren  $\mathbf{0} \leq \mathbf{c}_t \leq \mathbf{c}_1$  gelöst werden (vgl. Klein und Steinhardt (2008, S 102f)):

$$V_t(\mathbf{c}_t) = \sum_{j \in \mathcal{J}} P_j \cdot \max\{V_{t+1}(\mathbf{c}_t), r_j + V_{t+1}(\mathbf{c}_t - \mathbf{A}_j)\} + P_0 \cdot V_{t+1}(\mathbf{c}_t) \quad (2.1)$$

unter den Nebenbedingungen

$$V_{T+1}(\mathbf{c}_{T+1}) = 0 \quad \forall \mathbf{c}_{T+1} \geq \mathbf{0} \quad (2.2)$$

$$V_t(\mathbf{c}_t) = -\infty \quad \text{falls } \exists c_{th} < 0 \ \forall h \in \mathcal{H}, t = 1, \dots, T \quad (2.3)$$

Die ersten Summanden der Wertfunktion repräsentieren jeweils eine eintreffende Anfrage für Produkt  $j$ , wobei das erste Element des Maximum-Operators die Ablehnung und das zweite die Annahme der Anfrage codiert. Der letzte Summand bildet den Fall ab, dass keine Anfrage eintrifft. Es wird deutlich, dass der Verzicht auf die explizite Modellierung des Kundenwahlverhaltens den Vorteil eines sehr kleinen Entscheidungsraums mit sich bringt. Da diese Vereinfachung jedoch auch einige nachteilige Aspekte aufweist (vgl. Kap. 3), wird nun gezeigt in welcher Weise ein Kundenwahlmodell in die exakte Formulierung des Kapazitätssteuerungsproblems integriert werden kann.

### 2.3.2 Choice-based-Netzwerk-RM-Problem

Ausgangspunkt der Modellerweiterung ist die Annahme, dass die Entscheidung des Kunden nicht vor, sondern während des Buchungsvorgangs durch einen Vergleich sämtlicher

angebotener Produkte gefällt wird (vgl. z. B. Miranda Bront et al. (2009, S. 772)). Daraus ergeben sich zwei unmittelbare Konsequenzen: Erstens müssen die Kaufwahrscheinlichkeiten der einzelnen Produkte nun in Abhängigkeit des Produktangebots, aus dem ein ankommender Kunde wählen kann, formuliert werden. Diese Teilmenge aller Produkte wird als Offerset  $S \subseteq \mathcal{J}$  bezeichnet, die Kaufwahrscheinlichkeiten entsprechend mit  $P_j(S)$  (vgl. Strauss et al. (2018, 367f)). Zweitens ist eine Unterscheidung notwendig zwischen den Perioden, in denen kein Kunde eintrifft, und denen, in welchen zwar ein Kunde eintrifft, sich aber gegen einen Kauf entscheidet, um beispielsweise das Angebot eines Wettbewerbers zu nutzen. Der erste Fall wird durch die Einführung einer Ankunfts-wahrscheinlichkeit  $\lambda$  abgebildet, der zweite durch eine Nichtkaufalternative  $j = 0$ , die in jedem Offerset enthalten sein muss (vgl. van Ryzin und Vulcano (2015, S. 284)).<sup>4</sup> Es gilt der Zusammenhang  $P_0(S) = 1 - \sum_{j \in S} P_j(S)$ . Somit ergibt sich mit  $\mathcal{J}(\mathbf{c}_t) = \{j \in \mathcal{J} | \mathbf{A}_j \leq \mathbf{c}_t\}$  folgende Wertfunktion (vgl. Strauss et al. (2018, S. 377)):

$$\begin{aligned}
 V_t(\mathbf{c}_t) &= \\
 \max_{S \subseteq \mathcal{J}(\mathbf{c}_t)} &\left\{ \sum_{j \in S} \lambda P_j(S) (r_j + V_{t+1}(\mathbf{c}_t - \mathbf{A}_j)) + (\lambda P_0(S) + 1 - \lambda) V_{t+1}(\mathbf{c}_t) \right\} \\
 &= \max_{S \subseteq \mathcal{J}(\mathbf{c}_t)} \left\{ \sum_{j \in S} \lambda P_j(S) (r_j - \Delta_j V_{t+1}(\mathbf{c}_t)) \right\} + V_{t+1}(\mathbf{c}_t)
 \end{aligned} \tag{2.4}$$

unter den Nebenbedingungen<sup>5</sup>

$$V_{T+1}(\mathbf{c}_{T+1}) = 0 \quad \forall \mathbf{c}_{T+1} \geq \mathbf{0} \tag{2.5}$$

Aus der Differenz des Erlöses  $r_j$  und der Opportunitätskosten  $\Delta_j V_{t+1}(\mathbf{c}_t) = (V_{t+1}(\mathbf{c}_t) - V_{t+1}(\mathbf{c}_t - \mathbf{A}_j))$  ergibt sich der sogenannte Displacement-Adjusted-Revenue (DAR)  $\bar{r}_j = r_j - \Delta_j V_{t+1}(\mathbf{c}_t)$  von Produkt  $j$  (vgl. Strauss et al. (2018, S. 377)). Die optimale Politik beinhaltet nun nicht mehr nur eine Annahme-/Ablehnungsentscheidung, sondern die Entscheidung darüber welches Offerset in welcher Periode gewählt werden sollte, um den erwarteten Gesamterlös zu maximieren.

<sup>4</sup> Im Folgenden wird die Nichtkaufalternative als Produkt aufgefasst, d. h. es gilt  $\mathcal{J} = \{0, \dots, n\}$ .

<sup>5</sup> Auf die Randbedingungen des Typs (2.3) kann verzichtet werden, indem in jeder Periode nur diejenigen Offersets  $S \subseteq \mathcal{J}(\mathbf{c}_t) \subseteq \mathcal{J}$  in zulässiger Weise angeboten werden können, für die noch ausreichend Kapazität vorhanden ist.

Die Berechnung der Parameter  $\lambda$  und  $P_j(S)$  kann nur auf Basis eines Discrete-Choice-Modells erfolgen, das den Entscheidungsprozess des Kunden, der jetzt von Bedeutung ist, so exakt wie möglich abbildet. Ein ungeeignetes Discrete-Choice-Modell führt zu fehlerhaften Inputdaten für die Kapazitätssteuerung und damit zwangsläufig zu einer schlechteren Qualität der ermittelten Politik oder anderer Steuerungsinstrumente wie der Bid-Preise (vgl. Berbeglia et al. (2018, S. 2) und Abschnitt 3.3).

### 2.3.3 Assortment-Optimierungsproblem

Dass die Genauigkeit des Discrete-Choice-Modells zwar eine notwendige, aber keine hinreichende Bedingung für ein gutes Gesamtverfahren zur Lösung des Netzwerk-RM-Problems ist, wird bei der Betrachtung des Maximum-Operators in der Wertfunktion ersichtlich. Wie im klassischen Netzwerk-RM-Problem wird durch die Berechnung dieses Maximums in jeder Periode die optimale Entscheidung bestimmt. Der Entscheidungsraum ist nun jedoch nicht mehr binär, sondern besteht in jeder Periode  $t$  aus  $2^{|\mathcal{J}(c_t)|}$  zulässigen Offersets (vgl. Strauss et al. (2018, S. 377)). Wie effizient dieses komplexe Teilproblem des Choice-based-Netzwerk-RM-Problems, das auch als Assortment-Optimierungsproblem bezeichnet wird, gelöst werden kann, hängt bei gegebenen Opportunitätskosten von dem Modell ab, welches der Berechnung der Kaufwahrscheinlichkeiten  $P_j(S)$  zugrunde liegt (vgl. Berbeglia et al. (2018, Appendix A8)). Folglich wird auch die exakte Lösung des übergeordneten Netzwerk-RM-Problems aufwendiger, wenn nicht sogar praktisch unmöglich. Die Anwendung der Discrete-Choice-Modelle wird in dieser Arbeit im Kontext der Assortment-Optimierung vorgestellt, weil ein Vergleich verschiedener Modelle anhand dieses Problems einen guten Überblick über die Auswirkungen der Modellauswahl auf Lösungsqualität und Lösbarkeit grundlegender RM-Probleme liefert.

### 2.3.4 Einführung des Rechenbeispiels

Zum Ende dieses Kapitels soll das didaktische Beispiel vorgestellt werden, anhand dessen die Datenbasis sowie, in den folgenden Kapiteln, die Verfahren zur Parameterschätzung veranschaulicht werden.

Betrachtet wird eine Fluggesellschaft, die das Städtepaar Augsburg - München mit zwei täglichen Flügen verbindet, wobei die Hinflüge und Rückflüge jeweils getrennt gesteuert

werden.<sup>6</sup> Der morgendliche Flug  $h = 1$  startet um 10 Uhr, der Abendflug  $h = 2$  um 18 Uhr von Augsburg nach München. Jeder Flug wird in den Buchungsklassen  $N$  (Normalpreis; Stornierung oder Umbuchung möglich) und  $E$  (Ermäßigter Preis; keine Stornierung oder Umbuchung) angeboten. Insgesamt ergeben sich so vier Produkte  $j = 1, \dots, 4$  mit den Bezeichnungen  $1N$ ,  $1E$ ,  $2N$  und  $2E$ . Die Nichtkaufalternative wird mit dem Index  $j = 0$  bezeichnet. Da die Airline für den Ticketvertrieb an ein GDS angebunden ist, bietet sie jedes Produkt zu dem folgenden konstanten Preis an:  $r_1 = 200$ ,  $r_2 = 100$ ,  $r_3 = 130$  und  $r_4 = 60$ . Die Preisunterschiede begründen sich einerseits in der grundsätzlich höheren Nachfrage nach der Verbindung am Vormittag und andererseits in der Preisdifferenzierung, welche die Airline einsetzt. Konkret soll u. a. durch die ermäßigten Tarife die Auslastung gesteigert werden. Die Restriktionen bzgl. Stornierung und Umbuchung sollen durch den Mechanismus der Selbstselektion dafür sorgen, dass zahlungskräftige Geschäftsreisende trotzdem zum Normalpreis buchen (vgl. Klein und Steinhardt (2008, S 44)). Aus ihrem internen OLTP-System kann die Airline Daten zu vergangenen Buchungsprozessen abrufen und für die Modellierung des Kundenwahlverhaltens nutzen. Der Datensatz zu jeder Periode enthält das jeweilige Offersset und das verkaufte Produkt bzw. die Information, dass der angekommene Kunde kein Produkt gekauft hat oder während der Periode kein Kunde angekommen ist.<sup>7</sup> Abgesehen von deren Kaufentscheidung liegen der Airline keine weiteren Daten zu den Kunden vor. Insbesondere ist die Segmentzugehörigkeit der Kunden nicht bekannt.

Die Datengrundlage des Rechenbeispiels umfasst zehn Transaktionen  $d = 1, \dots, 10$ , die jeweils die Informationen  $(j_d, S_d)$  enthalten. Das in Transaktion  $d$  gekaufte Produkt  $j \in \mathcal{J}$  wird mit dem Parameter  $j_d$  bezeichnet, das angebotene Offersset mit  $S_d$ . Vereinfachend wird von einer Ankunftsrate  $\lambda = 1$  ausgegangen, sodass jeder Datensatz  $d$  die Entscheidung eines tatsächlich eingetroffenen Kunden abbildet. Für die Generierung der in Tabelle 1 angegebenen Transaktionsdaten wurde angenommen, dass die Kunden in zwei homogene Segmente  $l \in \mathcal{L} = \{1, 2\}$  unterteilt werden können. Das Segment  $l = 1$ , zu dem 40% der Kunden gehören, umfasst Geschäftsreisende, die Wert auf Flexibilität legen, wenig preissensitiv sind und den Vormittagsflug bevorzugen. Daher präferiert dieses Segment das Produkt  $j = 1$ , ist aber auch bereit anstelle dessen das Produkt  $j = 2$  und, falls dieses ebenfalls nicht verfügbar

---

<sup>6</sup> Basierend auf dieser Annahme werden im Rahmen des Beispiels nur die beiden Hinflüge betrachtet.

<sup>7</sup> Die Unterscheidung von Perioden, in denen kein Kunde eingetroffen ist, und Perioden, in denen ein Kunde eingetroffen ist, sich aber gegen einen Kauf entschieden hat, ist nur dann möglich, wenn die Ankunft eines Kunden informationstechnisch separat erfasst werden kann. Im sogenannten Censored-Demand-Fall, in dem diese Erfassung nicht möglich ist, ergeben sich komplexere Parameterschätzprobleme. Diese können nicht mit den Verfahren für den Uncensored-Demand-Fall gelöst werden, die in dieser Arbeit ausschließlich vorgestellt werden (vgl. z. B. van Ryzin und Volcano (2015, S. 297)).

ist, das Produkt  $j = 4$  zu kaufen. Ist keines dieser Produkte verfügbar, buchen die Kunden dieses Segments bei einem Wettbewerber. Dieses Wahlverhalten kann durch die Präferenzliste  $\sigma_1 = (1,2,4,0,3)$  beschrieben werden. Die übrigen 60% der Kunden bilden das Segment der Privatreisenden  $l = 2$ . Dieses weist eine hohe Preissensitivität auf und ist daher nur bereit, die beiden günstigsten Produkte zu kaufen, wobei  $j = 4$  gegenüber  $j = 2$  in der Präferenzliste  $\sigma_2 = (4,2,0,3,1)$  bevorzugt wird.<sup>8</sup> Die in Tabelle 1 enthaltene Information  $l_d$  über das Segment des für die jeweilige Transaktion  $d$  verantwortlichen Kunden dient lediglich der Veranschaulichung und steht in der Realität nicht zur Verfügung.

$d$	1	2	3	4	5	6	7	8	9	10
$S_d$	{0,1,2 , 3,4}	{0,3,4}	{0,2,4}	{0,1,3}	{0,1,2 , 3,4}	{0,1,4}	{0,2,3}	{0,1,4}	{0,1,2 , 4}	{0,1}
$j_d$	1	4	2	0	4	1	2	4	1	0
$l_d$	1	2	1	2	2	1	2	2	1	2

**Tabelle 1: Transaktionsdaten des fortlaufenden Anwendungsbeispiels**

In Tabelle 2 und Tabelle 3 sind die Daten zu Produktangebot und Kundenpopulation des Rechenbeispiels zusammengefasst.

Produktindex $j$	1	2	3	4
Bezeichnung	1N	1E	2N	2E
Buchungsklasse	N	E	N	E
Flugverbindung $h$	1	1	2	2
Abflugzeit	10 Uhr	10 Uhr	18 Uhr	18 Uhr
Flugpreis $r_j$	200	100	130	60

**Tabelle 2: Daten der Produkte des fortlaufenden Anwendungsbeispiels**

<sup>8</sup> Das für die Datengenerierung verwendete Modell gehört zur Klasse der Rank-based-Modelle, die in Kap. 6 ausführlich behandelt wird. Da die Produkte ab der Nichtkaufalternative unter keinen Umständen gekauft werden und deren Reihenfolge damit irrelevant ist, können die Präferenzlisten auch in verkürzter Form  $\sigma_1 = (1,2,4)$  und  $\sigma_2 = (4,2)$  dargestellt werden.

Segmentindex $l$	1	2
Bezeichnung	Geschäftsreisende	Privatreisende
Präferenzliste $\sigma_l$	(1,2,4,0,3)	(4,2,0,3,1)
Anteil	0,4	0,6

Tabelle 3: Daten der Kundensegmente des fortlaufenden Anwendungsbeispiels

### 3 Grundlagen der Discrete-Choice-Modelle

Im Fokus des dritten Kapitels steht nun die eigentliche Modellierung des Kundenwahlverhaltens. Nachdem im ersten Abschnitt der Begriff des Offersets weiter formalisiert wird, folgt die Einführung des trivialen Independent-Demand-Modells, anhand dessen die Folgen einer zu stark vereinfachenden Modellierung des Kundenwahlverhaltens veranschaulicht werden. Anschließend werden die Substitutionseffekte, die von einem Wahlmodell abgebildet werden müssen, näher beschrieben. Der mathematische Hintergrund aller Discrete-Choice-Modelle, die Random-Utility-Theory (RUT), wird im vierten Abschnitt des Kapitels eingeführt. In den beiden darauffolgenden Abschnitten werden die Grundprinzipien der Modellspezifikation und die Methoden der Parameterschätzung, jeweils bezogen auf bestimmte Modellklassen, erläutert. Den Abschluss des Kapitels bildet ein Abschnitt zu den Kriterien, mithilfe derer ein geeignetes Modell identifiziert werden kann.

#### 3.1 Spezifikation des Offersets

Als Offerset, bzw. allgemein als Choice-Set, wird die Menge der Alternativen bezeichnet, aus denen der Entscheidungsträger, d. h. im Airline-RM der Passagier, wählen kann. Damit der Entscheidungsprozess mithilfe eines Discrete-Choice-Modells abgebildet werden kann, muss jedes Offerset folgende Eigenschaften aufweisen (vgl. Train (2009, S. 11-14)):

- Genau eine Alternative wird gewählt: Damit diese Eigenschaft erfüllt ist, müssen die Alternativen so definiert sein, dass sie einander ausschließen und alle möglichen Entscheidungen abdecken. Der erste Teil der Eigenschaft wird im Fall des Offersets im Airline-RM dadurch sichergestellt, dass ausschließlich Einzelbuchungen betrachtet werden, jeder Kunde also höchstens ein Produkt erwirbt (vgl. Klein und Steinhardt



(2008), S. 85 und 96). Um den zweiten Teil zu erfüllen, reicht es nicht aus, alle angebotenen Produkte als Alternativen zu definieren, denn jeder Kunde hat die Möglichkeit keines dieser Produkte zu wählen (vgl. z. B. Berbeglia et al. (2018, S. 5)). Für die Kunden stellen die Produkte der Wettbewerber oder das Verlassen des Marktes daher ebenfalls Alternativen dar. Für welche der letztgenannten Alternativen sich ein Kunde entscheidet, ist für die Airline nicht beobachtbar und auch nicht relevant, da keine dieser Entscheidungen mit einem Erlös oder Kapazitätsverbrauch einhergeht. Die sogenannte Nichtkaufalternative ( $j = 0$ ), die in jedem Offerset enthalten sein muss, fasst diese Entscheidungen zusammen und garantiert daher, dass jeder Kunde eine im Offerset repräsentierte Entscheidung trifft (Kauf oder Nichtkauf).

- Die Menge der Alternativen ist endlich: Diese Eigenschaft schließt die Anwendung der Discrete-Choice-Modelle in Situationen aus, in denen potentiell unendlich viele Entscheidungen getroffen werden können. Dies ist der Fall, wenn die Variable, welche die Entscheidung codiert, kontinuierlich ist. Ein hypothetisches Beispiel hierfür ist die Versteigerung von Flugtickets, bei der ein Kunde über die Höhe seines Gebots entscheidet. Da ein Offerset jedoch stets aus einer endlichen Menge an Produkten und der Nichtkaufalternative besteht, ist diese Bedingung in der Kapazitätssteuerung immer erfüllt.

## 3.2 Substitutionsverhalten

Substitutionsverhalten liegt immer dann vor, wenn ein Kunde eines bestimmten Segments nicht das Produkt kauft, das er grundsätzlich gegenüber allen anderen präferiert (vgl. Kök und Fisher (2007, S. 1001)). Im Airline-RM bedeutet dies häufig, dass die Selbstselektion nicht greift, d. h. der Kunde ein Produkt kauft, welches im Rahmen der Preisdifferenzierung nicht für das Segment vorgesehen ist, dem er angehört. Grundsätzlich werden drei Arten von Substitutionsverhalten unterschieden, die sich anhand der in Abschnitt 2.3.4 vorgestellten Beispielinstantz veranschaulichen lassen (vgl. Vulcano et al. (2010, S. 372)):

- Buy-up: Ein Kunde bucht einen Flug in einer teureren Buchungsklasse, obwohl er eigentlich eine günstigere Buchungsklasse bevorzugt. In der Beispielinstantz liegt

dieser Fall vor, wenn ein Privatreisender anstelle des Produkts  $j = 2$  (Ermäßigter Preis) das Produkt  $j = 1$  (Normalpreis) kauft.<sup>9</sup>

- Buy-down: Ein Kunde kauft ein Produkt in einer günstigeren Buchungsklasse, obwohl er grundsätzlich bereit ist den gleichen Flug in einer teureren Buchungsklasse zu buchen. Ein Beispiel hierfür stellt die Transaktion  $d = 3$  des Rechenbeispiels dar, in der ein Geschäftsreisender aufgrund der Nichtverfügbarkeit von Produkt  $j = 1$  (Normalpreis) das Produkt  $j = 2$  (Ermäßigter Preis) kauft.
- Diversion: Ein Kunde bucht einen alternativen Flug, um ein Ticket in seiner bevorzugten Buchungsklasse erwerben zu können. Dieses Substitutionsverhalten tritt in der Transaktion  $d = 7$  des Rechenbeispiels auf. Der Privatreisende entscheidet sich hier für Produkt  $j = 2$ , d. h. für den Vormittagsflug, da der abendliche Flug nicht mehr zum ermäßigten Preis (Produkt  $j = 4$ ) verfügbar ist.

Mit dem Einsatz eines Discrete-Choice-Modells wird das Ziel verfolgt dieses Substitutionsverhalten so exakt wie möglich abzubilden. Im folgenden Abschnitt soll nun zunächst ein Modell betrachtet werden, das auf die Abbildung des Substitutionsverhaltens gänzlich verzichtet.

### 3.3 Independent Demand

Das Independent-Demand-Modell ergibt sich aus der Annahme, dass die Nachfrage nach den einzelnen Produkten unabhängig von der Verfügbarkeit anderer Produkte ist. Wie in Kap. 2 beschrieben, resultiert diese Annahme in den Kaufwahrscheinlichkeiten  $P_j$  für alle Produkte  $j$ , die nicht von dem jeweiligen Offerset  $S$  abhängen, und liegt somit dem traditionellen Problem der Kapazitätssteuerung zugrunde. Obwohl die Annahme die Existenz von Kundenwahlverhalten de facto ausschließt, kann diese als einfaches Discrete-Choice-Modell interpretiert werden, mit dem die Wahrscheinlichkeiten  $P_j(S)$  für den Kauf eines Produkts  $j$  abhängig von Offerset  $S$  berechnet werden können (vgl. Berbeglia et al. (2018, S. 9)):

$$P_j(S) = \begin{cases} s_j & , \forall j \in S \setminus \{0\} \\ s_j + \sum_{k \in J \setminus S} s_k & , \text{für } j = 0 \\ 0 & , \text{sonst} \end{cases} \quad (3.1)$$

---

<sup>9</sup> Falls die Kunden des Segments  $l = 2$  deterministisch anhand der Präferenzliste  $\sigma_2 = (4,2,0,3,1)$  entscheiden, tritt der Fall in der Beispielinstantz nur dann tatsächlich auf, wenn das unzulässige Offerset  $S = \{1\}$  angeboten wird (Eigenschaft der Vollständigkeit nicht erfüllt). Alle anderen Offersets, die weder  $j = 4$  noch  $j = 2$  enthalten, haben einen Nichtkauf zur Folge.

Wird für die Parameter  $s_j$  für die Produkte  $j = 0, \dots, 4$  des Anwendungsbeispiels  $\mathbf{s} = (0,1; 0,3; 0,2; 0,1; 0,3)$  angenommen, ergeben sich beispielhaft folgende Kaufwahrscheinlichkeiten für das Offersset  $S = (0,1,4)$ :

$$P_1(S) = s_1 = 0,3; P_4(S) = 0,3; P_0(S) = s_0 + s_2 + s_3 = 0,4; P_2(S) = P_3(S) = 0 \quad (3.2)$$

Der wesentliche Vorteil dieses Modells ist die geringe Komplexität der Spezifikation, der Parameterschätzung und der Assortment-Optimierung (vgl. Talluri und van Ryzin (2004a, S. 302f)).

Sofern die Nachfrager nur in sehr geringem Maße Substitutionsverhalten zeigen, stellen die Kaufwahrscheinlichkeiten des Independent-Demand-Modells eine gute Näherung für die tatsächlichen Wahrscheinlichkeiten dar. Im RM der FSNC war dies lange Zeit der Fall, da die Tarife mit vielen unterschiedlichen Produktrestriktionen versehen waren und die Preisdifferenzierung nahezu ideal umsetzbar war (vgl. Belobaba (2009b, S. 85)). In der Airline-Industrie, aber auch in anderen Branchen, führten Veränderungen des Marktumfelds in den vergangenen Jahrzehnten jedoch dazu, dass die Substitutionseffekte umfangreicher und bedeutender wurden. In der Literatur wird dies hauptsächlich auf die folgenden Aspekte zurückgeführt:

- Wachstum der LCC: Der Markteintritt und das rasche Wachstum der LCC hat mehrere Mechanismen zur Folge, die Substitutionsverhalten fördern. Zum einen entsteht dadurch ein intensiverer Wettbewerb, wodurch sich für die Kunden insgesamt mehr Wahlmöglichkeiten hinsichtlich der Flugverbindung und des Flugpreises bieten (vgl. Strauss et al. (2018, S. 375) und Vulcano et al. (2010, S. 372)). Zum anderen sind die Tarife der LCC sehr viel einfacher strukturiert als die der FSNC. Insbesondere verzichten die LCC auf einen Großteil der gängigen Produktrestriktionen, was die Attraktivität ihrer Produkte für viele Kunden erhöht. Um nicht noch mehr Marktanteile an die LCC zu verlieren, sind die FSNC daher gezwungen ihre eigene Tarifstruktur ebenfalls zu vereinfachen, wodurch die Effektivität der Preisdifferenzierung stark ab- und damit das Substitutionsverhalten zunimmt (vgl. Belobaba (2009b, S. 87)).
- Digitaler Vertrieb: Durch die Digitalisierung können die Airlines neue Vertriebskanäle, wie z. B. Vergleichsportale oder den Direktvertrieb über die eigene Webseite, erschließen. Für die Kunden erhöht sich dadurch die Informationstransparenz und der Vergleich alternativer Angebote wird vereinfacht (vgl. Belobaba (2009b, S. 85)). So wird es wahrscheinlicher, dass die Kunden den Buchungsprozess nicht mit dem

Wunsch nach einem konkreten Produkt beginnen, sondern verschiedene Angebote vergleichen, um die Kaufentscheidung erst danach auf Basis eines Nutzenkalküls zu treffen. Ein Privatreisender etwa, der seine Reisedaten zeitlich flexibel anpassen kann, wird nun sehr wahrscheinlich die Flugpreise mehrerer aufeinanderfolgender Abflugtage vergleichen und sich erst dann für das Abflugdatum mit dem günstigsten Preis entscheiden

- Sinkende Geschäftsreisebudgets: Als Folge der New-Economy-Krise im Jahr 2000 und der Weltfinanzkrise ab dem Jahr 2007 wurden die Budgets für Geschäftsreisen in vielen Firmen gekürzt, sodass die Angestellten vermehrt auf niedrigere Beförderung- und Buchungsklassen ausweichen mussten (vgl. Belobaba (2009b, S. 85) und Manager Magazin (2009)). Dies führt unmittelbar zu einer erhöhten Buy-down-Wahrscheinlichkeit, die das Independent-Demand-Modell nicht abbilden kann.

Die Ergebnisqualität des Independent-Demand-Modells und damit auch die der daran gekoppelten Kapazitätssteuerung wird unter den veränderten Rahmenbedingungen zunehmend schlechter. Besonders deutlich werden die Folgen dieser unpräzisen Modellierung des Kundenwahlverhaltens an dem sogenannten „Spiral-down-Effekt“ (vgl. Belobaba (2009b, S. 109)). Ursächlich für diesen Effekt ist die fehlende Abbildung des Buy-down-Verhaltens von Nachfragern mit einer hohen Zahlungsbereitschaft. Im Rechenbeispiel dieser Arbeit ist dies gleichbedeutend mit Geschäftsreisenden aus dem Segment  $l = 1$ , die anstelle des Produkts  $j = 1$  das Produkt  $j = 2$  erwerben. Da das Independent-Demand-Modell davon ausgeht, dass die Geschäftsreisenden ausschließlich den Kauf von Produkten der Buchungsklasse „Normalpreis“ in Erwägung ziehen, wird die Nachfrage nach Produkt  $j = 2$ , die durch Buy-down entsteht, fälschlicherweise als Nachfrage aus dem Segment der Privatreisenden klassifiziert. Es kommt also zu einer Unterschätzung der Nachfrage nach der hochpreisigen Buchungsklasse (Parameter  $s_1$  und  $s_3$ ) und einer Überschätzung der Nachfrage nach dem ermäßigten Tarif ( $s_2$  und  $s_4$ ). Im darauffolgenden Buchungshorizont wird als Folge dieser ungenauen Schätzung ein größerer Anteil der Kapazität in der ermäßigten Buchungsklasse angeboten. Da sich für die Geschäftsreisenden so wiederum mehr Buy-down-Möglichkeiten ergeben, gehen die Ticketverkäufe in der Buchungsklasse „Normalpreis“ weiter zurück. Diese Entwicklung führt letztendlich zu drastischen Erlöseinbußen für die Airline und in manchen Fällen sogar dazu, dass sämtliche Kapazität in der günstigsten Buchungsklasse angeboten wird (vgl. Cooper et al. (2006, S. 969)).

### 3.4 Random-Utility-Theory

Die Random-Utility-Theory (RUT) bildet die Grundlage für die meisten Discrete-Choice-Modelle. Im Zentrum steht die Annahme, dass die Kunden rational entscheiden und dabei danach streben, den subjektiven Nutzen ihrer Entscheidung zu maximieren. Diese Entscheidung treffen die Kunden, indem sie die angebotenen Alternativen hinsichtlich ihres Nutzens vergleichen und die Alternative mit dem größten Nutzen wählen. Im folgenden Abschnitt soll zunächst die Nutzenfunktion, von der die RUT ausgeht, formuliert und untersucht werden. Anschließend wird gezeigt wie daraus die Kaufwahrscheinlichkeiten  $P_j(S)$  abgeleitet werden. In den darauffolgenden Abschnitten werden die Implikationen zweier zentraler Grundsätze der Discrete-Choice-Modelle beschrieben, die sich ebenfalls aus der RUT ergeben.

#### 3.4.1 Nutzenfunktion

Der subjektive Nutzen, den ein Kunde einem Produkt  $j \in \mathcal{J}$  beimisst, wird mit dem Parameter  $U_j$  bezeichnet (vgl. Train (2009, S. 14f) für die folgenden Überlegungen). Aus Sicht der Fluggesellschaft handelt es sich dabei um eine Zufallsvariable, denn die Werte  $U_j$  sind ihr, im Unterschied zu dem Kunden selbst, nicht bekannt. Beobachten kann die Airline lediglich verschiedene Attribute  $\mathbf{x}_j$  der Produkte, wie etwa den Preis oder den Zeitpunkt des Abflugs, von denen der Nutzen des Kunden abhängt. Dadurch ist sie in der Lage eine Funktion  $V_j(\mathbf{x}_j)$  zu formulieren, welche als deterministische Komponente von  $U_j$  angesehen werden kann (vgl. Strauss et al. (2018, S. 377)). Die Struktur des deterministischen Nutzens  $V_j(\mathbf{x}_j)$  ist für jedes Modell unterschiedlich, da die Funktion nicht nur von den Attributwerten, sondern auch von Modellparametern abhängig ist, deren Werte statistisch geschätzt werden müssen. Sämtliche Einflussfaktoren, die für die Airline nicht beobachtbar sind, werden in der stochastischen Nutzenkomponente  $\varepsilon_j$  zusammengefasst, welche einer bestimmten Verteilung unterliegt. Zu diesen Faktoren werden u. a. Messfehler bzgl. der deterministischen Komponente, nicht beobachtbare Faktoren, wie die subjektive Komfortwahrnehmung einer Flugverbindung, und die Variabilität der Kundenpräferenzen gezählt (vgl. Cascetta (2009, S. 94f)). Letztere kann etwa dadurch entstehen, dass Kunden in einer bestimmten Entscheidungssituation mehr oder weniger preissensitiv sind als üblich. Der Gesamtnutzen eines Produkts  $j$  ergibt sich somit aus der Summe der beiden Komponenten gemäß:

$$U_j = V_j + \varepsilon_j \tag{3.3}$$

Wenn mehreren Kunden ein identisches Offersset angeboten wird, dann entspricht  $V_j$  dem Erwartungswert von  $U_j$ . Die Varianz von  $U_j$  hängt allein von der Varianz des stochastischen Nutzens ab, was in den folgenden Definitionen resultiert (vgl. Cascetta (2009, S. 91)):

$$E(U_j) = E(V_j) = V_j \text{ und } Var(U_j) = Var(\varepsilon_j) \text{ mit } E(\varepsilon_j) = Var(V_j) = 0 \quad (3.4)$$

Die Kaufentscheidung eines Kunden ergibt sich aus  $j = \operatorname{argmax}_{k \in S} \{U_k\}$ , d. h. der Nutzen des gewählten Produkts  $j$  ist größer als der aller anderen Produkte und es gilt  $U_j > U_k \forall k \neq j, k \in S$  (vgl. Train (2009, S. 14f)). Mithilfe dieses Ausdrucks lässt sich nun die Kaufwahrscheinlichkeit  $P_j(S)$  als bedingte Wahrscheinlichkeit dafür berechnen, dass der Kunde bei gegebenem Offersset  $S$  das Produkt  $j$  auswählt:

$$P_j(S) = P(U_j > U_k \forall k \neq j, k \in S) \quad (3.5)$$

Durch Einsetzen von (3.3) und Umformen der Ungleichung ergibt sich:

$$\begin{aligned} P_j(S) &= P(V_j + \varepsilon_j > V_k + \varepsilon_k \forall k \neq j, k \in S) \\ &= P(V_j - V_k > \varepsilon_k - \varepsilon_j \forall k \neq j, k \in S) \end{aligned} \quad (3.6)$$

Folglich hängt die Kaufwahrscheinlichkeit von Produkt  $j$  aus Sicht der Fluggesellschaft einerseits von den deterministischen Nutzenkomponenten  $V_k$  der übrigen im Offersset  $S$  enthaltenen Produkte und andererseits von der multivariaten Wahrscheinlichkeitsverteilung ab, welcher der Vektor  $\boldsymbol{\varepsilon} = (\varepsilon_0, \dots, \varepsilon_n)^T$  folgt. Da die Differenzen  $\Delta V_{jk} = V_j - V_k$  bekannt und die Differenzen  $\Delta \varepsilon_{kj} = \varepsilon_k - \varepsilon_j$  unsicher sind, stellt der Ausdruck (3.6) eine multivariate Verteilungsfunktion der Form  $F_{\Delta \boldsymbol{\varepsilon}}(\Delta \mathbf{V}) = P(\Delta \boldsymbol{\varepsilon} < \Delta \mathbf{V})$  dar (vgl. Cramer und Kamps (2017, S. 196)). Daher kann  $P_j(S)$  auch in Form eines multidimensionalen Integrals über die Dichtefunktion  $f(\boldsymbol{\varepsilon})$  des Zufallsvektors  $\boldsymbol{\varepsilon}$  ausgedrückt werden, wobei die Bedingung in (3.6) mithilfe einer Indikatorfunktion  $\mathbf{I}$  geprüft wird (vgl. Train (2009, S. 15)):

$$\begin{aligned} P_j(S) &= P(\varepsilon_k - \varepsilon_j < V_j - V_k \forall k \neq j, k \in S) \\ &= \int_{\boldsymbol{\varepsilon}} \mathbf{I}(\varepsilon_k - \varepsilon_j < V_j - V_k \forall k \neq j, k \in S) f(\boldsymbol{\varepsilon}) d\boldsymbol{\varepsilon} \end{aligned} \quad (3.7)$$

Somit kann ein Discrete-Choice-Modell allgemein als bedingte, multivariate Wahrscheinlichkeitsverteilung definiert werden, die unter der Annahme eines bestimmten Offersets  $S$  jedem enthaltenen Produkt eine Kaufwahrscheinlichkeit zuordnet (vgl. Farias et al. (2013, S. 305)). Bezüglich dieser Kaufwahrscheinlichkeiten können Axiome formuliert werden, die jedes Discrete-Choice-Modell erfüllen muss (vgl. Berbeglia et al. (2018, S. 5)):

$$\bullet \quad P_j(S) \geq 0 \quad \forall j \in S \subseteq \mathcal{J} \quad (3.8)$$

$$\bullet \quad P_j(S) = 0 \quad \forall j \notin S \subseteq \mathcal{J}, j \in \mathcal{J} \quad (3.9)$$

$$\bullet \quad \sum_{j \in S} P_j(S) = 1 \quad \forall S \subseteq \mathcal{J} \quad (3.10)$$

Die sogenannten Random-Utility-Modelle (RUM) werden aus dieser allgemeinen Formulierung abgeleitet, indem jeweils eine bestimmte Dichtefunktion  $f(\boldsymbol{\varepsilon})$  angenommen wird. Diese Annahme hat Auswirkungen darauf, ob für das Integral in Gleichung (3.7) eine Lösung in geschlossener Form angegeben werden kann (vgl. Train (2009, S. 16)). Daraus ergeben sich Unterschiede hinsichtlich der Komplexität des Assortment-Optimierungsproblems.

Gemäß der Definition, dass sich verschiedene RUM in Abhängigkeit von  $f(\boldsymbol{\varepsilon})$  ergeben, sind von den in dieser Arbeit betrachteten Modellen das MNL-Modell und das LC-MNL-Modell der Klasse der RUM zuzurechnen. In der Literatur finden sich jedoch auch weiter gefasste Definitionen. Als RUM im weiteren Sinne werden Modelle bezeichnet, die lediglich kompatibel mit den Grundannahmen der RUT sind. So definieren beispielsweise Berbeglia et al. (2018) jede Wahrscheinlichkeitsverteilung  $P_j(S)$  als RUM, die sich aus einem nutzen- bzw. präferenzbasierten Modell ergibt. Dieser Definition folgend können auch das Rank-based-Modell und das Markov-Chain-Modell als RUM bezeichnet werden. Um zwischen der eng gefassten und der weiter gefassten Definition differenzieren zu können, werden im weiteren Verlauf der Arbeit nur die RUM im engeren Sinne auch als solche bezeichnet. Für Modelle, die nicht aus der Annahme einer konkreten Dichtefunktion  $f(\boldsymbol{\varepsilon})$  der stochastischen Nutzenkomponente einer Nutzenfunktion hervorgehen, wird der Begriff RUT-kompatibles Modell verwendet.

### 3.4.2 Irrelevanz des absoluten Nutzens und der Skalierung

Im Nutzenkalkül des Kunden spielen die absoluten Nutzenwerte  $U_j$  keine Rolle. Erkennbar wird dies durch die Addition einer Konstante  $K \in \mathbb{R}$  zu allen Nutzenparametern und die Tatsache, dass die Kaufentscheidung des Kunden sich dadurch nicht ändert (vgl. Train (2009, S. 19f)):

$$j = \operatorname{argmax}_{k \in S} \{U_k + K\} = \operatorname{argmax}_{k \in S} \{U_k\} \quad (3.11)$$

Auch für die Fluggesellschaft, welche lediglich Kaufwahrscheinlichkeiten anhand der beobachtbaren Nutzenkomponenten  $V_j$  und in Abhängigkeit der Verteilung von  $\boldsymbol{\varepsilon}$  ermitteln kann, sind die absoluten Werte dieser Parameter irrelevant:

$$\begin{aligned}
P_j(S) &= P\left((\varepsilon_k + K) - (\varepsilon_j + K) < (V_j + K) - (V_k + K) \forall k \neq j, k \in S\right) \\
&= P(\varepsilon_k - \varepsilon_j < V_j - V_k \forall k \neq j, k \in S)
\end{aligned} \tag{3.12}$$

Sowohl aus Sicht des Kunden als auch aus Sicht der Airline sind also nur die Nutzendifferenzen zwischen den Alternativen und nicht die absoluten Nutzenwerte von Bedeutung. Dieser Grundsatz muss bei der Spezifikation der RUM, die den Abschnitten 4.2 und 5.2 ausführlich behandelt wird, beachtet werden, da die Differenzen der deterministischen und der stochastischen Nutzenkomponenten nur bezüglich einer Referenz berechnet werden können (vgl. Train (2009, S. 20)).

Eine ähnliche Argumentation führt zu dem zweiten wichtigen Grundsatz, der Irrelevanz der Skalierung. Werden die Nutzenparameter  $U_j$  mit einer positiven Konstante  $K \in \mathbb{R}^+$  multipliziert, bleibt die Entscheidung des Kunden unverändert (vgl. Train (2009, S. 23)):

$$j = \operatorname{argmax}_{k \in S} \{U_k \cdot K\} = \operatorname{argmax}_{k \in S} \{U_k\} \tag{3.13}$$

Gleiches gilt für die Berechnung der Kaufwahrscheinlichkeiten aus Sicht der Fluggesellschaft, wobei die Konstanten erneut gekürzt werden können:

$$\begin{aligned}
P_j(S) &= P(\varepsilon_k K - \varepsilon_j K < V_j K - V_k K \forall k \neq j, k \in S) \\
&= P(\varepsilon_k - \varepsilon_j < V_j - V_k \forall k \neq j, k \in S)
\end{aligned} \tag{3.14}$$

Aus dem Grundsatz der Irrelevanz der Skalierung folgt die Notwendigkeit einer Normalisierung des Nutzens (vgl. Train (2009, S. 24)). Diese erfolgt meist indirekt über die Normalisierung der Varianz der stochastischen Nutzenkomponenten  $\operatorname{Var}(\varepsilon_j) = \sigma^2$ . Durch die Multiplikation mit  $\frac{K}{\sigma^2}$  kann diese für alle Produkte  $j$  auf einen bestimmten Wert  $K$  gesetzt werden. Aufgrund der Beziehung  $\frac{K}{\sigma^2} \operatorname{Var}(\varepsilon_j) = \operatorname{Var}\left(\frac{\sqrt{K}}{\sigma} \varepsilon_j\right)$  entspricht dies einer Änderung der stochastischen Nutzenkomponenten um den Faktor  $\frac{\sqrt{K}}{\sigma}$ . Da (3.3) weiterhin gilt, ist dies mit einer Änderung des Gesamtnutzens und der deterministischen Komponente um den gleichen Faktor verbunden. Die genaue Vorgehensweise bei der Normalisierung hängt von der Verteilung der stochastischen Nutzenkomponente ab und ist demzufolge für jedes RUM unterschiedlich (vgl. Train (2009, S. 24-28)). Die Normalisierung muss daher bei der Modellspezifikation sowie der Interpretation und dem Vergleich der durch verschiedene Modelle geschätzten Parameter beachtet werden (vgl. Train (2009, S. 29)).



## 3.5 Spezifikation der Discrete-Choice-Modelle

Der Begriff der Spezifikation bezeichnet im Zusammenhang mit Discrete-Choice-Modellen die Festlegung der mathematischen Modellstruktur. Diese beinhaltet zwei wesentliche Schritte: Zuerst wird ein Modell ausgewählt, das den funktionalen Zusammenhang der zu schätzenden Parameter bestimmt.<sup>10</sup> Im zweiten Schritt werden dann die einzelnen Parameter definiert, die das Modell beinhalten soll (vgl. Cascetta (2009, S. 521)). In dieser Arbeit wird der Begriff Spezifikation aus Gründen der Klarheit nur für letzteren Schritt gebraucht und für ersteren der Begriff Modellauswahl verwendet. Discrete-Choice-Modelle lassen sich in die Klassen der parametrischen und der nicht-parametrischen Modelle einteilen, die sich hinsichtlich ihrer Spezifikation fundamental unterscheiden und in den folgenden Abschnitten untersucht werden sollen.

### 3.5.1 Parametrische Modelle

Die mathematische Struktur parametrische Modelle wird grundsätzlich durch den Modellierer, d. h., bezogen auf den in dieser Arbeit betrachteten Anwendungsfall, durch die Airline festgelegt (vgl. Berbeglia et al. (2018, S. 3) für die folgenden Überlegungen). Deswegen wird die Struktur vor dem Schätzverfahren, welches auf den Schritt der Spezifikation folgt, als gegeben angenommen. Diese kann somit während des Schätzprozesses auch dann nicht mehr verändert werden, wenn dies zu einer exakteren Beschreibung der Datenbasis führte. Der Spezifikationsschritt ist im Fall der parametrischen Modelle entscheidend für die Modellgüte, da die Airline die Faktoren identifizieren muss, welche die Entscheidung des Kunden maßgeblich beeinflussen, um diese in Form entsprechender Parameter in das Modell aufzunehmen. Die denkbare Strategie, sämtliche in Frage kommenden Faktoren im Modell abzubilden, um sicherzugehen, dass alle entscheidenden Faktoren enthalten sind, ist allerdings nicht zielführend. Auf diese Weise verringert sich zwar der sogenannte Spezifikationsfehler und das reale Kundenwahlverhalten wird theoretisch genauer abgebildet, jedoch erhöhen sich im Gegenzug die Schätzungenauigkeiten, da eine große Zahl von Parametern anhand eines endlichen Datensatzes geschätzt werden müssen. Dieses Problem wird auch als Overfitting bezeichnet (vgl. van Ryzin und Vulcano (2015, S. 282)). Der umgekehrte Fall entsteht, wenn eine zu geringe Anzahl an Parametern gewählt wird. Um eine Parametermenge zu bestimmen, die in Bezug auf Spezifikations- und Schätzfehler

---

<sup>10</sup> Im Fall der RUM ist dies gleichbedeutend mit der Spezifikation der Verteilung der stochastischen Nutzenkomponente.

ausgewogen ist, wird ein hohes Maß an Erfahrung und die fortwährende Überprüfung und Anpassung der Spezifikation benötigt (vgl. Berbeglia et al. (2018, S. 3)). Ein weiterer wichtiger Spezifikationsgrundsatz besteht darin, die Aufnahme linear abhängiger Parameter zu vermeiden, weil diese nicht getrennt geschätzt werden können (vgl. Cascetta (2009, S. 521)). Ein Beispiel hierfür ist die Integration von Parametern für die Gesamtreisezeit, die reine Flugzeit und die Umsteigezeiten in ein parametrisches Modell. Da die Gesamtreisezeit aus der Summe der reinen Flugzeit und der Umsteigezeiten errechnet wird, bestehen lineare Abhängigkeiten zwischen diesen Parametern.

Den Herausforderungen bei der Spezifikation parametrischer Modelle steht jedoch auch ein großer Vorteil gegenüber: Dadurch, dass das Kundenwahlverhalten anhand relevanter Produktfaktoren modelliert wird, können die Ergebnisse der Parameterschätzung interpretiert werden, um daraus abzuleiten wie die Präferenzen der Kunden für bestimmte Produkte zustande kommen (vgl. van Ryzin und Vulcano (2015, S. 282)). Wird ein neues Produkt angeboten, so können die Auswirkungen auf die Kaufwahrscheinlichkeiten unmittelbar aus der bestehenden Parameterschätzung abgeleitet werden, da die grundsätzliche Gewichtung der Faktoren im Entscheidungsprozess des Kunden sich nicht ändert und auf das neue Produkt übertragen werden kann. Gleiches gilt für die Anpassung von Produktattributen wie etwa bei einer Preisänderung. In beiden Fällen ist es folglich nicht notwendig neue Transaktionsdaten zu dem veränderten Produktangebot zu generieren und die Parameter neu zu schätzen (vgl. Berbeglia et al. (2018, S. 3)).

### 3.5.2 Nicht-parametrische Modelle

Im Unterschied zu parametrischen Modellen werden die Parameter nicht-parametrischer Modelle nicht durch die Fluggesellschaft vorgegeben, sondern während der Parameterschätzung durch das Schätzverfahren bestimmt. Obwohl eine grundlegende Struktur durch die Auswahl eines Modells vorgegeben wird, sind Anzahl und Bedeutung der Parameter je nach verwendetem Datensatz variabel (vgl. Berbeglia et al. (2018, S. 3)). Einerseits bedeutet dies, dass sich die Aufgabe der Spezifikation für die Airline vereinfacht und die damit verbundene Gefahr von Spezifikationsfehlern sinkt (vgl. Farias et al. (2013, S. 306)). Sie ist deshalb noch vorhanden, weil die Airline, um das Schätzverfahren zu beschleunigen und Overfitting zu vermeiden, Restriktionen bzgl. der maximalen Parameteranzahl einführen kann (vgl. van Ryzin und Vulcano (2015, S. 282)). Andererseits können jedoch auch kaum Einsichten in das Entscheidungskalkül des Kunden gewonnen werden und es ist nicht möglich die Auswirkungen von Angebotsanpassungen zu beurteilen (vgl. Strauss et al. (2018, S. 379)).

### 3.6 Schätzung der Modellparameter

Der nächste Schritt nach Auswahl und Spezifikation eines Discrete-Choice-Modells ist die Schätzung der Modellparameter. Diese Schätzung erfolgt auf Basis sogenannter Trainingsdaten, die im Airline-RM aus historischen Transaktionen bestehen. Formal wird dieser Trainingsdatensatz durch eine Menge  $\{(j_1, S_1), \dots, (j_D, S_D)\}$  ausgedrückt, wobei  $S_d$  das in Transaktion  $d$  angebotene Offerset und  $j_d \in S_d$  das gewählte Produkt repräsentiert (vgl. Berbeglia et al. (2018, S. 6)). In Tabelle 1 wird der Trainingsdatensatz des Anwendungsbeispiels veranschaulicht, der  $D = 10$  Transaktionen umfasst. Das Ziel der Schätzung besteht darin, die Werte der Parameter, die allgemein durch den Vektor  $\boldsymbol{\beta}$  bezeichnet werden, so zu wählen, dass sie die beobachteten Transaktionen des Trainingsdatensatzes möglichst gut erklären (vgl. Berbeglia et al. (2018, S. 6)). Zu diesem Zweck kann eine Reihe statistischer Verfahren eingesetzt werden. In dieser Arbeit soll ausschließlich die Maximum-Likelihood-Methode vorgestellt werden, da diese zu den am häufigsten eingesetzten Methoden gehört (vgl. Cascetta (2009, S. 523)). Ähnlich wie im Fall des Assortment-Optimierungsproblems, variiert auch die Komplexität des Maximum-Likelihood-Schätzproblems je nach gewähltem Discrete-Choice-Modell. Daher werden im Folgenden Standardverfahren der nichtlinearen Optimierung vorgestellt, die zur Lösung einfacher Schätzprobleme eingesetzt werden. Die darauffolgenden Abschnitte behandeln approximative Verfahren für komplexere Schätzprobleme und Verfahren für nicht-parametrische Modelle.

#### 3.6.1 Maximum-Likelihood-Methode

Im Rahmen der Maximum-Likelihood-Methode wird das Problem der Parameterschätzung als nichtlineares Optimierungsproblem formuliert, welches die anhand des Modells berechnete Wahrscheinlichkeit dafür maximiert, dass die Kunden exakt das in den Trainingsdaten beobachtete Wahlverhalten zeigen (vgl. Berbeglia et al. (2018, S. 7) für die folgenden Überlegungen). Um diese Gesamtwahrscheinlichkeit auszudrücken, muss zunächst die Wahrscheinlichkeit für die Beobachtung der Kaufentscheidungen der einzelnen Transaktionen  $d$  formuliert werden. Diese ist durch die Parameter  $P(j_d|S_d, \boldsymbol{\beta})$  gegeben. Konkret handelt es sich dabei um die anhand des gewählten Modells mit dem Parameterset  $\boldsymbol{\beta}$  berechnete Wahrscheinlichkeit dafür, dass der in Transaktion  $d$  beobachtete Kauf von  $j_d$  stattfindet, wenn der Kunde aus dem Offerset  $S_d$  wählen kann. Die Zielfunktion des Schätzproblems, die auch als Likelihood-Funktion bezeichnet wird, ergibt sich somit aus dem Produkt der

Einzelwahrscheinlichkeiten  $P(j_d|S_d, \boldsymbol{\beta})$ , unter der Annahme, dass diese unabhängig und identisch verteilt sind:

$$L(\boldsymbol{\beta}) = \prod_{d=1}^D P(j_d|S_d, \boldsymbol{\beta}) \quad (3.15)$$

Sie kann als die von einem Discrete-Choice-Modell prognostizierte Gesamtwahrscheinlichkeit für die Beobachtung aller in den Trainingsdaten erfassten Kaufentscheidungen interpretiert werden. Durch die Anpassung der Modellparameter  $\boldsymbol{\beta}$  soll diese Gesamtwahrscheinlichkeit maximiert werden. Die Einzelwahrscheinlichkeiten  $P(j_d|S_d, \boldsymbol{\beta})$  sind in der Likelihood-Funktion multiplikativ verknüpft. Eine Vereinfachung des Maximierungsproblems kann daher erreicht werden, indem das Produkt in eine Summe transformiert wird. Eine solche Transformation ist durch die Anwendung des natürlichen Logarithmus auf die Funktion, gemäß der Regel  $\ln(a \cdot b) = \ln(a) + \ln(b)$ , möglich:

$$LL(\boldsymbol{\beta}) = \ln(L(\boldsymbol{\beta})) = \sum_{d=1}^D \ln(P(j_d|S_d, \boldsymbol{\beta})) \quad (3.16)$$

Aufgrund der Monotonieeigenschaften des natürlichen Logarithmus befindet sich das globale Maximum der Log-Likelihood-Funktion  $LL(\boldsymbol{\beta})$  an der gleichen Stelle wie das der Likelihood-Funktion  $L(\boldsymbol{\beta})$ . Somit ist ein Parameterset  $\boldsymbol{\beta}^*$ , das die Funktion  $L(\boldsymbol{\beta})$  maximiert, stets auch eine optimale Lösung für das Maximierungsproblem bzgl. der Funktion  $LL(\boldsymbol{\beta})$ . Das Problem der Parameterschätzung kann daher wie folgt formuliert werden:

$$\boldsymbol{\beta}^* = \underset{\boldsymbol{\beta}}{\operatorname{argmax}} \sum_{d=1}^D \ln(P(j_d|S_d, \boldsymbol{\beta})) \quad (3.17)$$

Der optimale Zielfunktionswert ist dabei stets nicht-positiv, da für alle Transaktionen  $d$   $0 < P(j_d|S_d, \boldsymbol{\beta}) \leq 1$  und somit  $\ln(P(j_d|S_d, \boldsymbol{\beta})) \leq 0$  gilt (vgl. Train (2009, S. 186)).

Wie effizient dieses nichtlineare Optimierungsproblem gelöst werden kann, hängt einerseits von der Struktur der Log-Likelihood-Funktion und damit von der Struktur des Discrete-Choice-Modells ab, da mit diesem die Kaufwahrscheinlichkeiten  $P(j_d|S_d, \boldsymbol{\beta})$  berechnet werden. Andererseits trägt auch eine große Anzahl von Parametern  $|\boldsymbol{\beta}|$  zur Problemkomplexität bei (vgl. Berbeglia et al. (2018, S. 7)). Diese Komplexitätsunterschiede müssen bei der Wahl eines geeigneten Lösungsverfahrens für das Schätzproblem beachtet werden.

### 3.6.2 Numerische Verfahren

Numerische Verfahren sind für die Lösung des Schätzproblems geeignet, wenn die Log-Likelihood-Funktion eine günstige Struktur aufweist, d. h. nur wenige nicht-konkave Bereiche aufweist oder im Idealfall global konkav ist (vgl. Berbeglia et al. (2018, S. 7)). Allen numerischen Verfahren liegt ein gemeinsames Prinzip zugrunde (vgl. Train (2009, S. 186)). Im Initialisierungsschritt wird ein Startvektor  $\boldsymbol{\beta}^0$  festgelegt. In jeder Iteration  $i = 1, \dots, I$  wird dann, ausgehend von dem aktuellen Parametervektor  $\boldsymbol{\beta}^i$ , ein neues Parameterset  $\boldsymbol{\beta}^{i+1}$  bestimmt, wobei  $LL(\boldsymbol{\beta}^i) < LL(\boldsymbol{\beta}^{i+1})$  gilt.<sup>11</sup> Die Aktualisierung der Parameter erfolgt anhand des Gradienten  $\mathbf{G}_i = \left( \frac{\partial LL(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \right)_{\boldsymbol{\beta}^i}$  und der Hessematrix  $\mathbf{H}_i = \left( \frac{\partial^2 LL(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^2} \right)_{\boldsymbol{\beta}^i}$  an der Stelle  $\boldsymbol{\beta}^i$ . Dem Gradienten, der die Steigung der Log-Likelihood-Funktion angibt, kann dabei die Information entnommen werden in welche Richtung die Parameter angepasst werden müssen, um eine verbessernde Schätzung zu erhalten. Die Hessematrix enthält Informationen zur Krümmung der Funktion, die zur Berechnung des Betrags genutzt werden, um den die Parameter angepasst werden.

In diesem Abschnitt soll nun exemplarisch das Newton-Raphson-Verfahren (NR-Verfahren) vorgestellt werden (vgl. Train (2009, S. 187-190) für die folgenden Überlegungen), dessen Erweiterungen, zu denen bspw. das BFGS-Verfahren gehört, in Standardsoftware zur nicht-linearen Optimierung, wie dem SciPy-Solver, eingesetzt werden (vgl. SciPy (2019)). In den Iterationen des NR-Verfahrens wird das neue Parameterset wie folgt berechnet:

$$\boldsymbol{\beta}^{i+1} = \boldsymbol{\beta}^i + \Delta(-\mathbf{H}_i^{-1})\mathbf{G}_i \quad (3.18)$$

Diese Vorschrift kann hergeleitet werden, indem das Taylor-Polynom zweiten Grades der Log-Likelihood-Funktion für den Punkt  $\boldsymbol{\beta}^i$  gebildet und maximiert wird:

$$LL(\boldsymbol{\beta}^{i+1}) = LL(\boldsymbol{\beta}^i) + (\boldsymbol{\beta}^{i+1} - \boldsymbol{\beta}^i)^T \mathbf{G}_i + \frac{1}{2}(\boldsymbol{\beta}^{i+1} - \boldsymbol{\beta}^i)^T \mathbf{H}_i(\boldsymbol{\beta}^{i+1} - \boldsymbol{\beta}^i) \quad (3.19)$$

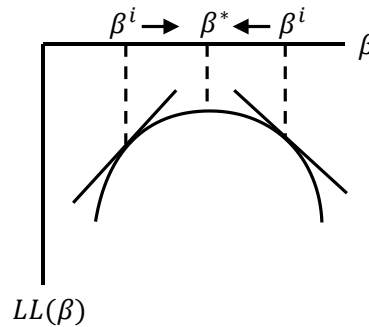
$$\frac{\partial LL(\boldsymbol{\beta}^{i+1})}{\partial \boldsymbol{\beta}^{i+1}} = \mathbf{G}_i + \mathbf{H}_i(\boldsymbol{\beta}^{i+1} - \boldsymbol{\beta}^i) = 0 \quad (3.20)$$

Bei Betrachtung von (3.18) ist unmittelbar ersichtlich, dass  $\boldsymbol{\beta}$  erhöht wird, wenn der Gradient, also die Steigung der Log-Likelihood-Funktion an der Stelle  $\boldsymbol{\beta}^i$ , positiv und die Funktion konkav, d. h.  $-\mathbf{H}_i^{-1}$  positiv definit ist. Ist der Gradient negativ, wird  $\boldsymbol{\beta}$  verringert.

---

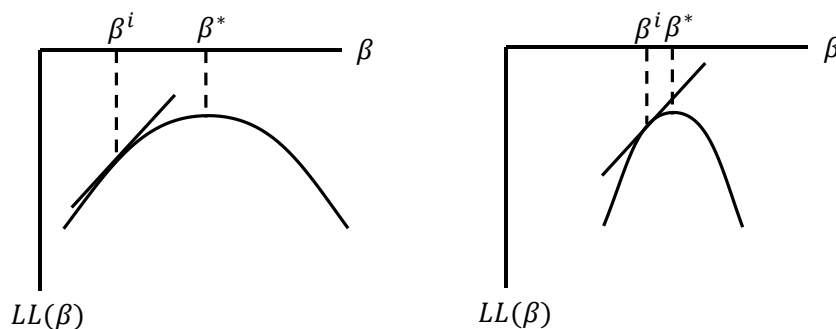
<sup>11</sup> Im Fall des Newton-Raphson-Verfahrens gilt diese Bedingung nur dann, wenn die Log-Likelihood-Funktion global konkav ist (vgl. Train (2009, S. 192)).

Abbildung 1 veranschaulicht diesen Zusammenhang für eine eindimensionale Log-Likelihood-Funktion  $LL(\beta)$ .



**Abbildung 1: Richtung der Anpassung der Parameterschätzung in Abhängigkeit des Gradienten (in Anlehnung an Train (2009, S. 188))**

Da der Gradient mit der invertierten Hessematrix von  $LL(\beta)$  an der Stelle  $\beta^i$  multipliziert wird, führt eine größere Krümmung zu einer geringeren Anpassung des Parametervektors. Dieser Zusammenhang sorgt dafür, dass das Maximum der Log-Likelihood-Funktion  $\beta^*$  nicht „übersprungen“ wird, denn je größer die Krümmung, desto näher liegt der aktuelle Parametervektor  $\beta^i$  an  $\beta^*$ . Das gleiche Ziel ist mit der Einführung des Parameters  $\Delta$  verbunden, der im Verlauf des Verfahrens dynamisch verändert wird und den Betrag der Anpassung von  $\beta$  zusätzlich reguliert, um die Anzahl der benötigten Iterationen zu verringern. Abbildung 2 veranschaulicht dies für  $LL(\beta)$ .



**Abbildung 2: Betrag der Anpassung der Parameterschätzung in Abhängigkeit der Hessematrix (in Anlehnung an Train (2009, S. 188))**

Über den Abbruch des Verfahrens kann anhand verschiedener Konvergenzkriterien entschieden werden (vgl. Train (2009, S. 191f)). Beispielsweise kann das Verfahren abgebrochen werden, wenn alle Elemente des Gradienten kleiner als ein festgelegter Schwellwert sind.

Am Beispiel des Newton-Verfahrens ist erkennbar, weshalb konvexe Bereiche der Log-Likelihood-Funktion deren numerische Optimierung erschweren (vgl. Train (2009, S. 191)).

Liegt der aktuelle Parametervektor  $\beta^i$  in einem konvexen Bereich mit positivem Gradienten, ist  $-H_i^{-1}$  negativ definit, was zu einer Anpassung des Parametervektors entgegen des Gradienten führt. Auch die Problematik lokaler Optima wird bei der Betrachtung des NR-Verfahrens deutlich. Wenn der Startvektor  $\beta^0$  ungünstig gewählt wird, besteht die Gefahr, dass das Verfahren in einem lokalen Maximum konvergiert. Zwar kann das Verfahren mehrfach mit unterschiedlichen Werten für  $\beta^0$  durchlaufen werden, jedoch ist auch diese Vorgehensweise bei einer größeren Anzahl lokaler Maxima nicht effizient (vgl. Train (2009, S. 199)).

Numerische Verfahren stoßen bei der Schätzung komplexer Discrete-Choice-Modelle jedoch nicht nur aufgrund der Struktur der Log-Likelihood-Funktion an ihre Grenzen. Ein weiterer limitierender Faktor ist die Anzahl der Modellparameter (vgl. Train (2009, S. 347)). Je mehr Parameter das Modell enthält desto mehr Dimensionen weist der Lösungsraum auf, sodass die Anzahl der bis zur Konvergenz nötigen Iterationen steigt. Zudem verlängert sich die Dauer einer einzelnen Iteration, weil in jeder Iteration sowohl der Gradient als auch die Inverse der Hesse-Matrix berechnet werden müssen. Besonders die Berechnung der Hesse-Matrix ist mit erheblichem Rechenaufwand verbunden (vgl. Train (2009, S. 192)). Die Konkavität der Log-Likelihood-Funktion und eine relativ geringe Parameteranzahl sind somit die zentralen Voraussetzungen für die Anwendbarkeit numerischer Verfahren, die nur von einfachen Discrete-Choice-Modellen erfüllt werden. In dieser Arbeit wird die numerische Lösung des Schätzproblems in Kap. 4 anhand des Multinomialen-Logit-Modells (MNL-Modells) veranschaulicht.

### 3.6.3 Expectation-Maximization-Verfahren

Das Expectation-Maximization-Verfahren (EM-Verfahren) wurde ursprünglich entwickelt, um die Maximum-Likelihood-Methode auf einen unvollständigen Datensatz anzuwenden, welcher nicht alle Informationen enthält, die zur Formulierung einer vollständigen, konkaven Log-Likelihood-Funktion benötigt werden (vgl. Dempster et al. (1977)). Eine solche Situation stellt etwa der Censored-Demand-Fall im Airline-RM dar, in dem die Fluggesellschaft nicht zwischen einem Nichtkauf und einer Periode, in der kein Kunde angekommen ist, unterscheiden kann. Ohne diese Information ergibt sich selbst für das einfache MNL-Modell eine komplexe, nicht-konkave Log-Likelihood-Funktion (vgl. Talluri und van Ryzin (2004b, S. 26)). Ähnlich wie die numerischen Verfahren maximiert das EM-Verfahren in iterativer Weise die Log-Likelihood-Funktion und bricht ab, sobald ein Konvergenzkriterium erfüllt ist. In jeder Iteration wird dazu die vollständige Log-Likelihood-Funktion

maximiert, welche auch die unbekannten Daten, repräsentiert durch den Vektor  $\mathbf{z}$ , enthält. Anstelle der tatsächlichen Werte dieser Parameter, wird in die Funktion jedoch deren Erwartungswert eingesetzt. Dieser wird vor dem Maximierungsschritt anhand der bekannten Daten sowie der aktuellen Modellparameterschätzung  $\boldsymbol{\beta}^i$  berechnet.

Neben der Anwendung bei einer lückenhaften Datengrundlage ist das EM-Verfahren auch dann einsetzbar, wenn bereits die Struktur der vollständigen Log-Likelihood-Funktion eines Discrete-Choice-Modells komplex ist. Voraussetzung dafür ist, dass sich die Struktur durch Einsetzen der Erwartungswerte der unbekannten Daten  $\mathbf{z}$  vereinfachen lässt. Bei diesen kann es sich in diesem Fall entweder um zusätzliche Informationen zu den beobachteten Transaktionen  $d$  oder eine Teilmenge der eigentlichen Modellparameter handeln (vgl. Train (2009, S. 349)). Die Modellierung dieser Informationen bzw. Modellparameter als unbekannte Daten, ermöglicht dann die bereits beschriebene Maximierung einer konkaven Log-Likelihood-Funktion, die neben den beobachteten Daten die geschätzten Erwartungswerte der unbekannten Daten enthält (vgl. Borman (2004, S. 5)).

Den Ausgangspunkt der Überlegungen zur vereinfachten Herleitung<sup>12</sup> des EM-Verfahrens bildet die durch das Choice-Modell prognostizierte Wahrscheinlichkeit  $P(j_d|S_d, \mathbf{z}, \boldsymbol{\beta})$  für die Beobachtung der Transaktion  $d$ . Diese hängt nun nicht mehr wie bisher nur von dem Offersset  $S_d$  und den Modellparametern  $\boldsymbol{\beta}$  ab. Zusätzlich besteht eine Abhängigkeit von den unbekannten Daten  $\mathbf{z}$ . Für deren Werte kann, wiederum in Abhängigkeit der Modellparameter, eine Dichtefunktion  $f(\mathbf{z}|\boldsymbol{\beta})$  formuliert werden. Die bzgl.  $\mathbf{z}$  unbedingte Wahrscheinlichkeit für die Beobachtung der Transaktion  $d$  ergibt sich nach dem Gesetz der totalen Wahrscheinlichkeit aus dem Integral der bedingten Wahrscheinlichkeit über  $f(\mathbf{z}|\boldsymbol{\beta})$ :

$$P(j_d|S_d, \boldsymbol{\beta}) = \int_{\mathbf{z}} P(j_d|S_d, \mathbf{z}, \boldsymbol{\beta}) f(\mathbf{z}|\boldsymbol{\beta}) d\mathbf{z} \quad (3.21)$$

Demzufolge ergibt sich für die bereits bekannte Log-Likelihood-Funktion mit komplexer Struktur:

$$LL(\boldsymbol{\beta}) = \sum_{d=1}^D \ln(P(j_d|S_d, \boldsymbol{\beta})) = \sum_{d=1}^D \ln \left( \int_{\mathbf{z}} P(j_d|S_d, \mathbf{z}, \boldsymbol{\beta}) f(\mathbf{z}|\boldsymbol{\beta}) d\mathbf{z} \right) \quad (3.22)$$

Diese soll nun mithilfe der unbekannten Daten vereinfacht werden. Obwohl die unbekannten Daten nicht direkt beobachtet werden können, enthält die Datenbasis, genau wie über die

---

<sup>12</sup> Auf eine ausführlichere Herleitung des EM-Verfahrens wird in dieser Arbeit verzichtet. Eine solche findet sich bspw. in Borman (2004). Die folgenden Ausführungen zur vereinfachten Herleitung basieren auf Train (2009, S. 348-350).



Parameter  $\boldsymbol{\beta}$ , Informationen über diese. Präzise formuliert, können anhand der Transaktionen  $(j_d, S_d)$  Rückschlüsse auf die Werte der unbekannten Daten  $\mathbf{z}$  gezogen werden. Um diesen Zusammenhang abzubilden, wird die bedingte Dichtefunktion von  $\mathbf{z}$  bzgl. der aktuellen Parameter und der jeweiligen Transaktionsdaten  $h(\mathbf{z}|j_d, S_d, \boldsymbol{\beta})$  eingeführt, die mithilfe des Satz von Bayes berechnet werden kann:

$$h(\mathbf{z}|j_d, S_d, \boldsymbol{\beta}) = \frac{P(j_d|S_d, \mathbf{z}, \boldsymbol{\beta})f(\mathbf{z}|\boldsymbol{\beta})}{P(j_d|S_d, \boldsymbol{\beta})} \quad (3.23)$$

Für jeden beliebigen Wert von  $\mathbf{z}$  ergibt sich dessen bedingte Wahrscheinlichkeitsdichte aus der Wahrscheinlichkeit für die tatsächlich beobachtete Transaktion multipliziert mit der Wahrscheinlichkeit dafür, dass die unbekannten Daten bei gegebener aktueller Parameterschätzung  $\boldsymbol{\beta}$  den Wert  $\mathbf{z}$  annehmen. Dieses Produkt wird durch das entsprechende Integral über alle  $\mathbf{z}$  normalisiert. Wird der Zähler von (3.23), also die gemeinsame Wahrscheinlichkeit der Beobachtung einer Transaktion  $d$  und eines bestimmten Wertes der unbekannten Daten  $\mathbf{z}$ , logarithmiert, so kann dieser als die Log-Likelihood der kombinierten Ereignisse interpretiert werden. In jeder Iteration des EM-Verfahrens wird nun ausgenutzt, dass eine Aussage darüber getroffen werden kann, wie wahrscheinlich jeder beliebige Wert von  $\mathbf{z}$  angesichts der beobachteten Transaktion  $d$  und der aktuellen Parameterschätzung  $\boldsymbol{\beta}^i$  ist. Diese Aussage ist, wie bereits beschrieben, in  $h(\mathbf{z}|j_d, S_d, \boldsymbol{\beta}^i)$  enthalten. Mithilfe dieser Dichtefunktion kann daher der kumulierte Erwartungswert (über die Dichte aller möglicher Werte der unbekannten Daten  $\mathbf{z}$  und alle Transaktionen  $d$ ) der gemeinsamen Log-Likelihood von Transaktion  $(j_d, S_d)$  und Vektor  $\mathbf{z}$  berechnet werden:

$$ELL(\boldsymbol{\beta}|\boldsymbol{\beta}^i) = \sum_{d=1}^D \int_{\mathbf{z}} h(\mathbf{z}|j_d, S_d, \boldsymbol{\beta}^i) \ln(P(j_d|S_d, \mathbf{z}, \boldsymbol{\beta})f(\mathbf{z}|\boldsymbol{\beta})) d\mathbf{z} \quad (3.24)$$

Die Modellparameter  $\boldsymbol{\beta}$  sind in dieser Funktion sowohl als Inputparameter zur Berechnung der Dichtefunktion  $h(\mathbf{z}|j_d, S_d, \boldsymbol{\beta}^i)$  als auch als Entscheidungsvariable innerhalb des Logarithmus vorhanden. Dementsprechend beginnt jede Iteration eines EM-Verfahrens mit der Berechnung von  $h(\mathbf{z}|j_d, S_d, \boldsymbol{\beta}^i)$  mittels des aktuellen Parametersets  $\boldsymbol{\beta}^i$ . Dieser Schritt wird als Expectation-Schritt (E-Schritt) bezeichnet. Anschließend wird im Maximierungs-Schritt (M-Schritt) die Funktion  $ELL(\boldsymbol{\beta}|\boldsymbol{\beta}^i)$  maximiert und dabei das neue Parameterset bestimmt:

$$\boldsymbol{\beta}^{i+1} = \underset{\boldsymbol{\beta}}{\operatorname{argmax}} ELL(\boldsymbol{\beta}|\boldsymbol{\beta}^i) \quad (3.25)$$

Nach der Initialisierung mit beliebigen Parameterwerten  $\beta^0$  findet ein EM-Verfahren in jeder Iteration einen verbessernden oder zumindest gleich guten Parametervektor, sodass  $LL(\beta^i) \leq LL(\beta^{i+1})$  in allen Iterationen  $i = 1, \dots, I - 1$  gilt (vgl. Borman (2004, S. 8)). Das Verfahren wird abgebrochen, sobald sich die Parameter oder der Zielfunktionswert von einer Iteration zur nächsten nur noch geringfügig verändern.

Das EM-Verfahren ist aufgrund der Flexibilität bei der Definition der unbekannten Daten für die Schätzung verschiedener komplexer Discrete-Choice-Modelle geeignet (vgl. Simsek und Topaloglu (2018, S. 749)). Im Rahmen dieser Arbeit werden EM-Verfahren für das LC-MNL-Modell und das Markov-Chain-Modell in den Kapiteln 5 und 7 vorgestellt.

### 3.6.4 Schätzverfahren für nicht-parametrische Modelle

Wie in Abschnitt 3.5.2 beschrieben, steht das Parameterset vor dem Schätzen eines nicht-parametrischen Modells nicht fest und wird anhand der Datengrundlage erst durch das Schätzverfahren bestimmt. Das um die Aufgabe der Parameterdefinition erweiterte Schätzproblem weist eine zwar eine konkave Log-Likelihood-Funktion auf. Die Problemkomplexität ist dennoch hoch, da jeder potentielle Parameter als Entscheidungsvariable des Problems repräsentiert ist (vgl. van Ryzin und Vulcano (2015, S. 286)). Diese Tatsache schließt eine Lösung mithilfe numerischer Verfahren aus (vgl. Abschnitt 3.6.2). Für Optimierungsprobleme, bei denen die Anzahl der Entscheidungsvariablen exponentiell mit der Größe der Problemistanz wächst, sind grundsätzlich zwei verschiedene Lösungsprinzipien geeignet. Zum einen kann das primale Problem direkt mithilfe eines Spaltengenerierungsverfahrens gelöst werden. Diesen Ansatz, angewandt auf die Log-Likelihood-Funktion, verfolgen van Ryzin und Vulcano (2015), deren Verfahren in Kap. 6 dieser Arbeit vorgestellt wird. Zum anderen kann die Technik des Constraint-Sampling auf das Problem in seiner dualen Form angewandt werden. Ein solches Verfahren wird von Farias et al. (2013) vorgeschlagen, allerdings für ein lineares Optimierungsproblem, welches eine untere Schranke für den erwarteten Erlös eines bestimmten Offersets berechnet. Die Parameter werden in diesem Verfahren nicht anhand des Maximum-Likelihood-Ansatzes geschätzt, weswegen das Verfahren in dieser Arbeit nicht betrachtet wird.

## 4 MNL-Modell

Das zur Klasse der RUM gehörende MNL-Modell wird, abgesehen von trivialen Modellen wie dem Independent-Demand-Modell, als das einfachste Discrete-Choice-Modell bezeichnet (vgl. Cascetta (2009, S. 95)). In den folgenden Abschnitten wird gezeigt auf welchen Annahmen dieses Modell basiert, in welcher Form die Spezifikation erfolgt und welche Vorteile wegen der einfachen Modellstruktur im Hinblick auf die Berechnung der Kaufwahrscheinlichkeiten und die Parameterschätzung entstehen. Im Gegenzug führt die Einfachheit des Modells zu einer ungenauen Modellierung des Kundenwahlverhaltens. Auch die daraus resultierenden Nachteile, welche den Einsatz komplexerer Modelle motivieren, werden betrachtet. Im letzten Abschnitt des Kapitels werden die zuvor auf theoretischer Ebene behandelten Konzepte anhand des Anwendungsbeispiels illustriert.

### 4.1 Annahmen und Modellstruktur

Wie allen RUM liegt auch dem MNL-Modell die Annahme einer bestimmten Verteilung der stochastischen Komponente in (3.3) zugrunde. Im Fall des MNL-Modells wird angenommen, dass alle  $\varepsilon_j$  stochastisch unabhängig sind und gemäß einer identischen Extremwertverteilung des Typs I, die auch als Gumbelverteilung bezeichnet wird, verteilt sind (vgl. Train (2009, S. 34)). Aus der in Abschnitt 3.4.1 hergeleiteten allgemeinen Formel für die Berechnung der Kaufwahrscheinlichkeiten  $P_j(S) = P(V_j - V_k > \varepsilon_k - \varepsilon_j \forall k \neq j, k \in S)$  für alle Produkte  $j$ , der Dichtefunktion  $f(\varepsilon_j) = e^{-\varepsilon_j} e^{-e^{-\varepsilon_j}}$  sowie der Verteilungsfunktion  $F(\varepsilon_j) = e^{-e^{-\varepsilon_j}}$  ergibt sich nach einigen algebraischen Operationen die geschlossene Form (vgl. Train (2009, S. 36)).<sup>13</sup>

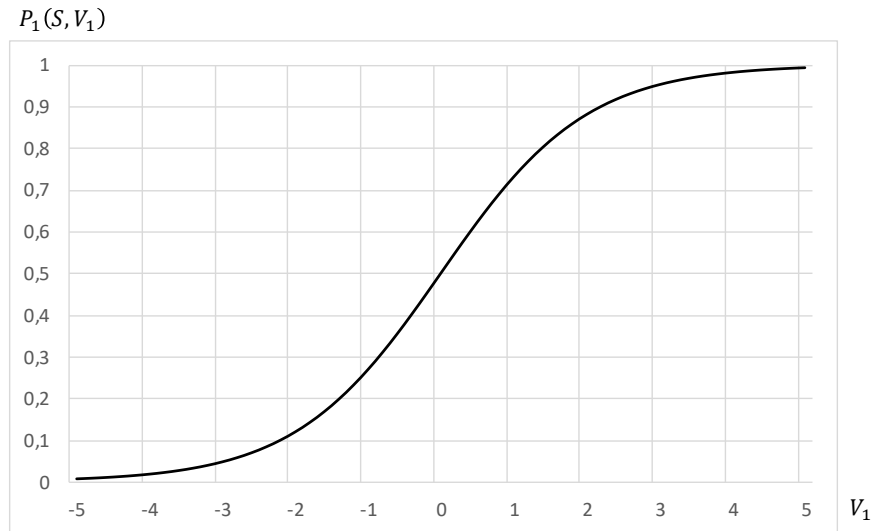
$$P_j(S) = \frac{e^{V_j}}{\sum_{k \in S} e^{V_k}} \quad (4.1)$$

Diese Berechnungsvorschrift erfüllt alle in Abschnitt 3.4.1 aufgelisteten Axiome. Wegen  $\lim_{V_j \rightarrow -\infty} e^{V_j} = 0$  sind die Kaufwahrscheinlichkeiten aller angebotenen Produkte positiv und Axiom (3.8) ist erfüllt. Wie ebenfalls unmittelbar erkennbar ist, stellt der Nenner die Summe des Zählers über alle angebotenen Produkte dar, wodurch die Summe aller Kaufwahrscheinlichkeiten in jedem Fall den Wert 1 ergibt (Axiom (3.10)). Die Auswirkung der

---

<sup>13</sup> Eine ausführliche Herleitung dieses Ausdrucks findet sich in Train (2009, S. 36 und S. 74f).

Nutzenänderung eines Produktes kann anhand eines Offersets  $S = \{0,1\}$  veranschaulicht werden, welches lediglich ein Produkt  $j = 1$  sowie die Nichtkaufalternative  $j = 0$  enthält, wobei  $V_0 = 0$  gesetzt wird (vgl. Train (2009, S. 38)).<sup>14</sup> Der Graph der Funktion  $P_1(S, V_1) = \frac{e^{V_1}}{1+e^{V_1}}$  ist in Abbildung 3 dargestellt.



**Abbildung 3: Verlauf der MNL-Kaufwahrscheinlichkeit  $P_1(S, V_1)$  in Abhängigkeit der deterministischen Nutzenkomponente  $V_1$  (in Anlehnung an Train (2009, S. 38))**

Eine Änderung von  $V_1$ , z. B. als Folge einer Preissenkung, hat die größte Auswirkung auf die Kaufwahrscheinlichkeit des Produktes  $j = 1$ , wenn das Produkt zunächst ähnlich attraktiv war, wie der Nichtkauf. Ist der Kauf dagegen bereits sehr wahrscheinlich, da der Kunde das Produkt gegenüber der Nichtkaufalternative deutlich präferiert, erhöht sich die Kaufwahrscheinlichkeit bei gleicher Nutzenerhöhung um einen geringeren Betrag. Diese Beziehung ist intuitiv plausibel, da für einen indifferenten Kunden bereits eine kleine Nutzenänderung den Ausschlag für oder gegen den Kauf des Produkts geben kann.

## 4.2 Spezifikation

Die Spezifikation des MNL-Modells erfolgt, wie bei allen RUM, über die strukturelle Definition der deterministischen Nutzenkomponente  $V_j(\mathbf{x}_j)$  als Funktion der Ausprägung  $x_{aj}$  verschiedener Attribute  $a = 1, \dots, A$  (vgl. Cascetta (2009, S. 93) für die folgenden Überlegungen). Obwohl grundsätzlich auch nichtlineare Spezifikationen möglich sind, wird  $V_j(\mathbf{x}_j)$  in der Regel als lineare Funktion modelliert, um die Komplexität des Modells gering zu halten. Die Ausprägung jedes Attributs wird im Nutzenkalkül des Kunden mit einem

<sup>14</sup> Die folgenden Überlegungen lassen sich ohne Weiteres auf Offersets mit mehreren Produkten übertragen.

Parameter  $\beta_a$  gewichtet. Aus der Addition der gewichteten Attributwerte ergibt sich so der deterministische Nutzen jedes Produkts:

$$V_j(\mathbf{x}_j) = \boldsymbol{\beta}^T \mathbf{x}_j = \sum_{a=1}^A \beta_a x_{aj} \quad (4.2)$$

Attribute können entweder generisch oder alternativenspezifisch modelliert werden. Generische Attribute sind in der Nutzenfunktion aller Produkte mit dem gleichen Gewicht  $\beta_a$ , aber produktspezifischen Attributwerten  $x_{aj}$  enthalten. Typischerweise werden auf diese Weise sämtliche Charakteristika der Produkte, wie Flugpreis, Reisezeit, Abflugzeit, Abflugtag, etc., modelliert. Wenn ein Attribut für alle Produkte den gleichen Wert aufweist und sich die Gewichte unterscheiden, muss dieses alternativenspezifisch modelliert werden.<sup>15</sup> Eine generelle Präferenz der Kunden für ein bestimmtes Produkt, die nicht durch die spezifizierten Produktattribute erfasst wird, wird ebenfalls auf diese Weise modelliert. Hierfür werden alternativenspezifische Konstanten (ASK)  $\beta_k$  für alle Produkte  $k = 1, \dots, n$  mit Ausnahme der Nichtkaufalternative eingeführt. Diese werden multiplikativ mit dem Wert des alternativenspezifischen Attributs  $x_{kj}$  verknüpft, welches lediglich als Dummy-Variable dient. Daher gilt  $x_{kj} = \begin{cases} 1, & \text{für } j = k \\ 0, & \text{sonst} \end{cases}$ . Aufgrund der Irrelevanz des absoluten Nutzens sind lediglich die Differenzen der ASK von Bedeutung für die Entscheidung des Kunden (vgl. Train (2009, S. 20f) für die folgenden Überlegungen). Aus diesem Grund wird eine der Konstanten auf den Wert 0 gesetzt, meist die der Nichtkaufalternative. Dies hat zur Folge, dass  $\beta_0$  nicht spezifiziert wird und die übrigen Konstanten die Präferenzen der Kunden relativ zur Nichtkaufalternative angeben. Wird auf die Einführung von ASK verzichtet, so kann dem Grundsatz der Irrelevanz des absoluten Nutzens auch dadurch Rechnung getragen werden, dass  $V_0 = 0$  gesetzt wird (vgl. Talluri und van Ryzin (2004b, S. 25)). Wenn nicht sämtliche beobachteten Nutzendifferenzen bereits durch die generischen Attribute erklärt werden können, was einer perfekten Spezifikation von  $V_j(\mathbf{x}_j)$  entspricht, erhöht die Verwendung von ASK die Exaktheit des Modells. Der Grund dafür ist, dass die stochastische Nutzenkomponente  $\varepsilon_j$  aller Alternativen  $j$  nur dann tatsächlich einen Erwartungswert von 0 aufweist. Die Bedingung  $E(\varepsilon_j) = 0$  gehört zu den Grundannahmen der RUT (vgl. Abschnitt 3.4.1) und ist

---

<sup>15</sup> Neben den Produktattributen können auch Attribute des Kunden, wie Größe, Alter oder Geschlecht, den deterministischen Nutzen eines Produkts beeinflussen. Diese werden als soziodemographische Attribute bezeichnet und meist alternativenspezifisch modelliert (vgl. Train (2009, S. 21)). Vereinfachend wird in dieser Arbeit jedoch angenommen, dass die Fluggesellschaft keine Attribute der Kunden beobachten kann und daher in der Spezifikation lediglich Produktattribute betrachtet werden.

ohne ASK in den meisten Fällen nicht erfüllt. Ähnlich verhält es sich mit der Annahme der Unabhängigkeit aller stochastischen Nutzenkomponenten  $\varepsilon_j$ . Dies ist zwar keine Grundannahme der RUT, aber die zentrale Annahme des MNL-Modells, deren Verletzung ebenfalls zu Ungenauigkeiten führt (vgl. Train (2009, S. 35f)). Auch diese Annahme ist bei einer perfekten Spezifikation der deterministischen Nutzenkomponente erfüllt, da sich die stochastischen Komponenten in diesem Fall nicht systematisch unterscheiden und lediglich weißes Rauschen repräsentieren. Spezifikationsfehler als Folge der Verletzung dieser Annahme lassen sich jedoch nicht durch eine Konstante ausgleichen, was die Bedeutung einer sorgfältigen Modellspezifikation für die Modellierungsgüte unterstreicht.

Zuletzt soll untersucht werden, welche Rolle der zweite Spezifikationsgrundsatz, die Irrelevanz der Skalierung, bei der Spezifikation des MNL-Modells spielt (vgl. Train (2009, S. 40f) für die folgenden Überlegungen). Hierfür wird zunächst die unskalierte Nutzenfunktion  $U'_j = \boldsymbol{\beta}'^T \mathbf{x}_j + \varepsilon'_j$  betrachtet. Die stochastischen Nutzenkomponenten werden als gumbelverteilt angenommen und die Varianz dieser Verteilung beträgt per Definition  $\frac{1}{6}\pi^2$ . Aus diesem Grund ist die Varianz der stochastischen Komponenten in der unskalierten Nutzenfunktion  $\text{Var}(\varepsilon'_j) = \sigma^2 \frac{1}{6}\pi^2$  stets ein Vielfaches dieses Werts. Wird nun  $\text{Var}(\varepsilon_j) = \frac{1}{6}\pi^2$  gesetzt, bewirkt dies eine Normalisierung des Modells, d. h. es wird mit dem Wert  $\frac{1}{\sigma}$  skaliert (vgl. Abschnitt 3.4.2). Der skalierte Nutzen  $U_j$  aller Produkte  $j$  ergibt sich folglich gemäß:

$$U_j = \boldsymbol{\beta}^T \mathbf{x}_j + \varepsilon_j = \frac{U'_j}{\sigma} = \frac{\boldsymbol{\beta}'^T}{\sigma} \mathbf{x}_j + \frac{\varepsilon'_j}{\sigma} \quad (4.3)$$

Für das Ergebnis der Parameterschätzung, welche im nächsten Abschnitt beschrieben wird, ergibt sich daraus eine wichtige Implikation: Lediglich der skalierte Parametervektor  $\boldsymbol{\beta}$  kann geschätzt werden. Die Gewichte der Attribute werden daher immer relativ zur Varianz der stochastischen Komponente  $\sigma^2$  geschätzt. Je höher diese Varianz, desto geringer werden die skalierten Gewichte  $\boldsymbol{\beta}$  und umgekehrt. Die tatsächlichen Gewichte  $\boldsymbol{\beta}'$ , die nicht direkt geschätzt werden können, bleiben dabei konstant. Interessant ist in diesem Zusammenhang das Verhalten des Modells für die Fälle  $\sigma^2 \rightarrow 0$  und  $\sigma^2 \rightarrow \infty$ . In ersterem Fall kann die Airline die Kaufentscheidung jedes Kunden mit Sicherheit angeben, da die stochastische Nutzenkomponente verschwindet und  $U_j = V_j$  für alle Produkte  $j$  gilt. Im zweiten Fall besteht für die Airline die größtmögliche Unsicherheit bzgl. der Kaufentscheidung der Kunden, sodass  $P_j(S) = \frac{1}{|S|}$  für alle Produkte  $j$  und Offersets  $S$  gilt (vgl. Strauss et al. (2018, S. 378)). Die Verhältnisse zweier Gewichte  $\beta_a$  und  $\beta_{a'}$  mit  $a = 1, \dots, A$  und  $a \neq a'$ , aus denen

beispielsweise abgeleitet werden kann, wie hoch die Zahlungsbereitschaft eines Kunden für eine Verkürzung der Reisezeit um einen bestimmten Wert ist, sind unabhängig von  $\sigma^2$ .

Das MNL-Modell kann das Kundenwahlverhalten in den meisten Anwendungsfällen, selbst bei nahezu perfekter Spezifikation, nur unzureichend genau abbilden. Zum einen ist das MNL-Modell nicht geeignet, heterogene Präferenzen der Kunden zu modellieren. Zum anderen besitzt das Modell die sogenannte IIA-Eigenschaft (engl.: independence from irrelevant alternatives) und kann daher nur proportionales Substitutionsverhalten abbilden (vgl. Train (2009, S. 42)). Diese beiden zwei Hauptursachen für die fehlende Exaktheit des Modells werden im Folgenden genauer erläutert.

Wie in Abschnitt 2.1 beschrieben, liegt eine heterogene Nachfragesituation vor, wenn sich die Präferenzen und damit die Nutzenfunktionen der Kunden unterscheiden. Im Kontext der RUM bedeutet dies, dass jeder Kunde die Produktattribute mit individuell unterschiedlichen Gewichten  $\beta$  versieht. Beispielsweise ist der Flugpreis für einen Privatreisenden in der Regel sehr viel bedeutender als der Abflugzeitpunkt, wohingegen ein Geschäftsreisender Wert auf einen Abflug am Vormittag und weniger Wert auf einen günstigen Flugpreis legt. Aufgrund dieser unterschiedlichen Gewichtung der Produktattribute ist es sehr wahrscheinlich, dass sich Kunden aus diesen Segmenten, auch wenn sie aus dem gleichen Offerset wählen können, nicht für das gleiche Produkt entscheiden. Je nachdem, wie hoch der Anteil der Segmente, innerhalb derer die Präferenzen als weitgehend homogen angenommen werden können, an der Gesamtnachfrage ist, ändern sich so die realen, aggregierten Kaufwahrscheinlichkeiten. Dies kann das MNL-Modell jedoch nicht abbilden, da es davon ausgeht, dass alle Kunden auf Basis der gleichen Präferenzen, d. h. des gleichen Gewichtsvektors  $\beta$  entscheiden. Letztendlich entstehen so Spezifikationsfehler, denn die mittels des MNL-Modells geschätzten Gewichte entsprechen häufig nicht den tatsächlichen gemittelten Kundenpräferenzen (vgl. Train (2009, S. 44)). Außerdem lassen sich aus der Schätzung des MNL-Modells keine Informationen über die Verteilung der unterschiedlichen Präferenzen gewinnen.

Die IIA-Eigenschaft äußert sich darin, dass das Verhältnis  $P_{j'}(S)/P_j(S)$  der Kaufwahrscheinlichkeiten zweier beliebiger Produkte  $j$  und  $j'$  stets konstant bleibt und damit unabhängig vom restlichen Produktangebot innerhalb eines Offersets  $S$  ist. Dass das MNL-Modell diese Eigenschaft besitzt, kann mit geringem Aufwand wie folgt gezeigt werden (vgl. Train (2009, S. 45)):

$$\frac{P_{j'}(S)}{P_j(S)} = \frac{\frac{e^{V_{j'}}}{\sum_{k \in S} e^{V_k}}}{\frac{e^{V_j}}{\sum_{k \in S} e^{V_k}}} = \frac{e^{V_{j'}}}{e^{V_j}} = e^{V_{j'} - V_j} \quad (4.4)$$

Der Term  $e^{V_{j'} - V_j}$  bleibt konstant, egal wie viele weitere Produkte in  $S$  enthalten sind und welche Attributwerte bzw. welchen Nutzen diese aufweisen. Da es die IIA-Eigenschaft besitzt, unterstellt das MNL-Modell eine proportionale Verlagerung der Nachfrage, falls sich das Offersset  $S$  ändert, d. h. beispielsweise ein zusätzliches Produkt  $k$  angeboten wird (vgl. Train (2009, S. 47)). Das bedeutet, dass sich sämtliche Kaufwahrscheinlichkeiten um den gleichen Faktor  $K$  verändern. Auch die Gültigkeit dieser Implikation der IIA-Eigenschaft lässt sich problemlos zeigen. Ausgangspunkt ist ein Offersset  $S = \{j, j'\}$ . Durch das Anbieten des Produktes  $k$  in einem Offersset  $S' = \{j, j', k\}$  steigt dessen Kaufwahrscheinlichkeit nach (3.8) und (3.9) von  $P_k(S) = 0$  auf  $P_k(S') > 0$ . Gemäß der IIA-Eigenschaft muss für das Verhältnis der Kaufwahrscheinlichkeiten der Produkte  $j$  und  $j'$  gelten:

$$\frac{P_{j'}(S)}{P_j(S)} = \frac{P_{j'}(S')}{P_j(S')} \Leftrightarrow \frac{P_j(S')}{P_j(S)} = \frac{P_{j'}(S')}{P_{j'}(S)} = K \quad (4.5)$$

In vielen Situationen zeigen die Nachfrager nicht das von dem MNL-Modell angenommene proportionale Substitutionsverhalten, was dazu führt, dass sich die tatsächliche Kaufwahrscheinlichkeit der Produkte unterschiedlich stark verändert. Das MNL-Modell überschätzt somit die Kaufwahrscheinlichkeiten einiger Produkte, während die Kaufwahrscheinlichkeiten der übrigen Produkte unterschätzt werden (vgl. Train (2009, S. 47)).

### 4.3 Numerische Maximierung der Log-Likelihood-Funktion

Nach der Spezifikation der deterministischen Nutzenkomponente müssen die Gewichte der Attribute, repräsentiert durch den Vektor  $\boldsymbol{\beta}$ , geschätzt werden. Dazu steht ein Trainingsdatensatz mit den Transaktionen  $d = 1, \dots, D$  der Form  $(j_d, S_d)$  zur Verfügung. Bei der Anwendung des MNL-Modells können die Kaufwahrscheinlichkeiten  $P_j(S)$  in geschlossener Form angegeben werden. Die Formulierung der Log-Likelihood-Funktion des MNL-Modells ist daher sehr einfach möglich, indem der Ausdruck für  $P_j(S)$  in die Log-Likelihood-Funktion in ihrer allgemeinen Form (3.16) eingesetzt wird (vgl. Abschnitt 3.6.1):

$$LL(\boldsymbol{\beta}) = \sum_{d=1}^D \ln(P(j_d | S_d, \boldsymbol{\beta})) = \sum_{d=1}^D \ln(P_{j_d}(S_d)) = \sum_{d=1}^D \ln\left(\frac{e^{\boldsymbol{\beta}^T x_{j_d}}}{\sum_{k \in S_d} e^{\boldsymbol{\beta}^T x_k}}\right) \quad (4.6)$$



Numerische Verfahren, wie das in Abschnitt 3.6.2 vorgestellte NR-Verfahren, eignen sich sehr gut für die Maximierung dieser Funktion (vgl. Berbeglia et al. (2018, S. 10)). Zum einen ist die Anzahl der Parameter unabhängig von Anzahl der Produkte  $|J|$  und steigt lediglich linear mit der Zahl der spezifizierten Produktattribute. Zum anderen ist die Funktion global konkav. Der geschätzte Parametervektor des MNL-Modells  $\beta^*$  ist die optimale Lösung des folgenden Log-Likelihood-Maximierungsproblems:

$$\beta^* = \operatorname{argmax}_{\beta} \sum_{d=1}^D \ln \left( \frac{e^{\beta^T x_{jd}}}{\sum_{i \in S_d} e^{\beta^T x_i}} \right) \quad (4.7)$$

#### 4.4 Exemplarische Spezifikation und Schätzung des MNL-Modells

Die Spezifikation des MNL-Modells, das darauffolgende Schätzverfahren und die IIA-Eigenschaft sollen nun anhand des fortlaufenden Anwendungsbeispiels veranschaulicht werden. Die Fluggesellschaft geht davon aus, dass für den Entscheidungsprozess der Kunden lediglich der Flugpreis und die Abflugzeit relevant sind. Für jede dieser beiden Produkteigenschaften spezifiziert die Airline daher ein Attribut  $a = 1, 2$ , wobei  $x_{1j}$  den Preis und  $x_{2j}$  die Abflugzeit eines Produkts  $j$  angibt. Gewichtet werden diese Attribute mit dem Parameter  $\beta_1$  bzw.  $\beta_2$ . Die deterministische Nutzenkomponente für das Rechenbeispiel ergibt sich demzufolge gemäß  $V_j(x_{1j}, x_{2j}) = \beta_1 x_{1j} + \beta_2 x_{2j}$ . Da die Airline auf zunächst auf ASK verzichtet, wird  $V_0 = 0$  gesetzt. Aus dieser Modellspezifikation ergibt sich das folgende zweidimensionale Schätzproblem mit  $\beta_1, \beta_2 \in \mathbb{R}$ :

$$\operatorname{Max} LL(\beta_1, \beta_2) = \sum_{d=1}^{10} \ln \left( P_{jd}(S_d) \right) = \sum_{d=1}^{10} \ln \left( \frac{e^{\beta_1 x_{1jd} + \beta_2 x_{2jd}}}{\sum_{k \in S_d} e^{\beta_1 x_{1k} + \beta_2 x_{2k}}} \right) \quad (4.8)$$

Dieses Optimierungsproblem kann mithilfe von Standardsoftware, wie dem SciPy-Solver, gelöst werden. Aufgrund der Tatsache, dass das Problem nur zwei Entscheidungsvariablen beinhaltet, ist auch eine graphische Optimierung möglich. Zu diesem Zweck wird ein Diagramm der Log-Likelihood-Funktion in Form einer Heatmap mit Isoquanten erstellt (Abbildung 4).

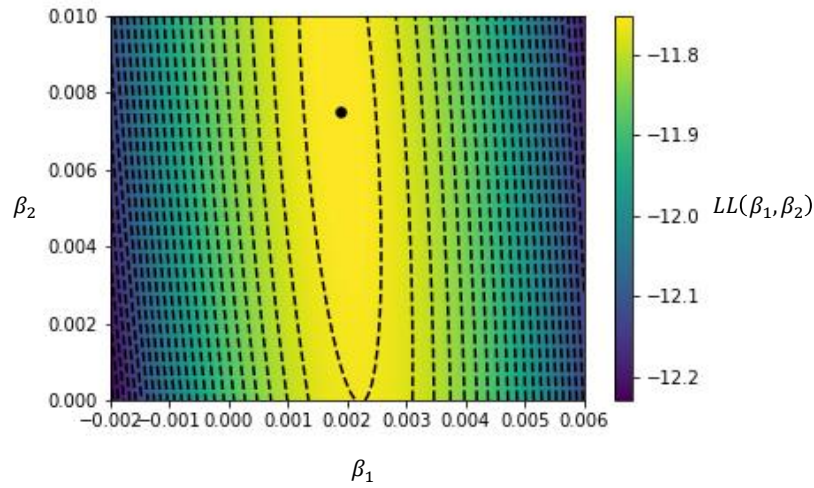


Abbildung 4: Diagramm der MNL-Log-Likelihood-Funktion  $LL(\beta_1, \beta_2)$

Aus der Abbildung lässt sich die optimale Lösung  $\beta^* = (0,002; 0,008)$  direkt ablesen. Die Log-Likelihood, d. h. der Zielfunktionswert, für dieses optimale Parameterset kann durch das Einsetzen von  $\beta^*$  in  $LL(\beta)$  berechnet werden. Hierfür werden zunächst die sogenannten Präferenzgewichte  $v_j = e^{\beta_1 x_{1j} + \beta_2 x_{2j}}$  der Produkte berechnet, wodurch sich die Log-Likelihood-Funktion vereinfacht (vgl. Miranda Bront et al. (2009, S. 772)):

$$LL(\beta_1, \beta_2) = \sum_{d=1}^{10} \ln \left( \frac{e^{\beta_1 x_{1j_d} + \beta_2 x_{2j_d}}}{\sum_{k \in S_d} e^{\beta_1 x_{1k} + \beta_2 x_{2k}}} \right) = \sum_{d=1}^{10} \ln \left( \frac{v_{j_d}}{\sum_{k \in S_d} v_k} \right) \quad (4.9)$$

Für das Produkt  $j = 1$  ergibt sich beispielsweise ein Präferenzgewicht von  $v_1 = e^{0.002 \cdot 200 + 0.008 \cdot 10} = 1,62$ . Die Präferenzgewichte aller Produkte sind in Tabelle 4 aufgelistet.

$j$	0	1	2	3	4
$v_j$	1	1,62	1,32	1,50	1,30

Tabelle 4: Geschätzte Präferenzgewichte  $v_j$  der Produkte  $j = 0, \dots, 4$  ohne ASK

Mithilfe der Präferenzgewichte können anschließend die prognostizierten Kaufwahrscheinlichkeiten  $P_j(S_d)$  für das in jeder Transaktion  $d$  tatsächlich gekaufte Produkt  $j_d$  berechnet werden. In der Transaktion  $d = 1$  wird das Offerset  $S_1 = \{0,1,2,3,4\}$  angeboten und das Produkt  $j_1 = 1$  gekauft. Somit ergibt sich die Kaufwahrscheinlichkeit  $P_1(S_1) = \frac{v_1}{\sum_{k \in S_1} v_k} =$

$\frac{v_1}{v_0+v_1+v_2+v_3+v_4} = \frac{1,62}{1+1,62+1,32+1,50+1,30} = 0,24$ . Tabelle 5 können die Kaufwahrscheinlichkeiten  $P_{j_d}(S_d)$  zu allen Transaktionen  $d$  entnommen werden.

$d$	1	2	3	4	5	6	7	8	9	10
$P_{j_d}(S_d)$	0,24	0,34	0,36	0,24	0,19	0,41	0,35	0,33	0,31	0,38

**Tabelle 5: Kaufwahrscheinlichkeiten  $P_{j_d}(S_d)$  der Produkte  $j_d$  in den Transaktionen  $d = 1, \dots, 10$**

Durch Einsetzen dieser Werte in die Log-Likelihood-Funktion ergibt sich:

$$\begin{aligned} LL(\boldsymbol{\beta}^*) &= \ln(P_1(S_1)) + \ln(P_4(S_2)) + \dots + \ln(P_0(S_{10})) \\ &= \ln(0,24) + \ln(0,34) + \dots + \ln(0,38) = -11,80 \end{aligned} \quad (4.10)$$

Um die Güte dieser Modellspezifikation beurteilen zu können, wird eine erneute Parameterschätzung durchgeführt, in diesem Fall jedoch mit den ASK  $\beta_1, \dots, \beta_4$  sowie weiterhin den Gewichten  $\beta_5$  und  $\beta_6$  für die Attribute Preis und Abflugzeit. Die deterministische Nutzenkomponente ist nun durch  $V_j(x_{1j}, x_{2j}) = \beta_1 x_{1j} + \beta_2 x_{2j} + \dots + \beta_6 x_{6j}$  gegeben, mit den Dummy-Attributwerten  $x_{1j}, \dots, x_{4j}$ . Da dieses Schätzproblem sechs Entscheidungsvariablen beinhaltet, wird die Lösung mithilfe des SciPy-Solvers bestimmt. Das optimale Parameterset ist nun  $\boldsymbol{\beta}_{ASK}^* = (2,898; 3,641; -10,349; 6,698; 0,007; -0,340)$  mit der Log-Likelihood  $LL(\boldsymbol{\beta}_{ASK}^*) = -9,47$ . Die Präferenzgewichte, welche sich daraus ergeben, sind in Tabelle 6 aufgelistet.

$j$	0	1	2	3	4
$v_j$	1	2,45	2,56	0,00	2,71

**Tabelle 6: Geschätzte Präferenzgewichte  $v_j$  der Produkte  $j = 0, \dots, 4$  mit ASK**

Auffällig ist, dass die Werte der ASK  $\beta_1, \dots, \beta_4$  deutlich höher sind als die Gewichte der Attribute Preis und Abflugzeit  $\beta_5$  und  $\beta_6$ . Dies deutet darauf hin, dass die gewählten Attribute nur eine geringe Relevanz für die Entscheidungen der Kunden besitzen und daher ein großer Spezifikationsfehler vorliegt. Zudem wird der Flugpreis positiv gewichtet, d. h. ein höherer Preis geht, laut der Schätzung, mit einem höheren Nutzen für die Kunden einher. Auch dieses unplausible Ergebnis ist auf eine fehlerhafte Spezifikation zurückzuführen, da die Geschäftsreisenden zwar tatsächlich die höherpreisige Buchungsklasse bevorzugen, dies jedoch aufgrund der Flexibilität tun, welche diese Produkte bieten. Die Aufnahme von

Attributen, welche die Flexibilitätsunterschiede modellieren, ist eine Möglichkeit diesen Fehler zu beheben.

Nicht nur die deutliche Verbesserung der Log-Likelihood von  $LL(\boldsymbol{\beta}^*) = -11,80$  auf  $LL(\boldsymbol{\beta}_{ASK}^*) = -9,47$ , auch der Vergleich der Präferenzgewichte zeigt, dass die zweite Schätzung eine deutlich höhere Modellierungsgüte aufweist. Obwohl etwa Produkt  $j = 3$  in keiner Transaktion gewählt wird, ist das Präferenzgewicht  $v_3$  in der ersten Schätzung größer als das von Produkt  $j = 4$ , welches dreimal gekauft wird. Die zweite Schätzung ergibt dagegen den plausiblen Wert  $v_3 \approx 0$ .

Abschließend soll die IIA-Eigenschaft des MNL-Modells numerisch untersucht werden. Dazu wird zunächst das Offersset  $S = \{0,2,4\}$  betrachtet, für dessen Produkte sich, unter Verwendung des Parametersets  $\boldsymbol{\beta}_{ASK}^*$ , die Kaufwahrscheinlichkeiten  $P_0(S) = \frac{1}{1+2,56+2,71} = \frac{100}{627}$ ,  $P_2(S) = \frac{2,56}{1+2,56+2,71} = \frac{241}{627}$  und  $P_4(S) = \frac{2,71}{1+2,56+2,71} = \frac{271}{627}$  ergeben. Aus der Erweiterung des Offersets  $S$  um Produkt  $k = 1$  resultiert das Offersset  $S' = \{0,1,2,4\}$ . Die neuen Kaufwahrscheinlichkeiten betragen  $P_0(S') = \frac{100}{872}$ ,  $P_1(S') = \frac{245}{872}$ ,  $P_2(S') = \frac{256}{872}$  und  $P_4(S') = \frac{271}{872}$ . Durch die Aufnahme des zusätzlichen Produkts verringern sich die Kaufwahrscheinlichkeiten von  $j = 0,1,4$  jeweils um  $\left(1 - \frac{P_j(S')}{P_j(S)}\right) \cdot 100 = \left(1 - \frac{627}{872}\right) \cdot 100 \approx 28$  Prozent.

Tatsächlich ist jedoch eher ein nicht-proportionales Substitutionsverhalten zu erwarten. Falls das Offersset  $S$  angeboten wird, entscheidet sich sehr wahrscheinlich ein großer Teil der preissensitiven Privatreisenden für Produkt  $j = 4$ . Es ist davon auszugehen, dass nur sehr wenige Privatreisende, denen das Offersset  $S'$  angeboten wird, anstelle des günstigsten Produkts  $j = 4$ , das teuerste Produkt  $j = 1$  wählen. Deutlich realistischer ist ein Buy-up der Geschäftsreisenden, die aus Offersset  $S$  Produkt  $j = 2$  gewählt hätten. Die Kaufwahrscheinlichkeit für Produkt  $j = 4$  sollte demnach in der Realität annähernd gleichbleiben, wohingegen für die Kaufwahrscheinlichkeit für Produkt  $j = 2$  eine stärkere Verringerung zu erwarten ist, als durch das MNL-Modell prognostiziert. Diese Überlegungen zeigen, dass die IIA-Eigenschaft zu einer Überschätzung der Kaufwahrscheinlichkeiten ähnlicher Produkte führen kann (vgl. Strauss et al. (2018, S. 378)).

## 5 LC-MNL-Modell

Das LC-MNL-Modell gehört als Generalisierung des MNL-Modells und Spezialfall des Mixed-MNL-Modells ebenfalls zur Klasse der RUM (vgl. Berbeglia et al. (2018, S. 11 und

14)). Es ist strukturell deutlich komplexer als das MNL-Modell und kann daher sowohl heterogene Kundenpräferenzen als auch nicht-proportionales Substitutionsverhalten abbilden. In diesem Kapitel wird das LC-MNL-Modell in Abgrenzung zu den verwandten RUM hergeleitet. Darüber hinaus werden die Besonderheiten der Modellspezifikation beschrieben, die im Vergleich zum MNL-Modell bestehen. Da das Schätzproblem des LC-MNL-Modells sehr komplex ist, liegt der Schwerpunkt des Kapitels auf der Beschreibung und Veranschaulichung eines Lösungsverfahrens für dieses Optimierungsproblem.

## 5.1 Annahmen und Modellstruktur

Das LC-MNL-Modell leitet sich aus der Annahme ab, dass sich die Population der Nachfrager aus  $L$  Segmenten  $l = 1, \dots, L$  zusammensetzt (vgl. Strauss et al. (2018, S. 378) für die folgenden Überlegungen). Ferner wird angenommen, dass, wegen der Homogenität innerhalb dieser Segmente, das Wahlverhalten aller Kunden eines bestimmten Segments  $l$  durch ein MNL-Modell mit identischen Parametern  $\beta_l$  modelliert werden kann. Die Information, zu welchem Segment ein ankommender Kunde gehört, steht der Fluggesellschaft nicht zur Verfügung. Daher kann jeder Kunde lediglich mit einer gewissen Wahrscheinlichkeit einem bestimmten Segment und damit einem Parameterset  $\beta_l$  zugeordnet werden. Der Parametervektor wird somit zu einem Zufallsvektor. Im allgemeineren Mixed-MNL-Modell wird der Parametervektor  $\beta$  als kontinuierlich verteilt angenommen, woraus mit der Dichtefunktion  $f(\beta)$  folgender Ausdruck für die Kaufwahrscheinlichkeiten  $P_j(S)$  resultiert (vgl. Train (2009, S. 135)):

$$P_j(S) = \int \frac{e^{\beta^T x_j}}{\sum_{k \in S} e^{\beta^T x_k}} f(\beta) d\beta \quad (5.1)$$

Ein Mixed-MNL-Modell wird genau dann als LC-MNL-Modell bezeichnet, wenn der Parametervektor diskret verteilt ist, d. h.  $L$  verschiedene Werte  $\beta_l$ , jeweils mit einer Wahrscheinlichkeit  $s_l$ , annehmen kann.  $\beta_l$  ist somit der  $l$ -te Spaltenvektor der Parametermatrix  $\beta$ . In Ausdruck (5.1) lässt sich somit das Integral durch eine Summe über alle Segmente ersetzen und es ergibt sich:

$$P_j(S) = \sum_{l=1}^L s_l \frac{e^{\beta_l^T x_j}}{\sum_{k \in S} e^{\beta_l^T x_k}} \quad (5.2)$$

Diese Kaufwahrscheinlichkeit ist der gewichtete Durchschnitt von  $L$  MNL-Modellen mit unterschiedlichen Parametersets  $\beta_l$ . Die Gewichtung jedes MNL-Modells entspricht dem

Anteil des zugehörigen Segments an der Gesamtpopulation der Nachfrager. Ein MNL-Modell als Spezialfall des LC-MNL-Modells liegt dann vor, wenn  $L = 1$  und folglich  $s_1 = 1$  gilt. Zusätzlich zu den Attributgewichten  $\beta$  enthält das LC-MNL mit den Segmentanteilen  $\mathbf{s}$  ein weiteres Parameterset. Der Oberbegriff der zu spezifizierenden und zu schätzenden Modellparameter umfasst beide Parametertypen.

## 5.2 Spezifikation

Wie bei der Spezifikation des einfachen MNL-Modells, muss zuerst die, für alle Segmente identisch strukturierte, deterministische Nutzenkomponente spezifiziert werden (vgl. Strauss et al. (2018, S. 378)):

$$V_{lj}(\mathbf{x}_j) = \beta_l^T \mathbf{x}_j = \sum_{a=1}^A \beta_{al} x_{aj} \quad (5.3)$$

Die in Abschnitt 4.2 in Bezug auf das MNL-Modell angestellten Überlegungen zur Auswahl der Attribute sind vollständig auf das LC-MNL-Modell übertragbar. Darüber hinaus ist vor der Parameterschätzung die Spezifikation eines weiteren Modellelements erforderlich (vgl. Shen (2009, S. 2916) für die folgenden Überlegungen). Statistische Schätzverfahren können lediglich die Anteile der Segmente an der Gesamtpopulation  $\mathbf{s}$ , nicht aber die Anzahl der Segmente  $L$  bestimmen, weshalb diese im Rahmen der Spezifikation festgelegt werden muss. Hierbei kann sich die Airline zum einen auf Erfahrungswerte stützen. Zum anderen können Modelle mit unterschiedlich vielen Segmenten testweise geschätzt und anhand eines sogenannten Informationskriteriums, z. B. des Akaike-Informationskriteriums (AIC), verglichen werden (vgl. Shen (2009, S. 2916)). Mit dem gewählten Informationskriterium wird jede Spezifikation bewertet, wobei sich eine hohe Log-Likelihood positiv und eine hohe Gesamtparameteranzahl, die gleichbedeutend mit einer hohen Segmentanzahl ist, negativ auf die Bewertung auswirkt. Auf diese Weise wird eine Segmentanzahl ermittelt, welche die größtmögliche Ausgewogenheit zwischen Schätz- und Spezifikationsfehler gewährleistet. Neben der Segmentanzahl kann für jedes Segment auch ein sogenanntes Considerationset  $C_l \subseteq \mathcal{J}$  modelliert werden (vgl. Miranda Bront et al. (2009, S. 772)). Darin sind alle Produkte enthalten, zu deren Kauf die Kunden des Segments  $l$  grundsätzlich bereit sind.<sup>16</sup> Für

---

<sup>16</sup> Modellspezifikationen ohne Considerationsets sind ebenfalls möglich. In diesem Fall ziehen alle Segmente grundsätzlich den Kauf aller Produkte in Erwägung und es gilt  $C_l = \mathcal{J} \forall l = 1, \dots, L$ .

Modellspezifikationen mit Considerationsets muss der Nenner in (5.2) angepasst werden und es ergibt sich für die Kaufwahrscheinlichkeiten  $P_j(S)$  aller Produkte  $j \in S$ :

$$P_j(S) = \sum_{l=1}^L s_l \frac{e^{\beta_l^T x_j}}{\sum_{k \in S \cap C_l} e^{\beta_l^T x_k}} \quad (5.4)$$

Wenn die Considerationsets aller Segmente  $l$  und  $l'$  die Bedingung  $C_l \cap C_{l'} = \emptyset$ , mit  $l \neq l'$ , erfüllen, wird das Modell als LC-MNL-Modell mit disjunkten Considerationsets bezeichnet. Anderenfalls liegt ein LC-MNL mit überlappenden Considerationsets vor (vgl. Miranda Bront et al. (2009, S. 772)). Diese Unterscheidung ist für die Komplexität des Assortment-Optimierungsproblem relevant, denn dieses ist nur für LC-MNL-Modelle mit disjunkten Considerationsets in polynomieller Zeit lösbar (vgl. Miranda Bront et al. (2009, S. 770)).

### 5.3 EM-Verfahren von Train (2009)

Die Kaufwahrscheinlichkeiten  $P_{j_d}(S_d)$  des LC-MNL-Modells lassen sich über eine geschlossene Form als gewichtetes Mittel der segmentspezifischen MNL-Modelle berechnen. Wird diese in (3.16) eingesetzt, ergibt sich folgende, von den Parametersets  $\beta$  und  $s$  abhängige Log-Likelihood-Funktion (vgl. Train (2009, S. 357)):

$$LL(\beta, s) = \sum_{d=1}^D \ln(P_{j_d}(S_d)) = \sum_{d=1}^D \ln\left(\sum_{l=1}^L s_l \frac{e^{\beta_l^T x_j}}{\sum_{k \in S} e^{\beta_l^T x_k}}\right) \quad (5.5)$$

Das Maximum-Likelihood-Schätzproblem des LC-MNL-Modells lautet folglich:

$$\text{Max } LL(\beta, s) \quad (5.6)$$

unter den Nebenbedingungen

$$\sum_{l=1}^L s_l = 1 \quad (5.7)$$

$$s_l \geq 0 \quad \forall l = 1, \dots, L \quad (5.8)$$

Verglichen mit dem MNL-Modell ist dieses Optimierungsproblem in zweierlei Hinsicht komplexer: Die Anzahl der Parameter erhöht sich von  $A$  auf  $A \cdot L + L$  und die Log-Likelihood-Funktion ist nicht mehr global konkav (vgl. Berbeglia et al. (2018, S. 11)). Numerische Verfahren sind unter diesen Bedingungen ineffizient, daher wird in dieser Arbeit das EM-Verfahren von Train (2009) als mögliches Lösungsverfahren vorgestellt (vgl.

Train (2009, S. 357-360) für die folgenden Überlegungen). Die Lösungsstrategie dieses Verfahrens beruht darauf, im Maximierungsschritt jeder Iteration die verbesserten Parametervektoren  $\beta^{i+1}$  und  $s^{i+1}$  über separate Optimierungsprobleme zu bestimmen. Das komplexe Ausgangsproblem wird dadurch in zwei vergleichsweise einfach zu lösende Teilprobleme zerlegt. Für eines dieser Probleme, die Schätzung der Parameter  $s^{i+1}$ , kann die optimale Lösung in geschlossener Form angegeben werden. Die Parameter  $\beta^{i+1}$  können durch die Schätzung eines einfachen MNL-Modells für jedes Segment ermittelt werden, d. h. durch das Lösen global konkaver Optimierungsprobleme, wofür dann effiziente numerische Standardverfahren eingesetzt werden können.

Das zentrale Charakteristikum eines EM-Verfahrens ist die Definition der unbekannten Daten  $\mathbf{z}$ . In dem hier betrachteten Anwendungsfall liegt keine Censored-Demand-Situation vor, denn alle für die Schätzung benötigten Informationen sind in den Transaktionsdatensätzen enthalten. Es wird also lediglich ausgenutzt, dass sich mithilfe eines EM-Verfahrens eine vollständige, aber dennoch komplex strukturierte Log-Likelihood-Funktion wie beschrieben vereinfachen lässt, falls zusätzliche Daten als gegeben erachtet werden. Im Rahmen des EM-Verfahrens von Train (2009) werden die Segmentzugehörigkeiten der in den Transaktionen der Trainingsdaten angekommenen Kunden als unbekannte Daten definiert (vgl. Train (2009, S. 359)). Somit wird davon ausgegangen, dass die Attributgewichte jedes Kunden  $\beta_l$ , die sich einzig aus dem Segment  $l$  ableiten, dem der Kunde angehört, bekannt sind. Nicht beobachtbare Daten dürfen generell nur dann als unbekannte Daten eines EM-Verfahrens definiert werden, wenn die vorhandenen Trainingsdaten Informationen über ihre Werte liefern (vgl. Train (2009, S. 357)). Diese Bedingung wird von den Daten zur Segmentzugehörigkeit erfüllt. Je höher die Präferenz eines Kundensegments, ausgedrückt durch dessen Attributgewichte  $\beta_l$ , für das in Transaktion  $d$  gekaufte Produkt  $j_d$  im Vergleich zu den übrigen Produkten des Offersets  $S_d$  ist, desto wahrscheinlicher ist es, dass der angekommene Kunde zu Segment  $l$  gehört. Die Wahrscheinlichkeit hierfür lässt sich natürlich nur in Abhängigkeit der aktuellen Schätzungen für die Segmentanteile  $s^i$  und die Matrix mit den Parametersets der segmentspezifischen MNL-Modelle  $\beta^i$  angeben. Sie kann folglich genau durch jene bzgl. der Transaktionsdaten und der aktuellen Parameterschätzung bedingte Dichtefunktion der unbekannten Daten  $h(\beta_l | j_d, S_d, \beta^i, s^i)$  beschrieben werden, die im E-Schritt jedes EM-Verfahrens berechnet werden muss (vgl. Train (2009, S. 357)). Wie in Abschnitt 3.6.3 erläutert, werden dazu eine Dichtefunktion  $f(\beta_l | \beta^i, s^i)$  sowie die Kaufwahrscheinlichkeiten  $P(j_d | S_d, \beta_l, \beta^i, s^i)$  und  $P(j_d | S_d, \beta^i, s^i)$  benötigt. Aus den



Grundannahmen des LC-MNL folgt, dass  $f(\boldsymbol{\beta}_l | \boldsymbol{\beta}^i, \mathbf{s}^i)$  diskret ist und nur für  $\boldsymbol{\beta}_l = \boldsymbol{\beta}_l^i$  positive Werte annimmt. Diese Werte können dem Vektor  $\mathbf{s}^i$  entnommen werden, weil die bzgl.  $j_d$  und  $S_d$  unbedingte Wahrscheinlichkeit, dass ein Kunde Segment  $l$  angehört, d. h. anhand von  $\boldsymbol{\beta}_l^i$  entscheidet, stets dem Anteil  $s_l^i$  dieses Segments an der Gesamtpopulation entspricht.  $P(j_d | S_d, \boldsymbol{\beta}_l, \boldsymbol{\beta}^i, \mathbf{s}^i)$  ist in dem hier betrachteten Fall die Kaufwahrscheinlichkeit des in  $d$  tatsächlich gekauften Produkts  $j_d$  unter der Bedingung, dass der Kunde, der diesen Kauf getätigt hat, dem Segment  $l$  angehört und somit anhand der Parameter  $\boldsymbol{\beta}_l$  entschieden hat. Diese Wahrscheinlichkeit wird durch ein einfaches MNL-Modell mit diesem Parametervektor berechnet und im Folgenden mit  $P_{j_d}^{i, MNL}(S_d)$  bezeichnet. Im Nenner des Ausdrucks für  $h(\boldsymbol{\beta}_l | j_d, S_d, \boldsymbol{\beta}^i, \mathbf{s}^i)$  stehen die Kaufwahrscheinlichkeiten  $P(j_d | S_d, \boldsymbol{\beta}^i, \mathbf{s}^i)$ , die durch das LC-MNL-Modell (Gleichung (5.2)) gegeben sind. Für  $h(\boldsymbol{\beta}_l | j_d, S_d, \boldsymbol{\beta}^i, \mathbf{s}^i)$  ergibt sich somit folgende Formulierung:

$$h(\boldsymbol{\beta}_l | j_d, S_d, \boldsymbol{\beta}^i, \mathbf{s}^i) = \frac{s_l^i \cdot P_{j_d}^{i, MNL}(S_d)}{P_{j_d}(S_d)} = s_l^i \frac{e^{\boldsymbol{\beta}_l^{i,T} \mathbf{x}_{j_d}}}{\sum_{k \in S_d} e^{\boldsymbol{\beta}_l^{i,T} \mathbf{x}_k}} \cdot \left( \sum_{l=1}^L s_l^i \frac{e^{\boldsymbol{\beta}_l^{i,T} \mathbf{x}_{j_d}}}{\sum_{k \in S_d} e^{\boldsymbol{\beta}_l^{i,T} \mathbf{x}_k}} \right)^{-1} \quad (5.9)$$

Unter Verwendung der unbekannten Daten lässt sich die erwartete Log-Likelihood wie folgt ausdrücken (vgl. Train (2009, S. 358)):

$$ELL(\boldsymbol{\beta}, \mathbf{s} | \boldsymbol{\beta}^i, \mathbf{s}^i) = \sum_{d=1}^D \sum_{l=1}^L h(\boldsymbol{\beta}_l | j_d, S_d, \boldsymbol{\beta}^i, \mathbf{s}^i) \cdot \ln \left( s_l \frac{e^{\boldsymbol{\beta}_l^T \mathbf{x}_{j_d}}}{\sum_{k \in S_d} e^{\boldsymbol{\beta}_l^T \mathbf{x}_k}} \right) \quad (5.10)$$

Zu beachten ist hier, wie bereits bei der Herleitung der allgemeinen Form dieses Ausdrucks in Abschnitt 3.6.2, dass die aktuelle Parameterschätzung  $\boldsymbol{\beta}^i$  bzw.  $\mathbf{s}^i$  lediglich zur Berechnung der Dichtefunktion der unbekannten Daten  $h(\boldsymbol{\beta}_l | j_d, S_d, \boldsymbol{\beta}^i, \mathbf{s}^i)$  verwendet wird. Zugleich finden sich die beiden Parametertypen als Entscheidungsvariablen der erwarteten Log-Likelihood-Funktion innerhalb des Logarithmus wieder. Diese immer noch komplex strukturierte Zielfunktion soll nun in je eine Zielfunktion für die Schätzung der neuen Attributgewichte  $\boldsymbol{\beta}^{i+1}$  und der neuen Segmentanteile  $\mathbf{s}^{i+1}$  zerlegt werden. Hierfür wird das Logarithmusgesetz  $\ln(a \cdot b) = \ln(a) + \ln(b)$  genutzt. Dessen Anwendung auf den Logarithmus in (5.10) ermöglicht folgende Umformung der erwarteten Log-Likelihood-Funktion (vgl. Train (2009, S. 358f)):

$$ELL(\boldsymbol{\beta}, \mathbf{s} | \boldsymbol{\beta}^i, \mathbf{s}^i) = \sum_{d=1}^D \sum_{l=1}^L h(\boldsymbol{\beta}_l | j_d, S_d, \boldsymbol{\beta}^i, \mathbf{s}^i) \cdot \ln(s_l)$$

$$+ \sum_{d=1}^D \sum_{l=1}^L h(\boldsymbol{\beta}_l | j_d, S_d, \boldsymbol{\beta}^i, \mathbf{s}^i) \cdot \ln \left( \frac{e^{\boldsymbol{\beta}_l^T \mathbf{x}_{j_d}}}{\sum_{k \in S_d} e^{\boldsymbol{\beta}_l^T \mathbf{x}_k}} \right) \quad (5.11)$$

Beide Summanden dieser Funktion sind lediglich von einem der beiden Parametertypen abhängig, da im ersten Summanden ausschließlich die Segmentanteile  $\mathbf{s}$  und im zweiten Summanden ausschließlich die Attributgewichte  $\boldsymbol{\beta}$  als Entscheidungsvariablen enthalten sind. Die separate Maximierung der beiden Summanden in Form der Zielfunktionen

$$ELL(\mathbf{s} | \boldsymbol{\beta}^i, \mathbf{s}^i) = \sum_{d=1}^D \sum_{l=1}^L h(\boldsymbol{\beta}_l | j_d, S_d, \boldsymbol{\beta}^i, \mathbf{s}^i) \cdot \ln(s_l) \quad \text{und} \quad (5.12)$$

$$ELL(\boldsymbol{\beta} | \boldsymbol{\beta}^i, \mathbf{s}^i) = \sum_{d=1}^D \sum_{l=1}^L h(\boldsymbol{\beta}_l | j_d, S_d, \boldsymbol{\beta}^i, \mathbf{s}^i) \cdot \ln \left( \frac{e^{\boldsymbol{\beta}_l^T \mathbf{x}_{j_d}}}{\sum_{k \in S_d} e^{\boldsymbol{\beta}_l^T \mathbf{x}_k}} \right) \quad (5.13)$$

ergibt daher die gleichen optimalen Parameterwerte  $\boldsymbol{\beta}^{i+1}$  bzw.  $\mathbf{s}^{i+1}$  wie die Maximierung von (5.10) (vgl. Train (2009, S. 359f)). Bei der Maximierung von  $ELL(\mathbf{s} | \boldsymbol{\beta}^i, \mathbf{s}^i)$  muss zusätzlich die Nebenbedingung  $\sum_{l=1}^L s_l = 1$  berücksichtigt werden, da  $\mathbf{s}$  ansonsten beliebig hoch gewählt werden könnte und das Optimierungsproblem unbeschränkt wäre. Die Lösung von (5.12) kann dann in geschlossener Form angegeben werden:

$$s_l^{i+1} = \underset{s_l}{\operatorname{argmax}} ELL(s_l | \boldsymbol{\beta}^i, \mathbf{s}^i) = \frac{\sum_{d=1}^D h(\boldsymbol{\beta}_l | j_d, S_d, \boldsymbol{\beta}^i, \mathbf{s}^i)}{\sum_{d=1}^D \sum_{l'=1}^L h(\boldsymbol{\beta}_{l'} | j_d, S_d, \boldsymbol{\beta}^i, \mathbf{s}^i)} \quad \forall l = 1, \dots, L \quad (5.14)$$

Die Maximierung von  $ELL(\boldsymbol{\beta} | \boldsymbol{\beta}^i, \mathbf{s}^i)$  entspricht der Maximierung einer einfachen MNL-Log-Likelihood-Funktion für jedes Segment, wobei die MNL-Log-Likelihood für die Beobachtung jeder einzelnen Transaktion mit  $h(\boldsymbol{\beta}_l | j_d, S_d, \boldsymbol{\beta}^i, \mathbf{s}^i)$  gewichtet wird. Es ergibt sich für jedes Segment  $l$  folgendes Schätzproblem:

$$\begin{aligned} \boldsymbol{\beta}_l^{i+1} &= \underset{\boldsymbol{\beta}_l}{\operatorname{argmax}} ELL(\boldsymbol{\beta}_l | \boldsymbol{\beta}^i, \mathbf{s}^i) \\ &= \sum_{d=1}^D h(\boldsymbol{\beta}_l | j_d, S_d, \boldsymbol{\beta}^i, \mathbf{s}^i) \cdot \ln \left( \frac{e^{\boldsymbol{\beta}_l^T \mathbf{x}_{j_d}}}{\sum_{k \in S_d} e^{\boldsymbol{\beta}_l^T \mathbf{x}_k}} \right) \quad \forall l = 1, \dots, L \end{aligned} \quad (5.15)$$

Je wahrscheinlicher es ist, dass in Transaktion  $d$  ein Kunde des Segments  $l$  angekommen ist, desto höher wird die MNL-Log-Likelihood dieses Segments gewichtet. Ein hoher Zielfunktionswert wird dadurch nur dann erreicht, wenn die segmentspezifischen Attributgewichte eine hohe Kompatibilität mit denjenigen beobachteten Käufen aufweisen, die sehr wahrscheinlich von einem Kunden dieses Segments getätigt worden sind. Auch die Log-Likelihood-Funktion dieses gewichteten MNL-Modells ist global konkav und kann mithilfe

eines Standard-Solvers für nichtlineare Optimierungsprobleme effizient gelöst werden (vgl. Train (2009, S. 361)).

Der gesamte Ablauf des EM-Verfahrens von Train (2009) ist in Form eines Pseudocodes in Abbildung 5 dargestellt (vgl. Train (2009, S. 360)).

```

 $\gamma := \frac{D}{L}, i := 1, konv := False$ 
for  $l = 1, \dots, L$ :
     $s_l^0 := \frac{1}{L}$ 
     $\beta_l^0 := \operatorname{argmax}_{\beta} \sum_{d=\gamma(l-1)+1}^{\gamma l} \ln \left( \frac{e^{\beta^T x_{jd}}}{\sum_{k \in S_d} e^{\beta^T x_k}} \right)$ 
while  $konv = False$ :
    for  $l = 1, \dots, L$ :
        Berechne  $h(\beta_l | j_d, S_d, \beta^i, s^i)$  gemäß (5.9)
    for  $l = 1, \dots, L$ :
        Berechne  $s_l^{i+1}$  gemäß (5.14)
        Berechne  $\beta_l^{i+1}$  gemäß (5.15) mittels eines numerischen Optimierungsverfahrens
    if  $|\beta_l^{i+1} - \beta_l^i| < K \wedge |s_l^{i+1} - s_l^i| < K \forall l = 1, \dots, L$ :
         $konv := True$ 
     $i = i + 1$ 

```

Abbildung 5: Ablauf des EM-Verfahrens von Train (2009)

Im Initialisierungsschritt des Verfahrens werden Startwerte für die Modellparameter festgelegt. Die anfänglichen Anteile  $s_l^0$  der Segmente  $l$  an der Gesamtpopulation werden identisch gewählt. Die Attributgewichte  $\beta_l^0$  der Segmente  $l$  dagegen dürfen nicht identisch sein, da das Verfahren in diesem Fall in der Startlösung „steckenbleiben“ und nicht konvergieren würde. Für die Ermittlung der Startwerte dieses Parametertyps werden die Trainingsdaten deswegen willkürlich in  $L$  Teilmengen mit  $\gamma$  Elementen unterteilt. Diese dienen als Grundlage für die numerische Schätzung  $L$  einfacher MNL-Modelle, aus der für jedes Segment ein individueller Startvektor  $\beta_l^0$  hervorgeht. Der E-Schritt wird durch die erste for-Schleife in jeder Iteration repräsentiert, der M-Schritt durch die zweite for-Schleife. Am Ende jeder

Iteration wird geprüft, ob das Konvergenzkriterium erfüllt ist.<sup>17</sup> Ist dies der Fall, so wird der Steuerungsparameter *konv* auf *True* gesetzt und das Verfahren abgebrochen.

## 5.4 Exemplarische Spezifikation und Schätzung des LC-MNL-Modells

Die deterministische Nutzenkomponente wird in diesem Anwendungsbeispiel genau wie in der exemplarischen Spezifikation des einfachen MNL-Modells in Abschnitt 4.4 modelliert. Die Kunden entscheiden lediglich auf Basis des Flugpreises  $x_{1j}$  und der Abflugzeit  $x_{2j}$  eines Produktes  $j$ . Die Airline modelliert zudem keine ASK. Nach dem Vergleich vergangener Spezifikationen anhand verschiedener Informationskriterien hat sich die Airline für ein LC-MNL-Modell mit  $L = 2$  Kundensegmenten entschieden, die sich beide für das gesamte Produktangebot interessieren. Für die Considerationsets gilt somit  $C_1 = C_2 = \mathcal{J}$ .  $V_{1j}(\mathbf{x}_j) = \boldsymbol{\beta}_1^T \mathbf{x}_j = \beta_{11}x_{1j} + \beta_{21}x_{2j}$  und  $V_{2j}(\mathbf{x}_j) = \boldsymbol{\beta}_2^T \mathbf{x}_j = \beta_{12}x_{1j} + \beta_{22}x_{2j}$  sind die deterministischen Nutzenkomponenten der beiden Segmente  $l = 1$  und  $l = 2$ . Daraus ergeben sich der Vektor der Segmentanteile und die Matrix der Attributgewichte, wobei jede Zeile ein Attribut und jede Spalte ein Segment widerspiegelt:

$$\mathbf{s} = \begin{pmatrix} s_1 \\ s_2 \end{pmatrix} \text{ und } \boldsymbol{\beta} = \begin{pmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{pmatrix} \quad (5.16)$$

Mithilfe des Schätzverfahrens muss folgende Log-Likelihood-Funktion, unter Einhaltung der Restriktionen  $s_1 + s_2 = 1$ ,  $s_1, s_2 \geq 0$  und  $\beta_{11}, \beta_{12}, \beta_{21}, \beta_{22} \in \mathbb{R}$ , maximiert werden:

$$\begin{aligned} LL(\boldsymbol{\beta}, \mathbf{s}) &= \sum_{d=1}^{10} \ln(P_{j_d}(S_d)) \\ &= \sum_{d=1}^{10} \ln \left( s_1 \frac{e^{\beta_{11}x_{1j_d} + \beta_{21}x_{2j_d}}}{\sum_{k \in S_d} e^{\beta_{11}x_{1k} + \beta_{21}x_{2k}}} + s_2 \frac{e^{\beta_{12}x_{1j_d} + \beta_{22}x_{2j_d}}}{\sum_{k \in S_d} e^{\beta_{12}x_{1k} + \beta_{22}x_{2k}}} \right) \end{aligned} \quad (5.17)$$

Bei der Betrachtung dieser beispielhaften Log-Likelihood-Funktion wird deutlich, weshalb sich deren Struktur vereinfacht, wenn das Segment  $l_d$  des für Transaktion  $d$  verantwortlichen Kunden bekannt ist. In diesem Fall muss innerhalb des Logarithmus nicht mehr die Summe über alle Segmente gebildet werden, wodurch die Log-Likelihood-Funktion strukturell derjenigen des einfachen MNL-Modells ähnelt:

---

<sup>17</sup> Train (2009) gibt kein konkretes Konvergenzkriterium an, daher wird in dieser Arbeit die Veränderung der Parameter zur Ermittlung der Konvergenz verwendet. Diese ist dann erreicht, wenn die Veränderung sämtlicher Parameter im Vergleich zur vorherigen Iteration geringer als die hinreichend klein gewählte Konstante  $K$  ist.

$$\begin{aligned}
LL(\boldsymbol{\beta}, \mathbf{s}) &= \sum_{d=1}^{10} \ln \left( s_{l_d} \frac{e^{\beta_{1l_d} x_{1j_d} + \beta_{2l_d} x_{2j_d}}}{\sum_{k \in S_d} e^{\beta_{1l_d} x_{1k} + \beta_{2l_d} x_{2k}}} \right) \\
&= \sum_{d=1}^{10} \ln(s_{l_d}) + \sum_{d=1}^{10} \ln \left( \frac{e^{\beta_{1l_d} x_{1j_d} + \beta_{2l_d} x_{2j_d}}}{\sum_{k \in S_d} e^{\beta_{1l_d} x_{1k} + \beta_{2l_d} x_{2k}}} \right)
\end{aligned} \tag{5.18}$$

Dies ermöglicht die im vorherigen Abschnitt beschriebene Separation der Parametertypen unter Ausnutzung eines Logarithmusgesetzes.

Das EM-Verfahren von Train (2009) beginnt mit der Bestimmung der Startparameter. Da die Fluggesellschaft davon ausgeht, dass die beiden Kundensegmente nicht gleich groß sind, wird  $\mathbf{s}^0 = (0,2; 0,8)^T$  gesetzt.<sup>18</sup> Zur Berechnung der Matrix  $\boldsymbol{\beta}^0$  werden aus dem Trainingsdatensatz  $L = 2$  Teilmengen mit je  $\gamma = \frac{D}{L} = 5$  Transaktionen gebildet. Die Transaktionen  $d = 1, \dots, 5$  dienen als Grundlage für die Schätzung von  $\boldsymbol{\beta}_1^0$  über ein einfaches MNL-Modell. Analog wird  $\boldsymbol{\beta}_2^0$  auf Basis der Transaktionen  $d = 6, \dots, 10$  geschätzt. Die beiden Schätzprobleme beinhalten jeweils nur zwei Parameter, daher kann die Optimierung, wie die Abbildungen 6 und 7 zeigen, auch graphisch erfolgen.

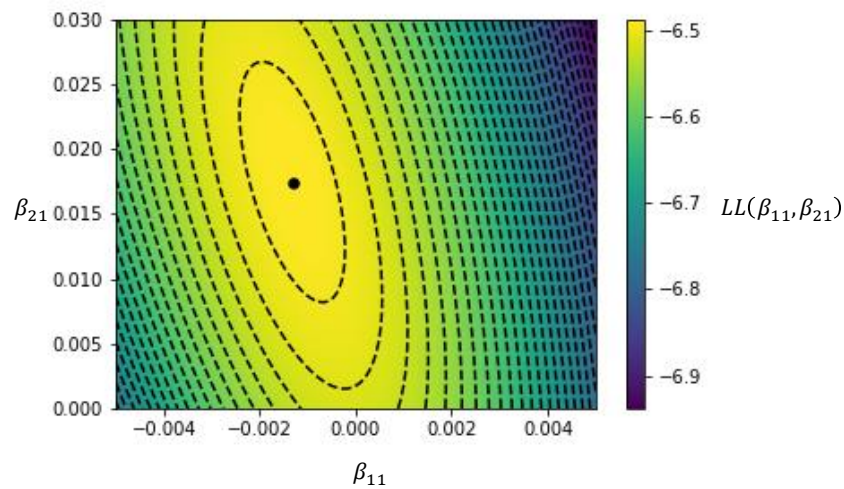


Abbildung 6: Diagramm der Log-Likelihood-Funktion  $LL(\beta_{11}, \beta_{21})$  mit  $d = 1, \dots, 5$

<sup>18</sup> Beliebige andere Startvektoren, wie bspw.  $\mathbf{s} = \begin{pmatrix} 0,5 \\ 0,5 \end{pmatrix}$  sind ebenfalls möglich, solange deren Elemente die Bedingungen  $s_1 + s_2 = 1$  und  $s_1, s_2 \geq 0$  erfüllen.

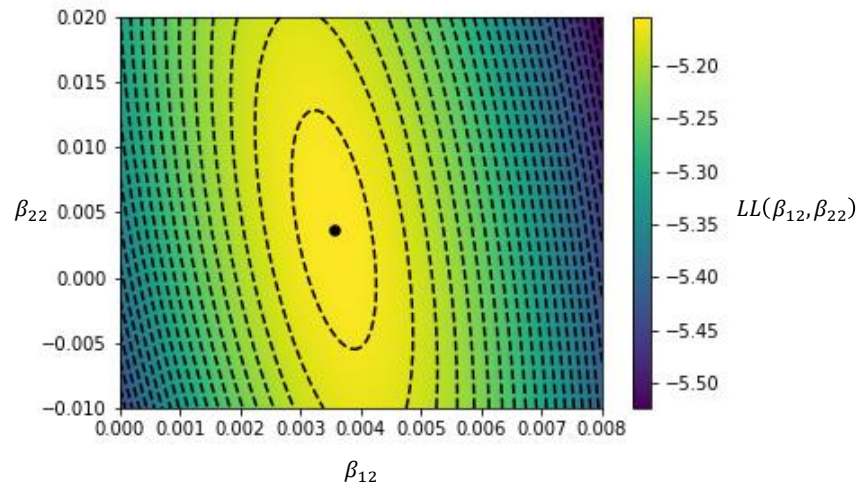


Abbildung 7: Diagramm der Log-Likelihood-Funktion  $LL(\beta_{12}, \beta_{22})$  mit  $d = 6, \dots, 10$

Die anfänglichen Attributgewichte betragen  $\beta_1^0 = (-0,001; 0,017)$  und  $\beta_2^0 = (0,004; 0,004)$ . Bereits in der Initialisierung deutet sich an, dass die Attributgewichte des Segments  $l = 1$  eher zu den Präferenzen der preissensitiven Privatreisenden passen und  $\beta_2^0$  kompatibler mit den Präferenzen der Geschäftsreisenden ist.

Nun werden im E-Schritt der ersten Iteration die Werte für  $h(\beta_l | j_d, S_d, \beta^0, s^0)$  ermittelt. Die dazu notwendige Berechnung wird an der Transaktion  $d = 1$  veranschaulicht und erfolgt analog für alle weiteren Transaktionen. Zunächst werden die Präferenzgewichte  $v_{lj}$  der beiden Segmente  $l = 1$  und  $l = 2$  berechnet (vgl. Tabelle 7).

$j$	0	1	2	3	4
$v_{1j}$	1	0,97	1,07	1,19	1,29
$v_{2j}$	1	2,32	1,55	1,81	1,37

Tabelle 7: Präferenzgewichte  $v_{lj}$  der Segmente  $l = 1, 2$  in der ersten Iteration  $i = 0$

Jetzt können die beiden segmentspezifischen MNL-Kaufwahrscheinlichkeiten  $P_{j_d}^{\beta_l^i, MNL}(S_d)$  für Transaktion  $d = 1$  und Produkt  $j_1 = 1$  berechnet und mit dem Segmentanteil  $s_l^0$  gewichtet werden, um den Zähler von  $h(\beta_l | j_d, S_d, \beta^0, s^0)$  zu ermitteln:

$$s_1^0 \cdot P_1^{\beta_1^0, MNL}(S_1) = 0,2 \cdot \frac{0,97}{1 + 0,97 + 1,07 + 1,19 + 1,29} = 0,0351 \text{ und} \quad (5.19)$$

$$s_2^0 \cdot P_1^{\beta_2^0, MNL}(S_1) = 0,2306 \quad (5.20)$$

Der Nenner ist für beide Segmente identisch und ergibt sich aus der Summe der beiden Zähler. Die Werte von  $h(\beta_l | j_d, S_d, \beta^0, s^0)$  für beide Segmente können nun berechnet werden:

$$h(\beta_1 | 1, S_1, \beta^0, s^0) = \frac{0,0351}{0,0351 + 0,2306} = 0,1321 \text{ und} \quad (5.21)$$

$$h(\beta_2 | 1, S_1, \beta^0, s^0) = 0,8679 \quad (5.22)$$

Die Interpretation dieses Ergebnisses zeigt, dass der für Transaktion  $d = 1$  verantwortliche Kunde mit hoher Wahrscheinlichkeit dem zweiten Segment angehört. Diese Beobachtung ist nicht nur darauf zurückzuführen, dass gemäß der Startparameter 80% der Kunden zu Segment  $l = 2$  gehören, sondern auch auf die hohe Präferenz dieser Kunden für Produkt  $j = 1$ . Dies ist direkt anhand der Präferenzgewichte zu erkennen, da  $v_{21} > v_{2j}$  und  $v_{11} < v_{1j} \forall j = 0,2,3,4$  gilt. Die in analoger Weise berechneten Werte von  $h(\beta_l | j_d, S_d, \beta^0, s^0)$  für die weiteren Transaktionen sind in Tabelle 8 aufgeführt.

$d$	1	2	3	4	5	6	7	8	9	10
$l = 1$	0,13	0,22	0,17	0,29	0,25	0,13	0,19	0,25	0,13	0,30
$l = 2$	0,87	0,78	0,83	0,71	0,75	0,87	0,81	0,75	0,87	0,70

Tabelle 8:  $h(\beta_l | j_d, S_d, \beta^0, s^0)$  für die Segmente  $l = 1, 2$  in der ersten Iteration  $i = 0$

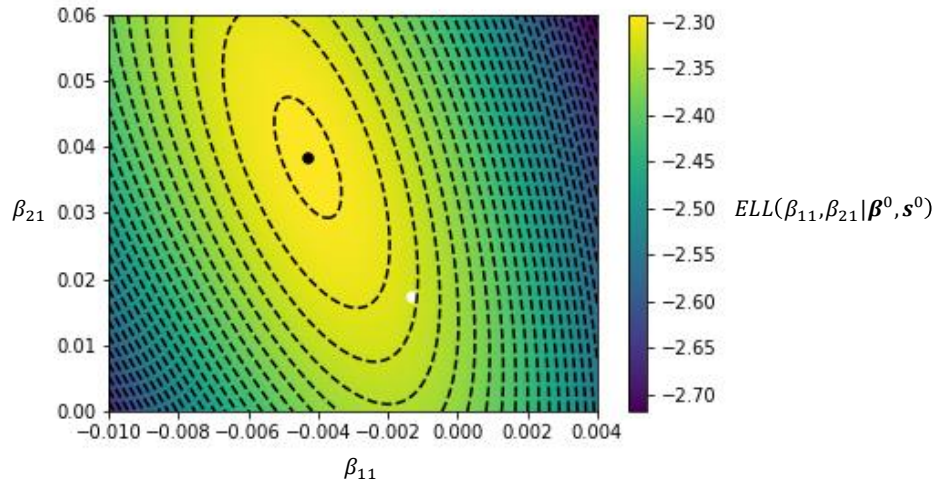
Mithilfe dieser Werte können im nun folgenden M-Schritt die verbessernden Modellparameter ermittelt werden. Zuerst werden  $s_1^1$  bzw.  $s_2^1$  berechnet, indem die Einträge in Zeile  $l = 1$  bzw.  $l = 2$  der Tabelle 8 aufsummiert und durch die Summe aller Einträge der gesamten Tabelle dividiert werden, die stets der Zahl der Transaktionen  $D$  entspricht:

$$s_1^1 = \frac{0,13 + 0,22 + \dots + 0,30}{D} = \frac{2,06}{10} = 0,206 \text{ und } s_2^1 = 0,794 \quad (5.23)$$

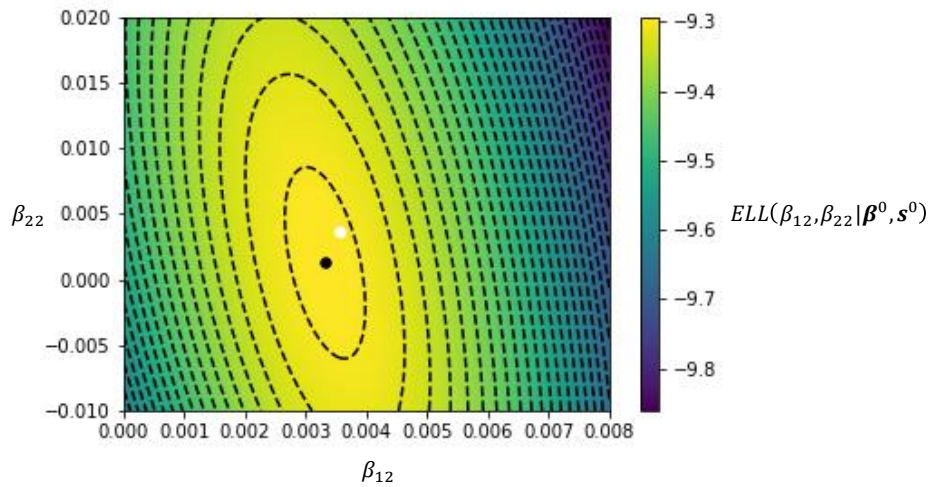
Im Vergleich zu den initialen Segmentanteilen lässt sich die Log-Likelihood folglich dadurch verbessern, dass der Anteil des ersten Segments, dessen Parameter eher die Präferenzen Privatreisender widerspiegeln, erhöht wird. Ein derartiges Ergebnis war zu erwarten, denn die Trainingsdaten wurden unter der Annahme generiert, dass 60% der Nachfrager zu dem Segment der Privatreisenden zu zählen sind.

Nach der Aktualisierung der Segmentanteile werden die verbessernden Attributgewichte der Segmente durch die Maximierung zweier MNL-Log-Likelihood-Funktionen ermittelt. Die zuvor im E-Schritt berechneten Parameter  $h(\beta_l | j_d, S_d, \beta^0, s^0)$  fungieren als Gewichte der Log-Likelihood-Werte der Transaktionen  $d$ . Zur Veranschaulichung wird erneut eine

graphische Schätzung anhand von Abbildung 8 für Segment  $l = 1$  und Abbildung 9 für Segment  $l = 2$  durchgeführt, welche  $\boldsymbol{\beta}_1^1 = (-0,004; 0,039)^T$  und  $\boldsymbol{\beta}_2^1 = (0,003; 0,001)^T$  ergibt.



**Abbildung 8:** Diagramm der erwarteten Log-Likelihood-Funktion  $ELL(\beta_{11}, \beta_{21} | \boldsymbol{\beta}^0, \boldsymbol{s}^0)$  in der ersten Iteration  $i = 0$



**Abbildung 9:** Diagramm der erwarteten Log-Likelihood-Funktion  $ELL(\beta_{12}, \beta_{22} | \boldsymbol{\beta}^0, \boldsymbol{s}^0)$  in der ersten Iteration  $i = 0$

Der Wert der Log-Likelihood-Funktion des LC-MNL-Modells verbessert sich dadurch von  $LL(\boldsymbol{\beta}^0, \boldsymbol{s}^0) = -11,73$  auf  $LL(\boldsymbol{\beta}^1, \boldsymbol{s}^1) = -11,66$ . In den Abbildungen 8 und 9 sind die Attributgewichte  $\boldsymbol{\beta}_1^0$  bzw.  $\boldsymbol{\beta}_2^0$  durch einen weißen Punkt gekennzeichnet, um die Veränderung der Parameterschätzung nach der ersten Iteration nachvollziehen zu können. Während sich die Attributgewichte von Segment  $l = 2$  kaum verändern, entwickeln sich die des ersten Segments in Richtung einer höheren Preissensitivität und einer höheren Präferenz für einen



späteren Abflugzeitpunkt. Auch diese Beobachtung zeigt, dass das Segment  $l = 1$  eher die Gruppe der Privatreisenden repräsentiert. Bestätigt wird dies durch Abbildung 10, in der die Attributgewichte beider Segmente in den Iterationen  $i = 0, \dots, 15$  dargestellt sind.

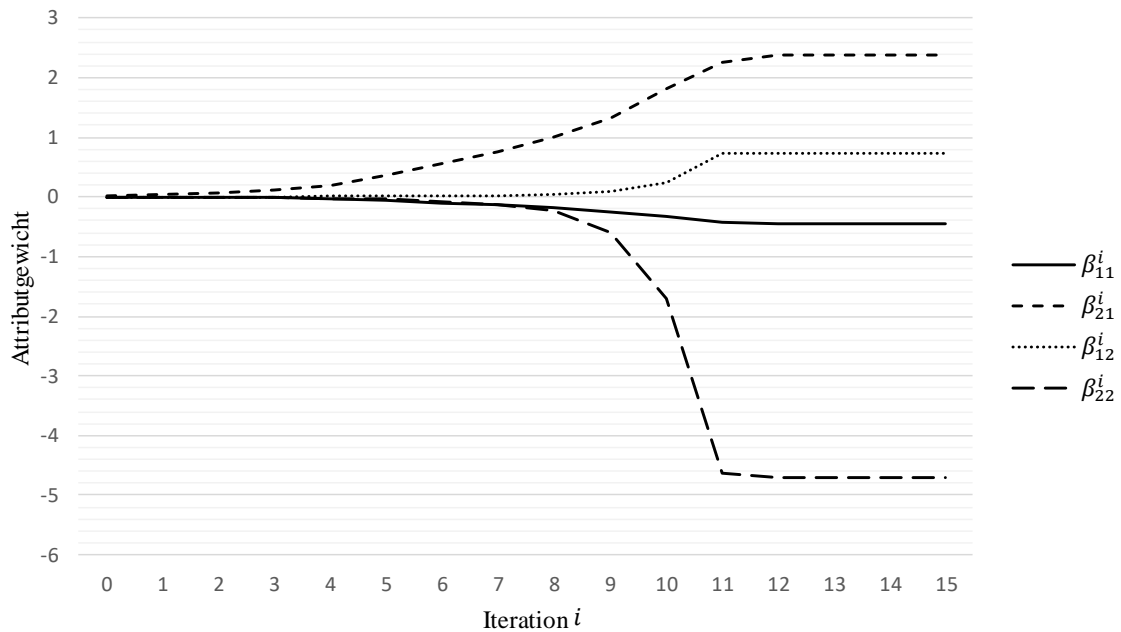


Abbildung 10: Attributgewichte der Segmente  $l = 1, 2$  in den Iterationen  $i = 0, \dots, 15$

Auch der geschätzte Anteil  $s_1^i$  des ersten Segments steigt im Verlauf des Verfahrens erwartungsgemäß weiter an (vgl. Abbildung 11).

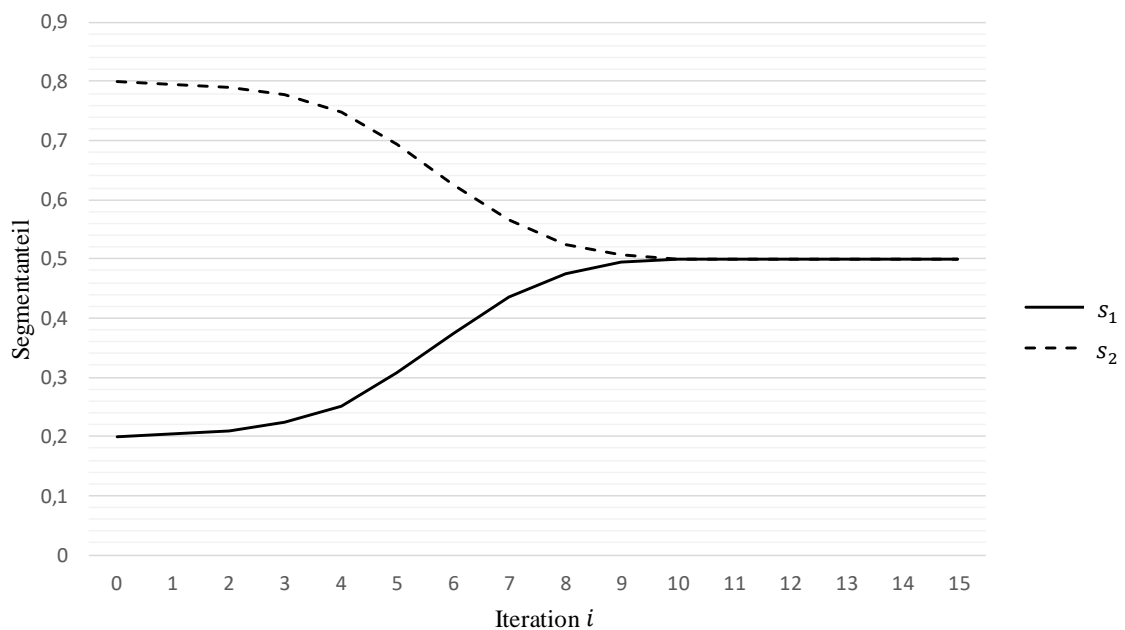
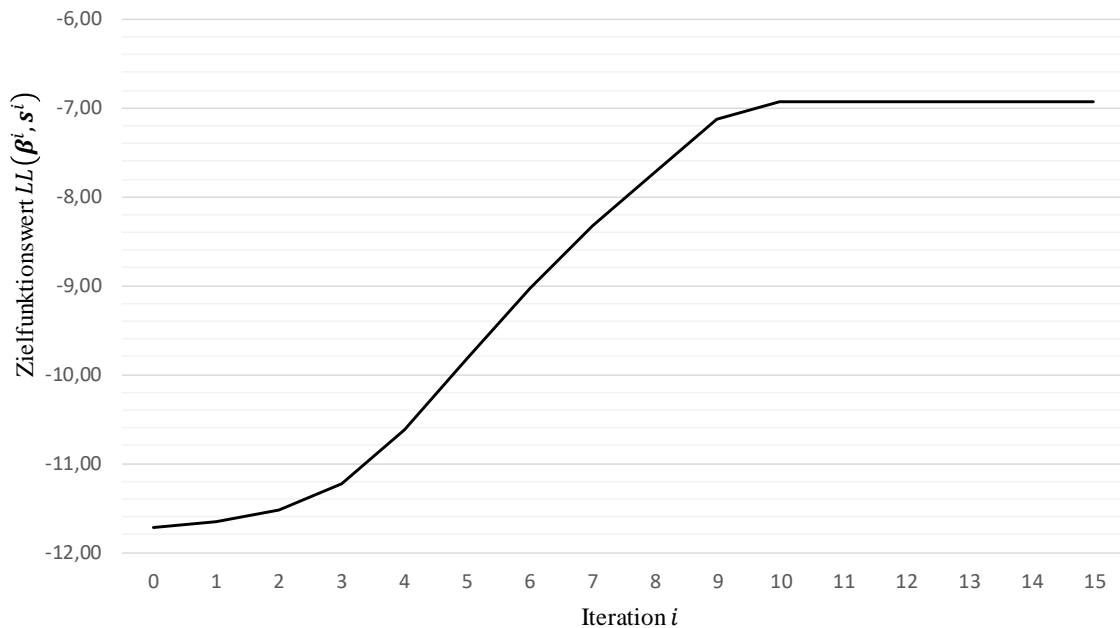


Abbildung 11: Anteile der Segmente  $l = 1, 2$  in den Iterationen  $i = 0, \dots, 15$

Abbildung 12 zeigt die Entwicklung des Zielfunktionswertes im Laufe des Verfahrens.



**Abbildung 12: Zielfunktionswert  $LL(\beta^i, s^i)$  in den Iterationen  $i = 0, \dots, 15$**

Wird das Verfahren nach der 15. Iteration abgebrochen, so lautet die Lösung des Schätzproblems  $\beta_1^{15} = (-0,447; 2,375)^T$ ,  $\beta_2^{15} = (0,739; -4,703)^T$  und  $s^{15} = (0,5; 0,5)^T$ . Die tatsächlichen Präferenzen der Kunden innerhalb der beiden Segmente spiegeln sich in dieser Schätzung deutlich erkennbar wider. Die Schätzung der Segmentanteile stimmt jedoch nicht mit den Daten des tatsächlichen Modells überein, mithilfe dessen die Trainingsdaten generiert wurden. Das LC-MNL-Modell überschätzt hier den Anteil des Segments der Geschäftsreisenden  $l = 2$ , der eigentlich bei 40 Prozent liegen sollte. Der Zielfunktionswert am Ende des Verfahrens beträgt  $LL(\beta^{15}, s^{15}) = -6,93$ . Somit ist die Güte des Modells deutlich höher als die des einfachen MNL-Modells mit gleicher Spezifikation.

## 6 Rank-based-Modell

Das Rank-based-Modell ist ein nicht-parametrisches Choice-Modell, das lediglich RUT-kompatibel ist (vgl. Farias et al. (2013, S. 308)). Grund dafür ist, dass die Kundenpräferenzen nicht durch Attributgewichte codiert werden, aus denen sich der explizite Nutzen jedes Produkts errechnen lässt, sondern in Form einer sogenannten Präferenzliste. Im ersten Abschnitt dieses Kapitels werden die Annahmen des Rank-based-Modells im Vergleich zu denen der

RUM betrachtet. Außerdem wird der Ausdruck für die Kaufwahrscheinlichkeiten hergeleitet. Anschließend werden die Herausforderungen bzgl. der Modellspezifikation erläutert, die für nicht-parametrische Modelle erst während des Schätzverfahrens erfolgt, welches im darauffolgenden Abschnitt vorgestellt wird. Abschließend wird dieses Verfahren anhand der Beispielinstantz veranschaulicht.

## 6.1 Annahmen und Modellstruktur

Das Rank-based-Modell geht davon aus, dass jedem Produkt  $j \in \mathcal{J}$  eine natürliche Zahl  $\sigma(j) = 1, \dots, n + 1$  zugeordnet wird, welche die Position des Produkts in der Präferenzliste  $\sigma$  angibt (vgl. van Ryzin und Vulcano (2015, S. 284f) für die folgenden Überlegungen). Je kleiner diese Zahl, desto höher ist die Präferenz des Kunden für dieses Produkt. Ein Produkt  $j$  wird folglich gegenüber einem alternativen Produkt  $k$  genau dann präferiert, wenn  $\sigma(j) < \sigma(k)$  ist. Aus einem Offerset  $S$  wählt ein Kunde stets das Produkt  $j = \underset{k \in S}{\operatorname{argmin}}\{\sigma(k)\}$ , das in seiner Präferenzliste am weitesten vorne positioniert ist. Relevant für die Modellierung des Kundenwahlverhaltens ist einzig die Präferenzreihenfolge der Produkte, die vor der Nichtkaufalternative  $j = 0$  platziert sind. Da diese zwingend in jedem Offerset enthalten sein muss, werden Produkte, gegenüber denen der Kunde die Nichtkaufoption präferiert, unter keinen Umständen gewählt und die Daten zur Präferenzreihenfolge dieser Produkte sind nicht von Bedeutung. Eine Präferenzliste, welche diese Produkte vernachlässigt, wird als verkürzte Präferenzliste bezeichnet und enthält genau  $\sigma(0) - 1$  Produkte.<sup>19</sup> Anhand ihrer Präferenzlisten können die Nachfrager segmentiert werden, wobei alle Kunden eines Segments  $l$  auf Basis einer identischen Präferenzliste  $\sigma_l$  entscheiden.<sup>20</sup> Genau wie im Fall des LC-MNL-Modells geben die Parameter  $s_l$  an, wie hoch der Anteil des Segments  $l$  an der gesamten Population der Nachfrager ist. Der Wert des Parameters  $s_l$  ist zugleich die Wahrscheinlichkeit dafür, dass ein beliebiger ankommender Kunde nach Präferenzliste  $\sigma_l$  entscheidet und somit zu Segment  $l$  zu zählen ist. Der Parametervektor  $\mathbf{s}$  ist daher letztendlich eine diskrete Wahrscheinlichkeitsverteilung über alle möglichen Präferenzlisten  $\sigma_l$ .

Um die Kaufwahrscheinlichkeiten  $P_j(S)$  für alle Produkte  $j$  gemäß dem Rank-based-Modell formulieren zu können, sind einige Vorüberlegungen nötig (vgl. van Ryzin und Vulcano (2015, S. 285)). Für ein gegebenes Offerset  $S$  und eine bekannte Präferenzliste  $\sigma$  kann, wie

---

<sup>19</sup> Im Folgenden werden sämtliche Präferenzlisten ausschließlich in verkürzter Form dargestellt.

<sup>20</sup> Der Begriff „Segment“ wird hier vereinfachend im Sinne eines Kundentyps verwendet. In der Realität können die Präferenzlisten der einzelnen Kunden sich dynamisch verändern, sodass ein Kunde nicht eindeutig einer Präferenzliste und damit einem Segment zugeordnet werden kann (vgl. Farias et al. (2013, S. 308)).

bereits beschrieben, die Kaufentscheidung des Kunden mit Sicherheit angegeben werden. Der Umkehrschluss ist aus Sicht der Fluggesellschaft jedoch nicht möglich, denn es gibt stets eine Reihe von Präferenzlisten  $\sigma$ , die der Entscheidung für ein bestimmtes Produkt  $j$  aus dem Offerset  $S$  zugrunde liegen können. Konkret sind dies sämtliche Präferenzlisten, in denen Produkt  $j$  vor allen anderen Produkten  $k \in S$ , mit  $k \neq j$ , des Offersets  $S$  positioniert ist, von dem Kunden also unter allen angebotenen Produkten präferiert wird. Diese werden als bzgl. der Kaufentscheidung kompatible Präferenzlisten bezeichnet und in der Menge  $\mathcal{M}_j(S)$  gespeichert. Formal wird diese Menge wie folgt definiert:

$$\mathcal{M}_j(S) = \{l: \sigma_l(j) < \sigma_l(k) \forall k \in S, k \neq j\} \quad (6.1)$$

Die Menge der kompatiblen Präferenzlisten ist essentiell für die Formulierung des Ausdrucks für die Kaufwahrscheinlichkeiten  $P_j(S)$  unter der Annahme eines Rank-based-Modells. Diese Wahrscheinlichkeit ist gleichzusetzen mit dem Anteil der Kunden an der gesamten Population der Nachfrager, die sich angesichts des Offersets  $S$  für Produkt  $j$  entscheiden, d. h. deren Präferenzliste mit dieser Kaufentscheidung kompatibel ist. Um diesen Anteil zu erhalten, müssen folglich die Anteile  $s_l$  aller Segmente  $l \in \mathcal{M}_j(S)$  mit kompatibler Präferenzliste aufsummiert werden und es ergibt sich folgender Ausdruck für die Kaufwahrscheinlichkeit  $P_j(S)$  aller Produkte  $j \in S$ :

$$P_j(S) = \sum_{l \in \mathcal{M}_j(S)} s_l \quad (6.2)$$

Voraussetzung für die Anwendbarkeit dieser Formel ist, dass die Präferenzen jedes Kunden zum Zeitpunkt des Kaufs transitiv sind und die Kunden stets gemäß einer eindeutigen Präferenzliste entscheiden (vgl. Farias et al. (2013, S. 308)). Ein Kunde muss seine Wahl dabei nicht notwendigerweise stets auf Basis der gleichen Präferenzliste treffen. Auch die Modellierung einer Wahrscheinlichkeitsverteilung, die angibt, welche Liste aus einer Menge potentieller Präferenzlisten der Entscheidung eines Kunden tatsächlich zugrunde liegt, ist zulässig.

Die Annahmen, die mit dem Rank-based-Modell einhergehen, sind daher konsistent mit denen der RUM, wenngleich sie allgemeiner sind. Gemäß den RUM entscheiden die Kunden ebenfalls präferenzbasiert. Diese Präferenzen ergeben sich nicht direkt aus einer Präferenzliste, sondern aus der Maximierung einer Nutzenfunktion. Es ist unmittelbar ersichtlich, dass sich aus dem Vergleich von Nutzenwerten  $U_j$  für alle Produkte  $j \in S$  eines Offersets  $S$  eine eindeutige Präferenzliste  $\sigma$  erzeugen lässt (vgl. Sher et al. (2011, S. 3)). Zur Definition der Nutzenfunktion eines RUM sind jedoch weitere Annahmen, bzgl. der stochastischen und

deterministischen Nutzenkomponenten dieser Nutzenwerte  $U_j$ , nötig. Da das Rank-based-Modell keine derartigen Annahmen erfordert, kann es als Generalisierung der RUM betrachtet werden (vgl. Farias et al. (2013, S. 308)). Rank-based-Modelle eignen sich aus diesem Grund besonders für Anwendungsfälle, in denen die Attributwerte  $x_{aj}$  einiger besonders relevanter Attribute  $a$  für alle Produkte  $j$  nicht oder nur schwer messbar sind (vgl. Sher et al. (2011, S. 2)). Dies trifft beispielsweise auf den Reisekomfort zu, den die Kunden mit einer bestimmten Flugverbindung assoziieren.

## 6.2 Spezifikation

Das Rank-based-Modell beinhaltet mit der relativen Häufigkeit  $s_l$  jeder Präferenzliste  $\sigma_l$  nur eine Parameterklasse, die durch den Vektor  $\mathbf{s}$  repräsentiert wird. Wenn die Airline davon ausgeht, dass alle denkbaren Präferenzlisten in der Population der Nachfrager tatsächlich auftreten, besteht ein möglicher Spezifikationsansatz darin, für jede dieser Präferenzlisten  $\sigma_l$  den Parameter  $s_l$  in das Schätzproblem aufzunehmen. Würde alle Präferenzlisten vernachlässigt, für die  $\sigma(0) = 1$  ist, so enthält der Parametervektor  $\mathbf{s}$  dieses Ansatzes  $L = \sum_{j=1}^n \binom{n}{j} j!$  Elemente (vgl. van Ryzin und Vulcano (2015, S. 284)). Neben den typischen Schwierigkeiten wie dem Overfitting und der rechenintensiven Parameterschätzung, die aus einer sehr großen Zahl von Modellparametern resultieren können (vgl. Abschnitt 3.5.1), besteht im Fall des Rank-based-Modells zusätzlich das Problem der Nicht-Identifizierbarkeit. Ein Modell wird dann als identifizierbar bezeichnet, wenn sich aus zwei verschiedenen Parametersets stets unterschiedliche Kaufwahrscheinlichkeiten für jede Kombination aus gekauftem Produkt  $j$  und Offerset  $S$  ergeben. Für das Rank-based-Modells muss daher folgende Bedingung für jede Kombination zweier Parametersets  $\mathbf{s} \neq \mathbf{s}'$  gelten (vgl. Sher et al. (2011, S. 6)):

$$P_j^{\mathbf{s}}(S) \neq P_j^{\mathbf{s}'}(S), \forall j \in \mathcal{J}, S \subseteq \mathcal{J} \quad (6.3)$$

Für Anwendungsfälle mit  $|\mathcal{J}| > 4$  erfüllt das Rank-based-Modell Bedingung (6.3) nicht und ist daher nicht identifizierbar. Vereinfacht ausgedrückt ist dies darauf zurückzuführen, dass die Zahl möglicher Präferenzlisten größer ist als die Zahl verschiedener Offersets  $S$  in den Trainingsdaten, selbst wenn Transaktionsdaten zu allen  $2^n$  Offersets vorliegen (vgl. van Ryzin und Vulcano (2015, S. 282)).<sup>21</sup> Die Eigenschaft der Nicht-Identifizierbarkeit ist deswegen problematisch, weil das Schätzproblem eines nicht identifizierbaren Modells nicht eindeutig

---

<sup>21</sup> Vgl. Sher et al. (2011) für eine ausführlicher Herleitung dieser Aussage.

lösbar ist. (vgl. Lewandowsky und Farrell (2011, S. 206)). Eine Möglichkeit des Umgangs mit einem nicht-identifizierbaren Modell ist die Bestimmung unterer und oberer Schranken für die Elemente des Parametervektors  $\mathbf{s}$ , die sogenannte partielle Identifikation des Modells (vgl. Sher et al. (2011, S. 9f)). Zur Berechnung der Kaufwahrscheinlichkeiten für die Assortment-Optimierung oder andere Anwendungsfelder der Discrete-Choice-Modelle im RM wird allerdings ein eindeutiger Parametervektor  $\mathbf{s}$  benötigt. Um diesen bestimmen zu können, darf nur eine Teilmenge der möglichen Präferenzlisten, idealerweise die der relevantesten Segmente, spezifiziert werden (vgl. van Ryzin und Vulcano (2015, S. 282)). So wird zugleich die Balance zwischen Spezifikationsfehlern und Schätzfehlern hergestellt (vgl. Abschnitt 3.5).

In Abschnitt 3.5.2 wurde bereits erläutert, dass die Spezifikation und die Schätzung nicht-parametrischer Modelle, zu denen das Rank-based-Modell gehört, integriert erfolgen müssen. Im nächsten Abschnitt wird exemplarisch das Spaltengenerierungsverfahren von van Ryzin und Vulcano (2015) vorgestellt. In diesem Verfahren wird das Spezifikationsproblem dadurch gelöst, dass in jeder Iteration, ausgehend von wenigen Startparametern, ein weiterer Parameter als Entscheidungsvariable in das Modell aufgenommen wird, solange dies zu einem signifikant verbesserten Log-Likelihood-Wert führt.

### 6.3 Spaltengenerierungsverfahren von van Ryzin und Vulcano (2015)

Die Zielfunktion des Schätzproblems ergibt sich für das Rank-based-Modell unter der Annahme einer Ankunftsrate von  $\lambda = 1$  durch Einsetzen von (6.2) in die allgemeine Form der Log-Likelihood-Funktion (3.16) (vgl. van Ryzin und Vulcano (2015, S. 286)):<sup>22</sup>

$$LL(\mathbf{s}) = \sum_{d=1}^D \ln \left( \sum_{l \in \mathcal{M}_{j_d}(s_d)} s_l \right) \quad (6.4)$$

Mit einer Nebenbedingung, die sicherstellt, dass sich die Anteile  $s_l$  der Segmente  $l = 1, \dots, L$  auf den Wert 1 summieren, ergibt sich ein global konkaves Maximum-Likelihood-Schätzproblem:

---

<sup>22</sup> Falls die Ankunftsrate der Kunden  $\lambda$  nicht bekannt ist, kann diese ebenfalls durch die Maximum-Likelihood-Methode geschätzt werden, indem  $\lambda$  in die Log-Likelihood-Funktion aufgenommen wird. Aus  $LL(\mathbf{s}, \lambda)$  lassen sich zwei separate Schätzprobleme für  $\mathbf{s}$  und  $\lambda$  formulieren, zweites ist in geschlossener Form lösbar (vgl. van Ryzin und Vulcano (2015, S. 286)) und wird in dieser Arbeit vernachlässigt, da  $\lambda = 1$  angenommen wird.

$$\text{Max } LL(\mathbf{s}) \quad (6.5)$$

unter den Nebenbedingungen

$$\sum_{l=1}^L s_l = 1 \quad (6.6)$$

$$s_l \geq 0 \quad \forall l = 1, \dots, L \quad (6.7)$$

Durch die Einführung der Hilfsvariablen  $y_d = \sum_{l \in \mathcal{M}_{j_d}(S_d)} s_l$  lässt sich eine äquivalente Zielfunktion ohne die Mengen  $\mathcal{M}_{j_d}(S_d)$  formulieren. Jeder Hilfsvariable  $y_d$  soll der kumulierte Anteil aller Segmente  $l$  zugewiesen werden, die eine bzgl. der Transaktion  $(j_d, S_d)$  kompatible Präferenzliste  $\sigma_l$  aufweisen. Zu diesem Zweck werden die binären Parameter  $\alpha_{dl} \in \{0; 1\}$  eingeführt, die genau dann den Wert 1 annehmen, wenn  $l \in \mathcal{M}_{j_d}(S_d)$  gilt. Eine zusätzliche Nebenbedingung pro Transaktion  $d = 1, \dots, D$  legt mithilfe der Parameter  $\alpha_{dl}$  die Werte der Hilfsvariablen  $y_d$  fest und es ergibt sich folgende Reformulierung des Schätzproblems (vgl. van Ryzin und Vulcano (2015, S. 286)):

$$\text{Max } LL(\mathbf{s}, \mathbf{y}) = \sum_{d=1}^D \ln(y_d) \quad (6.8)$$

unter den Nebenbedingungen

$$\sum_{l=1}^L s_l = 1 \quad (6.9)$$

$$\sum_{l=1}^L \alpha_{dl} s_l = y_d \quad \forall d = 1, \dots, D \quad (6.10)$$

$$s_l \geq 0 \quad \forall l = 1, \dots, L \quad (6.11)$$

Die Spaltengenerierung ist eine iterative Strategie, die typischerweise im Bereich der linearen Optimierung verwendet wird, jedoch auch für die Lösung nichtlinearer Optimierungsprobleme, wie dem vorliegenden Maximum-Likelihood-Schätzproblem, geeignet ist (vgl. Larsson et al. (2015, S. 166)). Bevor gezeigt wird wie dieses Problem mithilfe eines Spaltengenerierungsverfahrens gelöst werden kann, sollen zwei weitere transformierte Formen des Schätzproblems eingeführt werden, die für das Verfahren von Bedeutung sind:

das sogenannte Restricted-Master-Problem (RMP) und das duale Problem.

Um das RMP des Schätzproblems zu erhalten, wird eine Teilmenge  $\sigma = \{\sigma_1, \dots, \sigma_{\tilde{L}}\}$  mit  $\tilde{L} \ll L$  aus der Menge aller Präferenzlisten ausgewählt. Davon ausgehend wird der Variablenvektor  $\mathbf{s}$  so definiert, dass er lediglich Entscheidungsvariablen  $s_l$  mit  $l = 1, \dots, \tilde{L}$  enthält, die mit einer Präferenzliste  $\sigma_l \in \sigma$  korrespondieren. Daraus ergibt sich das RMP zu Problem (6.5) - (6.7) (vgl. van Ryzin und Vulcano (2015, S. 288)):

$$\text{Max } LL^\sigma(\mathbf{s}) \quad (6.12)$$

unter den Nebenbedingungen

$$\sum_{l=1}^{\tilde{L}} s_l = 1 \quad (6.13)$$

$$s_l \geq 0 \quad \forall l = 1, \dots, \tilde{L} \quad (6.14)$$

Auch die Formulierung (6.8) - (6.11) lässt sich analog in ein RMP überführen. Generell ist ein RMP mit deutlich geringerem Rechenaufwand lösbar als das ursprüngliche Problem mit der vollständigen Variablenmenge (vgl. Larsson et al. (2015, S. 173)). Das RMP des hier betrachteten Schätzproblems kann somit entweder mithilfe eines Standard-Solvers oder eines EM-Verfahrens gelöst werden (vgl. van Ryzin und Vulcano (2015, S. 288)). Wichtig ist außerdem, dass aus einer optimalen Lösung  $\bar{\mathbf{s}} = (\bar{s}_1, \dots, \bar{s}_{\tilde{L}})^T$  des RMP leicht eine zulässige, aber nicht notwendigerweise optimale Lösung  $\mathbf{s} = (s_1, \dots, s_L)^T$  des ursprünglichen Problems konstruiert werden kann. Hierfür wird  $s_1 = \bar{s}_1, \dots, s_{\tilde{L}} = \bar{s}_{\tilde{L}}$  und  $s_{\tilde{L}+1}, \dots, s_L = 0$  gesetzt.

Das duale Schätzproblem spielt für das Spaltengenerierungsverfahren ebenfalls eine wichtige Rolle, denn anhand der Werte der Dualvariablen lässt sich feststellen, ob eine zulässige Lösung  $(\mathbf{s}, \mathbf{y})$  des primalen Problems das Optimalitätskriterium erfüllt (vgl. van Ryzin und Vulcano (2015, S. 287)). Außerdem liefern diese Werte, ähnlich wie der Gradient im Fall der numerischen Lösungsverfahren, Informationen darüber, wie eine suboptimale Lösung angepasst werden muss, um zu einer verbessernden bzw. letztendlich optimalen Lösung zu gelangen (vgl. van Ryzin und Vulcano (2015, S. 283)). Das duale Problem zu (6.8) - (6.11) kann über die Formulierung der Lagrange-Funktion hergeleitet werden (vgl. van Ryzin und Vulcano (2015, S. 288f) für die folgenden Überlegungen):

$$\phi(\mathbf{s}, \mathbf{y}, \pi, \boldsymbol{\mu}) = \sum_{d=1}^D \ln(y_d) + \pi \left( \sum_{l=1}^{\tilde{L}} s_l - 1 \right) + \sum_{d=1}^D \mu_d \left( \sum_{l=1}^{\tilde{L}} \alpha_{dl} s_l - y_d \right) \quad (6.15)$$



Die Nebenbedingungen (6.10) werden hierfür jeweils mit einem Lagrange-Multiplikator  $\mu_d$  versehen und in die Zielfunktion aufgenommen. Gleiches gilt für Nebenbedingung (6.9), der der Lagrange-Multiplikator  $\pi$  zugeordnet wird.

Da die Lagrange-Multiplikatoren zugleich die Entscheidungsvariablen des dualen Problems sind, ergibt sich dieses mit  $\pi \in \mathbb{R}$ ,  $\boldsymbol{\mu} \in \mathbb{R}^T$  und  $\mathbf{s}, \mathbf{y} \geq \mathbf{0}$ :

$$\text{Min } \Theta(\pi, \boldsymbol{\mu}) = \sup_{\mathbf{s}, \mathbf{y}} \left\{ \sum_{d=1}^D \ln(y_d) + \pi \left( \sum_{l=1}^L s_l - 1 \right) + \sum_{d=1}^D \mu_d \left( \sum_{l=1}^L \alpha_{dl} s_l - y_d \right) \right\} \quad (6.16)$$

Gemäß der starken Dualität lässt sich folgendes Optimalitätskriterium formulieren: Die Lösungen  $(\mathbf{s}^*, \mathbf{y}^*)$  bzw.  $(\pi^*, \boldsymbol{\mu}^*)$  sind genau dann optimale Lösungen des primalen bzw. dualen Problems, wenn  $\phi(\mathbf{s}^*, \mathbf{y}^*, \pi^*, \boldsymbol{\mu}^*)$  ein Sattelpunkt der Lagrange-Funktion und damit ein Karush-Kuhn-Tucker- (KKT) Punkt ist. Das gleiche Kriterium gilt für die optimale Lösung  $(\bar{\mathbf{s}}, \bar{\mathbf{y}})$  bzw.  $(\bar{\pi}, \bar{\boldsymbol{\mu}})$  des primalen bzw. dualen RMP. Die KKT-Bedingungen, die für jeden KKT-Punkt erfüllt sind, können genutzt werden, um aus einer optimalen Lösung des primalen RMP die optimale duale Lösung zu errechnen. Zu den KKT-Bedingungen gehört unter anderem die Forderung, dass der Gradient der Lagrange-Funktion in einem globalen Optimum, d. h. in einem Sattelpunkt der Lagrange-Funktion, verschwindet. Nun wird angenommen, dass die optimale Lösung des dualen RMP  $(\bar{\pi}, \bar{\boldsymbol{\mu}})$  bekannt ist und  $(\bar{\mathbf{s}}, \bar{\mathbf{y}})$  berechnet werden soll. Sämtliche Ableitungen der aus dieser Annahme resultierenden Lagrange-Funktion  $\phi^\sigma(\mathbf{s}, \mathbf{y}, \bar{\pi}, \bar{\boldsymbol{\mu}})$  nach den primalen Entscheidungsvariablen  $\mathbf{s}$  und  $\mathbf{y}$  müssen aufgrund der KKT-Bedingung an der Stelle  $(\bar{\mathbf{s}}, \bar{\mathbf{y}}, \bar{\pi}, \bar{\boldsymbol{\mu}})$  den Wert 0 annehmen. Diese Forderung resultiert zum einen in folgendem Ausdruck für alle Transaktionen  $d = 1, \dots, D$ :

$$\left. \frac{\partial \phi^\sigma(\mathbf{s}, \mathbf{y}, \bar{\pi}, \bar{\boldsymbol{\mu}})}{\partial y_d} \right|_{(\mathbf{s}, \mathbf{y}) = (\bar{\mathbf{s}}, \bar{\mathbf{y}})} = \frac{1}{\bar{y}_d} - \bar{\mu}_d = 0 \Leftrightarrow \bar{\mu}_d = \frac{1}{\bar{y}_d} \quad (6.17)$$

Zum anderen ergibt sich folgender Ausdruck für alle Segmente  $l = 1, \dots, \tilde{L}$ , wobei die letzte Äquivalenz durch einige algebraische Umformungen gezeigt werden kann:<sup>23</sup>

$$\left. \frac{\partial \phi^\sigma(\mathbf{s}, \mathbf{y}, \bar{\pi}, \bar{\boldsymbol{\mu}})}{\partial s_l} \right|_{(\mathbf{s}, \mathbf{y}) = (\bar{\mathbf{s}}, \bar{\mathbf{y}})} = \bar{\pi} + \sum_{d=1}^D \bar{\mu}_d \alpha_{dl} = 0 \Leftrightarrow -\bar{\pi} = \sum_{d=1}^D \bar{\mu}_d \alpha_{dl} \Leftrightarrow -\bar{\pi} = D \quad (6.18)$$

Durch die Verwendung der Beziehungen (6.17) und (6.10) kann zu einer optimalen primalen Lösung  $(\bar{\mathbf{s}}, \bar{\mathbf{y}})$  des RMP mit sehr geringem Rechenaufwand auch die optimale duale Lösung  $(\bar{\pi}, \bar{\boldsymbol{\mu}})$  angegeben werden, die in jeder Iteration des Spaltengenerierungsverfahrens

<sup>23</sup> Vgl. van Ryzin und Vulcano (2015, S. 288) für die vollständige Herleitung.

benötigt wird.

Das Verfahren von van Ryzin und Vulcano (2015) beinhaltet folgende Schritte, die sich aus dem allgemeinen Ablauf eines Spaltengenerierungsverfahrens ableiten lassen und im Anschluss an diesen Überblick genauer erläutert werden (vgl. van Ryzin und Vulcano (2015, S. 290)).<sup>24</sup> In der Initialisierung wird der Variablenvektor  $\mathbf{s}^1$  des ersten RMP erstellt, indem die korrespondierende Teilmenge  $\boldsymbol{\sigma}^1$  aus der Menge aller Präferenzlisten ausgewählt wird. Die auf die Initialisierung folgenden Iterationen  $i = 1, \dots, I$  des Verfahrens können in drei Schritte gegliedert werden. Zuerst wird das aktuelle RMP mit dem Variablenvektor  $\mathbf{s}^i$  gelöst, um die Werte der Dualvariablen  $\bar{\boldsymbol{\mu}}^i$  gemäß (6.17) berechnen zu können. Anschließend wird eine Präferenzliste  $\sigma_{l'}$  generiert, deren korrespondierende Variable  $s_{l'}$  bei einer Aufnahme in die Menge der Variablen des RMP dazu führt, dass sich der Zielfunktionswert des RMP verbessert. Das damit verbundene lineare Optimierungsproblem wird als Subproblem der Spaltengenerierung bezeichnet und verwendet die aktuellen Werte der RMP-Dualvariablen  $\bar{\boldsymbol{\mu}}^i$  als Inputdaten. Das Subproblem steuert somit die schrittweise Spezifikation des Rank-based-Modells durch die Generierung relevanter Präferenzlisten. Im letzten Schritt wird geprüft, ob die Aufnahme der Variable  $s_{l'}$  den Zielfunktionswert des RMP tatsächlich verbessert und, falls dies der Fall ist, ob die Verbesserung statistisch signifikant ist. Wenn diese Prüfung ergibt, dass keine oder nur eine nicht signifikante Verbesserung erreicht werden kann, wird das Verfahren abgebrochen und die Lösung des RMP mit dem Variablenvektor  $\mathbf{s}^I$  wird als Lösung des ursprünglichen Schätzproblems  $\mathbf{s}^*$  ausgegeben. Ansonsten wird eine weitere Iteration  $i + 1$  durchgeführt und hierfür die Menge der einbezogenen Präferenzlisten gemäß  $\boldsymbol{\sigma}^{i+1} = \boldsymbol{\sigma}^i \cup \{\sigma_{l'}\} = \{\sigma_1, \dots, \sigma_{\bar{L}+1}\}$  aktualisiert. Das RMP der nächsten Iteration ergibt sich aus der Erweiterung von  $\mathbf{s}^i$  um  $s_{l'}$  und beinhaltet die Entscheidungsvariablen  $\mathbf{s}^{i+1}$ .

Vor Beginn der eigentlichen Spaltengenerierung muss, wie beschrieben, der initiale Variablenvektor  $\mathbf{s}^1$  spezifiziert werden (vgl. van Ryzin und Vulcano (2015, S. 288)). Dies ist der einzige Teilschritt der Spezifikation, der nicht anhand der Daten erfolgt, sondern manuell durch die Airline selbst. Dabei muss die Teilmenge  $\boldsymbol{\sigma}^1 = \{\sigma_1, \dots, \sigma_{\bar{L}^1}\}$  aller Präferenzlisten so gewählt werden, dass zu jeder Transaktion  $d = 1, \dots, D$  mindestens eine kompatible Präferenzliste enthalten ist. Formal bedeutet dies:

---

<sup>24</sup> Auf die allgemeine Darstellung des Prinzips der Spaltengenerierung wird in dieser Arbeit verzichtet. Eine solche Betrachtung, die auch die nichtlineare Optimierung einbezieht, findet sich in Larsson et al. (2015). Eine didaktische Darstellung im Kontext der linearen Optimierung ist in Desrosiers und Lübbecke (2005) enthalten.

$$|\mathcal{M}_{j_d}^{\sigma^1}(S_d)| \geq 1, \forall d = 1, \dots, D \text{ mit } \mathcal{M}_{j_d}^{\sigma^1}(S_d) = \{l: l \in \mathcal{M}_{j_d}(S_d), \sigma_l \in \sigma^1\} \quad (6.19)$$

Ist Bedingung (6.19) für eine Transaktion  $d$  verletzt, führt dies zu  $y_d = 0$  und einem unzulässigen Summanden  $\ln(0)$  in der Zielfunktion (6.12). Ein Spezifikationsansatz, der diese Bedingungen in jedem Fall erfüllt, ist der Independent-Demand-Ansatz. Diesem Ansatz folgend, wird für jedes Produkt  $j \in \mathcal{J} \setminus \{0\}$  eine Präferenzliste mit  $\sigma(j) = 1$  und  $\sigma(0) = 2$  erstellt, deren verkürzte Form lediglich Produkt  $j$  enthält.

Ziel jeder der nachfolgenden Iterationen  $i$  ist die Identifikation einer Präferenzliste  $\sigma_{l'}$ , die in das RMP aufgenommen werden sollte, um dessen Zielfunktionswert zu verbessern (vgl. van Ryzin und Vulcano (2015, S. 289) für die folgenden Überlegungen). Eine solche verbessernde Präferenzliste existiert dann, wenn noch nicht alle relevanten Präferenzlisten bereits in  $\sigma^i$  enthalten sind. Dies ist gleichbedeutend damit, dass die aus  $(\bar{s}^i, \bar{y}^i)$  konstruierte Lösung  $(s^i, y^i)$  für das ursprüngliche Schätzproblem (6.8) - (6.11) dieses zulässig, aber nicht optimal löst. Folglich ist der Punkt  $(s^i, y^i, \bar{\beta}^i, \bar{\mu}^i)$  kein KKT-Punkt der Lagrange-Funktion  $\phi(s, y, \bar{\beta}, \bar{\mu})$  des ursprünglichen Schätzproblems und es existiert mindestens ein Segment  $l = 1, \dots, L$ , für das die KKT-Bedingung (6.18) nicht erfüllt ist. Aufgrund der Optimalität der Lösung  $(\bar{s}^i, \bar{y}^i)$  bzgl. des RMP ist die Bedingung für die Segmente  $l = 1, \dots, \tilde{L}$  in jedem Fall erfüllt. D. h. die Bedingung wird durch ein Segment verletzt, welches zwar im ursprünglichen Problem, aber nicht im RMP enthalten ist. Formal muss daher für mindestens eines der Segmente  $l' = \tilde{L} + 1, \dots, L$  gelten:

$$\frac{\partial \phi(s, y, \bar{\pi}, \bar{\mu})}{\partial s_{l'}} = \bar{\pi} + \sum_{d=1}^D \bar{\mu}_d \alpha_{dl'} > 0 \Leftrightarrow \sum_{d=1}^D \bar{\mu}_d \alpha_{dl'} > D \text{ mit } -\bar{\pi} = D \quad (6.20)$$

Das Subproblem besteht also darin, ein Segment  $l'$  zu ermitteln, für das der Ausdruck  $\sum_{d=1}^D \bar{\mu}_d \alpha_{dl'}$  maximal wird. Dafür muss über die Struktur der Präferenzliste  $\sigma_{l'}$  des neu aufzunehmenden Segments  $l'$  entschieden werden. Hierfür werden die Parameter  $\alpha_{dl'}$  durch die binäre Hilfsvariable  $w_d$  für  $d = 1, \dots, D$  ersetzt, die genau dann den Wert 1 annimmt, wenn die generierte Präferenzliste  $\sigma_{l'} \in \mathcal{M}_{j_d}(S_d)$ , d. h. kompatibel mit der Transaktion  $d$  ist. Mithilfe einer Indikatorfunktion  $\mathbf{I}$  ergibt sich so zunächst folgendes Optimierungsproblem (vgl. van Ryzin und Vulcano (2015, S. 289f)):

$$\text{Max } SP(\mathbf{w}) = \sum_{d=1}^D \bar{\mu}_d w_d \quad (6.21)$$

unter den Nebenbedingungen

$$w_d \leq \mathbf{I}\{\sigma_{l'}(j_d) < \sigma_{l'}(k) \mid \forall k \in S, k \neq j_d\} \quad \forall d = 1, \dots, D \quad (6.22)$$

$$w_d \in \{0; 1\} \quad \forall d = 1, \dots, D \quad (6.23)$$

Um das Subproblem als gemischt-ganzzahliges Problem ohne die Indikatorfunktion formulieren zu können, werden die binären Entscheidungsvariablen  $x_{jk}$  eingeführt. Diese codieren die Produktreihenfolge innerhalb der Präferenzliste  $\sigma_{l'}$ . Eine Variable  $x_{jk}$  nimmt genau dann den Wert 1 an, wenn das Produkt  $j$  in der Präferenzliste vor Produkt  $k$  positioniert ist. Es ergibt sich das zu (6.21) - (6.23) äquivalente Optimierungsproblem:<sup>25</sup>

$$\text{Max } SP(\mathbf{x}, \mathbf{w}) = \sum_{d=1}^D \bar{\mu}_d w_d \quad (6.24)$$

unter den Nebenbedingungen

$$x_{ji} + x_{ij} = 1 \quad \forall j, i \in \mathcal{J}, j < i \quad (6.25)$$

$$x_{ji} + x_{ik} + x_{kj} \leq 2 \quad \forall j, i, k \in \mathcal{J}, j \neq i \neq k \quad (6.26)$$

$$\sum_{j=1}^n x_{j0} \geq 1 \quad (6.27)$$

$$w_d \leq x_{j_d, i} \quad \forall d = 1, \dots, D, i \in S_d, j_d \neq i \quad (6.28)$$

$$x_{ji}, w_d \in \{0; 1\} \quad \forall d = 1, \dots, D, j, i \in \mathcal{J} \quad (6.29)$$

Die ersten beiden Typen von Nebenbedingungen sorgen dafür, dass die Variablen  $x_{jk}$  in jeder zulässigen Lösung genau eine der möglichen Präferenzlisten codieren. Dafür muss einerseits sichergestellt sein, dass für jedes Produktpaar  $i$  und  $j$  eine eindeutige Präferenz ausgedrückt wird, d. h. entweder  $\sigma_{l'}(i) < \sigma_{l'}(j)$  oder  $\sigma_{l'}(i) > \sigma_{l'}(j)$  gilt. Andererseits

---

<sup>25</sup> Der Index  $i$  repräsentiert in diesem Optimierungsproblem nicht die aktuelle Iteration des Spaltengenerierungsverfahrens, sondern ein Produkt. Es gilt  $i \in \mathcal{J}$ .

müssen die Präferenzen, wie bereits in Abschnitt 6.1 geschildert, transitiv sein. Dies wird durch Nebenbedingung (6.26) garantiert. Eine weitere Nebenbedingung verhindert, dass eine irrelevante Präferenzliste mit  $\sigma_{l'}(0) = 1$ , d. h. der Nichtkaufalternative an erster Stelle, generiert wird. Der letzte Nebenbedingungstyp (6.28) hat die gleiche Funktion wie die Nebenbedingungen (6.22) und sorgt dafür, dass  $w_d$  nur dann auf den Wert 1 gesetzt werden darf, wenn die von den Variablen  $x_{jk}$  repräsentierte Präferenzliste  $\sigma_{l'}$  mit der Transaktion  $d$  kompatibel ist. Die Parameter  $\sigma_{l'}(j)$ , welche die Positionen der Produkte in der neu generierten Präferenzliste  $\sigma_{l'}$  angeben, können gemäß  $\sigma_{l'}(j) = \sum_{k \in J, k \neq j} x_{kj}^* + 1$  aus der optimalen Lösung des Subproblems  $(\mathbf{x}^*, \mathbf{w}^*)$  berechnet werden.

Im letzten Schritt jeder Iteration des Verfahrens von van Ryzin und Vulcano (2015) werden die beiden Abbruchkriterien geprüft (vgl. van Ryzin und Vulcano (2015, S. 290) für die folgenden Überlegungen).<sup>26</sup> Das erste Kriterium fordert für eine Fortsetzung des Verfahrens, dass aus der Lösung des Subproblems eine verbessernde Präferenzliste  $\sigma_{l'}$  resultiert. Dieses Kriterium lässt sich unmittelbar überprüfen: Nur dann, wenn der optimale Zielfunktionswert des Subproblems  $SP(\mathbf{x}^*, \mathbf{w}^*)$ , wie in Bedingung (6.20) formuliert, größer ist als die Anzahl der Transaktionen  $D$ , ist  $\sigma_{l'}$  eine verbessernde Präferenzliste. Ist diese Bedingung nicht erfüllt wird das Verfahren folglich abgebrochen. Erfüllt die generierte Präferenzliste  $\sigma_{l'}$  dieses Kriterium, wird anschließend das zweite Abbruchkriterium, die Signifikanz der Verbesserung, geprüft. Um festzustellen, ob die Verbesserung der Log-Likelihood durch Aufnahme der Variable  $s_{l'}$  statistisch signifikant ist, wird ein Log-Likelihood-Ratio-Test durchgeführt. Dieser Test wird allgemein eingesetzt, um die Nullhypothese  $H_0$ , dass ein oder mehrere Modellparameter den Wert 0 aufweisen, zu untersuchen (vgl. Train (2009, S. 70) für die folgenden Überlegungen).  $H_0$  kann somit als Nebenbedingung für die Parameterschätzung aufgefasst werden. Um herauszufinden, ob sich diese Nebenbedingung signifikant auf die Qualität der Parameterschätzung auswirkt, wird der Quotient  $\frac{L^{H_0}}{L}$  gebildet.  $L^{H_0}$  bezeichnet dabei den Zielfunktionswert, also die maximale Likelihood, der Schätzung unter Berücksichtigung der durch  $H_0$  implizierten Nebenbedingungen.  $L$  repräsentiert die optimale Likelihood des unbeschränkten Schätzproblems. Aus diesem Quotienten kann durch Logarithmieren die Teststatistik  $-2(LL^{H_0} - LL)$  abgeleitet werden, die einer  $\chi^2$ -Verteilung folgt. Die Anzahl der Freiheitsgrade entspricht der Anzahl der mit  $H_0$  verbundenen

---

<sup>26</sup> Über die beiden in dieser Arbeit beschriebenen Kriterien hinaus können weitere Abbruchkriterien, wie etwa das Erreichen einer maximalen Anzahl von Segmenten oder die Prüfung eines Informationskriteriums, in das Verfahren integriert werden (vgl. van Ryzin und Vulcano (2015, S. 290)).

Nebenbedingungen. Wenn der Wert der Teststatistik größer ist als der korrespondierende kritische Wert der  $\chi^2$ -Verteilung, dann wird  $H_0$  abgelehnt.

In jeder Iteration des Spaltengenerierungsverfahrens lautet  $H_0$ , dass die neu generierte Präferenzliste  $\sigma_{l'}$  irrelevant ist und somit  $s_{l'} = 0$  für den Anteil des Kundensegments mit dieser Präferenzliste an der Gesamtpopulation gilt (vgl. van Ryzin und Vulcano (2015, S. 290)). Dies ist im aktuellen RMP jeder Iteration  $i$  mit dem optimalen Zielfunktionswert  $LL^{\sigma^i}(\bar{s}^i)$  der Fall, welches das neu generierte Segment  $l'$  noch nicht enthält. Das unbeschränkte Schätzproblem wird durch das um  $\sigma_{l'}$  erweiterte RMP mit dem optimalen Zielfunktionswert  $LL^{\sigma^{i+1}}(\bar{s}^{i+1})$  der möglichen darauffolgenden Iteration repräsentiert. Für dieses gilt  $s_{l'} \geq 0$ . Daraus ergibt sich folgende Teststatistik:

$$-2 \left( LL^{\sigma^i}(\bar{s}^i) - LL^{\sigma^{i+1}}(\bar{s}^{i+1}) \right) \quad (6.30)$$

Da in dem beschränkten Schätzproblem lediglich eine zusätzliche Restriktion enthalten ist, weist die  $\chi^2$ -Verteilung der Teststatistik einen Freiheitsgrad auf. Der kritische Wert auf dem Signifikanzniveau von 5% liegt bei 3,842. Das Verfahren wird folglich fortgeführt, wenn  $-2 \left( LL^{\sigma^i}(\bar{s}^i) - LL^{\sigma^{i+1}}(\bar{s}^{i+1}) \right) > 3,842$  gilt. Abschließend wird in Abbildung 13 das gesamte Verfahren in Form eines Pseudocodes dargestellt.<sup>27</sup>

---

<sup>27</sup> In dieser Darstellung wird der Likelihood-Ratio-Test, außer in der ersten Iteration, jeweils zu Beginn der Iterationen durchgeführt. Damit wird vermieden, dass am Ende jeder Iteration bereits das RMP der darauffolgenden Iteration für den Likelihood-Ratio-Test gelöst werden muss, welches dann zu Beginn der Folgeiteration erneut gelöst werden muss, um die Werte der Dualvariablen zu berechnen.

```

i := 1, konv := False,  $\sigma^1 := \{\sigma_1, \dots, \sigma_{\tilde{L}^1}\}$ ,  $\mathbf{s}^1 := (s_1, \dots, s_{\tilde{L}^1})^T$ 
while konv = False:
     $\mathcal{M}_{j_d}^{\sigma^i}(S_d) := \{l: \sigma_l(j_d) < \sigma_l(k) \forall k \in S, k \neq j_d, \sigma_l \in \sigma^i\}$ 
    Berechne  $\bar{\mathbf{s}}^i$  und  $LL^{\sigma^i}(\bar{\mathbf{s}}^i)$  durch Lösen des aktuellen RMP
    if i ≠ 1 and  $-2 \left( LL^{\sigma^{i-1}}(\bar{\mathbf{s}}^{i-1}) - LL^{\sigma^i}(\bar{\mathbf{s}}^i) \right) < 3,842$ :
        break
    for d = 1, ..., D:
         $\bar{y}_d^i := \sum_{l \in \mathcal{M}_{j_d}^{\sigma^i}(S_d)} \bar{s}_l^i$ 
         $\bar{\mu}_d^i := (\bar{y}_d^i)^{-1}$ 
        Berechne  $\mathbf{x}^*$ ,  $\mathbf{w}^*$  und  $SP(\mathbf{x}^*, \mathbf{w}^*)$  durch Lösen des Subproblems mit  $\bar{\mu} = \bar{\mu}^i$ 
        for j ∈ J:
             $\sigma_{l'}(j) := \sum_{k \in J, k \neq j} x_{kj}^* + 1$ 
        if  $SP(\mathbf{x}^*, \mathbf{w}^*) > D$ :
             $\sigma^{i+1} := \sigma^i \cup \{\sigma_{l'}\} = \{\sigma_1, \dots, \sigma_{\tilde{L}^{i+1}}\}$ 
             $\mathbf{s}^{i+1} := (s_1, \dots, s_{\tilde{L}^{i+1}})^T$ 
        else:
            konv := True
    i = i + 1

```

Abbildung 13: Ablauf des Spaltengenerierungsverfahrens von van Ryzin und Vulcano (2015)

## 6.4 Exemplarische Spezifikation und Schätzung des Rank-based-Modells

Die initiale Spezifikation und der Ablauf einer Iteration des Schätzverfahrens von van Ryzin und Vulcano (2015) sollen nun anhand des Anwendungsbeispiels veranschaulicht werden. Dem Independent-Demand-Ansatz folgend, kann die Airline, um eine zulässige Startspezifikation zu erhalten, für jedes Produkt  $j = 1, \dots, 4$  der Beispielinstantz eine Präferenzliste der Form  $\sigma_j = (j)$  konstruieren. So wird  $\sigma^{1,ID} := \{(1), (2), (3), (4)\}$  und  $\mathbf{s}^{1,ID} := (s_1, \dots, s_4)^T$ ,

mit  $\tilde{L}^1 = 4$ , gesetzt. Die Auflistung der zu jeder Transaktion  $d = 1, \dots, 10$  kompatiblen Präferenzlisten  $\sigma_l \in \sigma^{1,ID}$  in Tabelle 9 zeigt, dass diese Spezifikation zulässig ist.

$d$	1	2	3	4	5	6	7	8	9	10
$\mathcal{M}_{j_d}^{\sigma^{1,ID}}(S_d)$	{1}	{4}	{2}	{2,4}	{4}	{1}	{2}	{4}	{1}	{2,3,4}
$\mathcal{M}_{j_d}^{\sigma^1}(S_d)$	{2}	{1}	{2}	{1}	{1}	{2}	{1}	{1}	{2}	{1}
$\mathcal{M}_{j_d}^{\sigma^2}(S_d)$	{2,3}	{1,3}	{2,3}	{1}	{1}	{2,3}	{1,3}	{1}	{2,3}	{1}

**Tabelle 9: Kompatible Präferenzlisten für  $d = 1, \dots, 10$  und  $\sigma^{1,ID}$ ,  $\sigma^1$  sowie  $\sigma^2$**

Die Fluggesellschaft verfügt jedoch über Daten aus Marktanalysen und nimmt an, dass  $\sigma^1 = \{(4,2), (1,3,2)\}$  eine geeignete Startspezifikation für das Modell darstellt. Dass auch diese zulässig ist, geht ebenfalls aus Tabelle 9 hervor.

In der ersten Iteration des Verfahrens wird nun das RMP formuliert und gelöst. Dafür ist zunächst die Ermittlung der Mengen  $\mathcal{M}_{j_d}^{\sigma^1}(S_d)$  für alle Transaktionen  $d$  erforderlich (vgl. Tabelle 9). So ist beispielsweise nur  $\sigma_2$  mit Transaktion  $d = 1$  kompatibel, da alle Produkte angeboten werden und das gekaufte Produkt  $j = 1$  an erster Stelle dieser Präferenzliste steht. Ein Kunde, der auf Basis von  $\sigma_1$  entscheidet, würde aus  $S_1$  das Produkt  $j = 4$  wählen und kann daher nicht für die erste Transaktion verantwortlich sein. Im nächsten Schritt wird das RMP in der Form (6.12 - 6.14) mit den Entscheidungsvariablen  $s_1$  und  $s_2$  formuliert:

$$\text{Max } LL^{\sigma^1}(s_1, s_2) = \ln(s_2) + \ln(s_1) + \dots + \ln(s_1) = 6 \ln(s_1) + 4 \ln(s_2) \quad (6.31)$$

unter den Nebenbedingungen

$$s_1 + s_2 = 1 \quad (6.32)$$

$$s_1, s_2 \geq 0 \quad (6.33)$$

In diesem Beispiel wird die Lösung  $\bar{s}^1 = (0,6; 0,4)$  des ersten RMP mit dem Zielfunktionswert  $LL^{\sigma^1}(0,6; 0,4) = -6,73$  graphisch anhand von Abbildung 14 ermittelt.



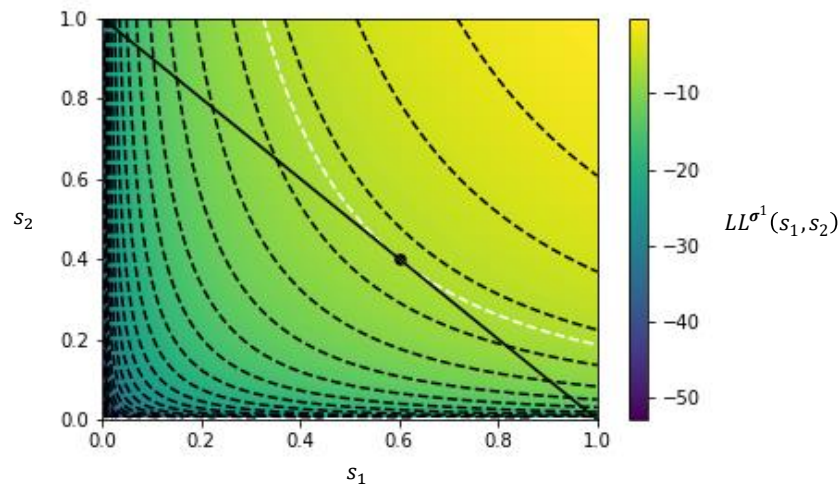


Abbildung 14: Diagramm des RMP (6.31) - (6.33) in der ersten Iteration  $i = 1$

Da in der ersten Iteration kein Signifikanztest durchgeführt werden muss, werden im Anschluss die optimalen Werte der Hilfsvariablen  $\bar{y}_d$  und mit diesen die Werte der Dualvariablen  $\bar{\mu}_d$  für alle Transaktionen  $d$  berechnet. Tabelle 10 stellt das Ergebnis dieser Berechnungen dar.

$d$	1	2	3	4	5	6	7	8	9	10
$\bar{y}_d$	0,4	0,6	0,4	0,6	0,6	0,4	0,6	0,6	0,4	0,6
$\bar{\mu}_d$	2,5	1,67	2,5	1,67	1,67	2,5	1,67	1,67	2,5	1,67

Tabelle 10: Optimale Werte der Variablen  $\bar{y}_d$  und  $\bar{\mu}_d$  für  $d = 1, \dots, 10$

Die Werte  $\bar{\mu}_d$  dienen als Parameter in der Zielfunktion des Subproblems. Da dieses insgesamt 115 Nebenbedingungen enthält, werden zu jedem der vier Nebenbedingungstypen nur einige Beispiele explizit formuliert.

$$\text{Max } 2,5w_1 + 1,67w_2 + \dots + 1,67w_{10} \quad (6.34)$$

unter den Nebenbedingungen

$$x_{01} + x_{10} = 1, x_{02} + x_{20} = 1, \dots, x_{34} + x_{43} = 1 \quad (6.35)$$

$$x_{01} + x_{12} + x_{20} \leq 2, x_{01} + x_{13} + x_{30} \leq 2, \dots, x_{23} + x_{34} + x_{42} \leq 2 \quad (6.36)$$

$$x_{10} + x_{20} + x_{30} + x_{40} \geq 1 \quad (6.37)$$

$$w_1 \leq x_{10}, w_1 \leq x_{12}, w_1 \leq x_{13}, w_1 \leq x_{14}, \dots, w_{10} \leq x_{01} \quad (6.38)$$

$$x_{ji}, w_d \in \{0; 1\} \forall d = 1, \dots, 10, j, i \in \{0, 1, 2, 3, 4\} \quad (6.39)$$

Dieses gemischt-ganzzahlige Optimierungsproblem wird nun durch einen Standard-Solver gelöst. Es ergibt sich folgende optimale Lösung:

$$\mathbf{x}^* = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \end{pmatrix} \text{ und } \mathbf{w}^* = (1, 1, 1, 0, 0, 1, 1, 0, 1, 0) \text{ mit } SP(\mathbf{x}^*, \mathbf{w}^*) = 13,33 \quad (6.40)$$

Die Position eines Produkts  $j = 0, \dots, 4$  in der neu generierten Präferenzliste  $\sigma_3 = (1, 2, 4, 3)$  wird durch Bildung der  $j$ -ten Spaltensumme von  $\mathbf{x}^*$ , zu der jeweils der Wert 1 addiert wird, bestimmt. Die Prüfung des ersten Abbruchkriteriums am Ende der Iteration zeigt, dass mit  $\sigma_3$  eine verbessernde Präferenzliste gefunden wurde, denn  $SP(\mathbf{x}^*, \mathbf{w}^*) = 13,33 > D = 10$ . Daher wird das Verfahren mit  $\sigma^2 = \{(4, 2), (1, 3, 2), (1, 2, 4, 3)\}$  und  $\mathbf{s}^2 = (s_1, s_2, s_3)$  zunächst fortgesetzt. Mit Beginn von Iteration  $i = 2$  werden die Mengen  $\mathcal{M}_{j_d}^{\sigma^2}(S_d)$  ermittelt, die in Tabelle 9 angegeben sind. Das aktualisierte RMP lautet demzufolge:

$$\begin{aligned} \text{Max } LL^{\sigma^1}(s_1, s_2) &= \ln(s_2 + s_3) + \ln(s_1 + s_3) + \dots + \ln(s_1) \\ &= 4 \ln(s_1) + 2 \ln(s_1 + s_3) + 4 \ln(s_2 + s_3) \end{aligned} \quad (6.41)$$

unter den Nebenbedingungen

$$s_1 + s_2 + s_3 = 1 \quad (6.42)$$

$$s_1, s_2, s_3 \geq 0 \quad (6.43)$$

Die optimale Lösung  $\bar{\mathbf{s}}^2 = (0,5; 0; 0,5)$  mit  $LL^{\sigma^2}(0,5; 0; 0,5) = -5,55$  des neuen RMP wird in dieser Iteration nun auch mithilfe eines Standard-Solvers für nichtlineare Optimierungsprobleme ermittelt. Im Anschluss wird das zweite Abbruchkriterium, d. h. die Signifikanz der Verbesserung des optimalen Zielfunktionswerts gegenüber dem des RMP in Iteration  $i = 1$ , geprüft. Für die Teststatistik des Likelihood-Ratio-Tests ergibt sich:

$$-2 \left( LL^{\sigma^1}(\bar{\mathbf{s}}^1) - LL^{\sigma^2}(\bar{\mathbf{s}}^2) \right) = 2,36 < 3,842 \quad (6.44)$$

Die Aufnahme von  $\sigma_3$  und des korrespondierenden Modellparameters  $s_3$  in das RMP führt somit zwar zu einer Verbesserung des Zielfunktionswertes, die jedoch nicht signifikant ist.

Daher wird das Verfahren abgebrochen und die zu Beginn von Iteration  $i = 1$  ermittelte Lösung des RMP wird als Lösung des Schätzproblems  $\mathbf{s}^* = (0,6; 0,4)$  mit  $LL(\mathbf{s}^*) = -6,73$  ausgegeben. Die initiale Modellspezifikation  $\boldsymbol{\sigma}^1 = \{(4,2), (1,3,2)\}$  wird somit beibehalten. Das Rank-based-Modell weist für die hier betrachtete Beispielinstantz eine deutlich höhere Modellierungsgüte als das einfache MNL-Modell und eine ähnlich hohe Güte wie das LC-MNL-Modell auf.

## 7 Markov-Chain-Modell

Das Markov-Chain-Modell ist ein komplexes, RUT-kompatibles Discrete-Choice-Modell, das zur Klasse der parametrischen Modelle gezählt werden kann (vgl. Berbeglia et al. (2018, S. 16f)), obwohl es auch einige typische Eigenschaften nicht-parametrischer Modelle aufweist. Der Entscheidungsprozess der Kunden wird in diesem Modell als Markov-Prozess modelliert, in dem für jedes Produkt ein Zustand existiert. Das Substitutionsverhalten der Kunden wird durch die Übergangswahrscheinlichkeiten zwischen diesen Zuständen abgebildet. Im folgenden Abschnitt werden die Modellstruktur und die bereits erwähnten Eigenschaften des Markov-Chain-Modells vorgestellt. Eine dieser Eigenschaften ist, dass die Anzahl und Art der zu schätzenden Modellparameter ausschließlich von der Produktanzahl des zu modellierenden Anwendungsfalls abhängt. Somit muss das Parameterset weder durch die Fluggesellschaft noch anhand der Trainingsdaten während des Schätzverfahrens spezifiziert werden. Der zweite Abschnitt befasst sich daher bereits mit dem EM-Verfahren von Simsek und Topaloglu (2018), welches im letzten Abschnitt des Kapitels veranschaulicht wird.

### 7.1 Annahmen und Modellstruktur

Zentrales Element des Markov-Chain-Modells ist eine Markov-Kette mit  $|\mathcal{J}|$  Zuständen, so dass jedes Produkt  $j \in \mathcal{J}$  mit einem Zustand  $X_t$  korrespondiert (vgl. Gupta und Hsu (2017, S. 2)). Der Entscheidungsprozess des Kunden wird in diesem Modell in diskrete Zeitpunkte  $t = 1, \dots, T$  unterteilt. Es wird angenommen, dass jeder Kunde eines der Produkte grundsätzlich gegenüber allen anderen bevorzugt. Die Wahrscheinlichkeit dafür, dass ein ankommender Kunde Produkt  $j \in \mathcal{J}$  global bevorzugt, d. h. zu Beginn des Entscheidungsprozesses den Kauf dieses Produkts beabsichtigt, ist durch den Parameter  $\lambda_j = P(X_1 = j)$  gegeben (vgl. Simsek und Topaloglu (2018, S. 750)). Wenn dieses Produkt im Offerset  $S$  enthalten ist, findet ein Kauf statt und der Kunde hat einen absorbierenden Zustand erreicht,

d. h. der Entscheidungsprozess ist beendet. Anderenfalls folgt ein Zustandsübergang zwischen  $t = 1$  und  $t = 2$ . Der Kunde entscheidet sich dann entweder für den Nichtkauf, was dem Zustand  $X_2 = 0$  entspricht, oder plant den Kauf eines alternativen Produkts  $k \neq j$ , wechselt also in den Zustand  $X_2 = k$ . Ist dieses alternative Produkt ebenfalls nicht verfügbar, entscheidet sich der Kunde wieder entweder für den Nichtkauf oder er zieht den Kauf eines weiteren alternativen Produkts in Erwägung. Der Prozess dauert an, bis in einem Zeitpunkt  $t$  ein absorbierender Zustand  $X_t = j$  mit  $j \in S$  erreicht wird und der Kunde somit eine Kaufentscheidung für Produkt  $j$  getroffen hat. Da  $0 \in S$  stets gilt, ist der Zustand  $X_t = 0$  in allen Entscheidungsprozessen absorbierend. Ausgehend von einem Zustand  $X_t = j$  ist die Wahrscheinlichkeit für einen Übergang in Zustand  $X_{t+1} = k$  durch den Parameter  $\rho_{jk}$  gegeben.<sup>28</sup> In Abhängigkeit des Offersets  $S$  werden diese Parameter für jede Entscheidungssituation wie folgt angepasst, damit die Zustände  $X_t = j$  mit  $j \in S$  die Eigenschaft eines absorbierenden Zustands erfüllen (vgl. Blanchet et al. (2016, S. 891)):

$$\rho_{jk}(S) = \begin{cases} 0 & , \text{ falls } j \in S, j \neq k \\ 1 & , \text{ falls } j \in S, j = k \\ \rho_{jk} & , \text{ sonst} \end{cases} \quad (7.1)$$

Die Übergangsmatrix  $\boldsymbol{\rho}$  bildet mit den Wahrscheinlichkeiten für den initialen Zustand  $\boldsymbol{\lambda}$  die Parametermenge des Markov-Chain-Modells (vgl. Gupta und Hsu (2017, S. 2)). Wie eingangs bereits erläutert, ergibt sich die Anzahl der Parameter beider Typen allein aus der Anzahl der Produkte. Der Parametervektor  $\boldsymbol{\lambda}$  weist  $|J|$  Elemente auf, die Parametermatrix  $\boldsymbol{\rho}$  enthält  $|J|^2$  Einträge. Daraus folgt, dass die Anzahl der Parameter nicht durch die Fluggesellschaft beeinflusst werden kann. So besteht keine Möglichkeit dem Overfitting-Problem durch die Begrenzung der Parameterzahl vorzubeugen. Das Markov-Chain-Modell ist deswegen ungeeignet für Anwendungsfälle mit einer großen Anzahl von Produkten und einem Trainingsdatensatz mit wenigen Transaktionen (vgl. Simsek und Topaloglu (2018, S. 757f)). Eine charakteristische Eigenschaft von Markov-Ketten ist deren Gedächtnislosigkeit, die in diesem Modell folglich auch für das Substitutionsverhalten der Kunden angenommen wird. Mit dieser Eigenschaft geht einher, dass die Übergangswahrscheinlichkeiten  $\rho_{jk}$ , ausgehend von einem aktuellen Zustand  $X_t = j$  für alle anderen Zustände  $X_{t+1} = k \in J$ , nicht von den Zuständen abhängen, die vor Zeitpunkt  $t$  durchlaufen wurden (vgl. Feldman und Topaloglu (2017, S. 1325)). Übertragen auf das Kundenwahlverhalten bedeutet dies: Ein Kunde, der

---

<sup>28</sup> Die Übergangsmatrix wird in der Literatur unterschiedlich definiert. So definieren etwa Blanchet et al. (2016) die Parameter  $\rho_{jk}$  nur für alle  $j, k \in J \setminus \{0\}$ . Die in dieser Arbeit verwendete Definition wird beispielsweise von Simsek und Topaloglu (2018) gebraucht.

Produkt  $j \notin S$  global bevorzugt, zeigt ein exakt gleiches Substitutionsverhalten wie ein Kunde, der eigentlich Produkt  $k \notin S$  gegenüber allen anderen Produkten bevorzugt und nun stattdessen den Kauf von Produkt  $j$  in Erwägung zieht (vgl. Blanchet et al. (2016, S. 889)). Neben der Gedächtnislosigkeit nimmt das Modell an, dass die Kunden die gesamte Produktpalette  $\mathcal{J}$  kennen und in ihr Entscheidungskalkül einbeziehen. Außerdem wird angenommen, dass der Kauf eines nicht verfügbaren Produkts  $j \notin S$  von einem Kunden im Laufe seines Entscheidungsprozesses mehrfach in Erwägung gezogen werden kann. Diese drei Voraussetzungen sind in der Realität selten erfüllt, weswegen das Modell den realen Entscheidungsprozess in den meisten Fällen nicht nachbilden kann. Für das Erreichen einer hohen Modellierungsgüte ist dies jedoch auch nicht erforderlich, denn diese hängt lediglich davon ab, wie exakt das Modell durch die berechneten Kaufwahrscheinlichkeiten  $P_j(S)$  für die Produkte  $j \in \mathcal{J}$  die letztendlichen Kaufentscheidungen der Kunden prognostizieren kann (vgl. Feldman und Topaloglu (2017, S. 1325)). Der Entscheidungsprozess wird gewissermaßen als „Black-Box“ modelliert. Dies erklärt, weshalb das Markov-Chain-Modell keine Rückschlüsse auf das Zustandekommen der Kaufentscheidung sowie die Reaktion der Kunden auf die Änderung von Produkteigenschaften zulässt und in diesem Punkt den nicht-parametrischen Modellen ähnlich ist.

Im Gegensatz zu den bisher betrachteten Modellen können die Kaufwahrscheinlichkeiten  $P_j(S)$  der Markov-Chain-Modells nicht über einen Ausdruck in geschlossener Form berechnet werden (vgl. Simsek und Topaloglu (2018, S. 750) für die folgenden Überlegungen). Stattdessen ist die Lösung eines Gleichungssystems nötig. Dieses berechnet die erwartete Anzahl von Besuchen  $\Theta(S) = (\Theta_0(S), \dots, \Theta_n(S))^T$  des Produkts  $j$ , d. h. des Auftretens des Zustands  $X_t = j$ , im Verlauf des Entscheidungsprozesses über alle Zeitpunkte  $t$ :

$$\Theta_j(S) = \lambda_j + \sum_{k \in \mathcal{J} \setminus S} \rho_{kj} \Theta_k(S) \quad \forall j \in \mathcal{J} \quad (7.2)$$

Wie die Struktur dieses Gleichungssystems zeigt, ist die erwartete Gesamtzahl an Besuchen von Produkt  $j$  die Summe aus der erwarteten Besuchsanzahl zu Beginn des Prozesses in  $t = 1$  und den Besuchen in den übrigen Zeitpunkten  $t = 2, \dots, T$  des Prozesses, ausgehend von den nicht absorbierenden Zuständen  $X_t = k \in \mathcal{J} \setminus S$ . Der letztgenannte Erwartungswert ergibt sich aus der erwarteten Besuchsanzahl aller Zustände  $X_t = k \in \mathcal{J} \setminus S$  in  $t = 2, \dots, T$ , multipliziert mit der jeweiligen Übergangswahrscheinlichkeit  $\rho_{kj}$  von Zustand  $X_t = k$  in Zustand  $X_{t+1} = j$ . Da der Prozess nach dem Besuch eines absorbierenden Zustands  $X_t = j \in S$  in einem beliebigen Zeitpunkt  $t$  mit dem Kauf von Produkt  $j$  endet, gilt  $\Theta_j(S) \leq 1$  für

alle angebotenen Produkte  $j \in S$ . Die erwartete Gesamtzahl an Besuchen der Zustände dieser Produkte ist somit gleichzusetzen mit der Wahrscheinlichkeit dafür, dass ein solcher Zustand im Verlauf des Entscheidungsprozesses überhaupt einmal besucht wird. Diese Wahrscheinlichkeit entspricht wiederum der Kaufwahrscheinlichkeit, denn im Falle eines Besuchs findet stets ein Kauf statt. Für die Kaufwahrscheinlichkeiten  $P_j(S)$  für alle Produkte  $j \in S$  gemäß dem Markov-Chain-Modell gilt folglich:<sup>29</sup>

$$P_j(S) = \Theta_j(S) \text{ mit } \Theta_j(S) = \lambda_j + \sum_{k \in \mathcal{J} \setminus S} \rho_{kj} \Theta_k(S) \quad \forall j \in \mathcal{J} \quad (7.3)$$

## 7.2 EM-Verfahren von Simsek und Topaloglu (2018)

Um das Markov-Chain-Modell anwenden zu können, müssen die Modellparameter  $(\lambda, \rho)$  mithilfe des Maximum-Likelihood-Ansatzes und auf Grundlage von Trainingsdaten geschätzt werden. Wie bereits im ersten Abschnitt dieses Kapitels erläutert, setzt sich die Menge aller Modellparameter aus den Wahrscheinlichkeiten für den initialen Zustand  $\lambda$  sowie den Übergangswahrscheinlichkeiten  $\rho$  zusammen. Aus dem Einsetzen des Ausdrucks (7.3) in (3.16) ergibt sich folgendes Maximum-Likelihood-Schätzproblem für das Markov-Chain-Modell (vgl. Simsek und Topaloglu (2018, S. 751)):

$$\text{Max } LL(\lambda, \rho) = \sum_{d=1}^D \ln(P_{j_d}(S_d)) = \sum_{d=1}^D \ln(\Theta_{j_d}(S_d)) \quad (7.4)$$

unter den Nebenbedingungen

$$\sum_{j \in \mathcal{J}} \lambda_j = 1 \quad (7.5)$$

$$\sum_{k \in \mathcal{J}} \rho_{jk} = 1 \quad \forall j \in \mathcal{J} \quad (7.6)$$

$$\Theta_j(S_d) = \lambda_j + \sum_{k \in \mathcal{J} \setminus S} \rho_{kj} \Theta_k(S_d) \quad \forall d = 1, \dots, D, j \in \mathcal{J} \quad (7.7)$$

$$\lambda_j \geq 0 \quad \forall j \in \mathcal{J} \quad (7.8)$$

---

<sup>29</sup> Für alle Produkte  $j \in \mathcal{J} \setminus S$  gilt der Zusammenhang  $P_j(S) = \Theta_j(S)$  nicht, da für diese Produkte gemäß Axiom (3.9), das in Abschnitt 3.4.1 vorgestellt wurde,  $P_j(S) = 0$  ist.

$$\rho_{jk} \geq 0 \quad \forall j, k \in \mathcal{J} \quad (7.9)$$

Die Nebenbedingungen (7.5) und (7.6) stellen sicher, dass sich die Wahrscheinlichkeiten für den initialen Zustand und die Übergangswahrscheinlichkeiten, ausgehend von jedem Produkt  $j$ , auf den Wert 1 summieren. Für das Schätzproblem erweist sich die Tatsache als problematisch, dass kein Ausdruck in geschlossener Form für die Kaufwahrscheinlichkeiten  $P_{j_d}(S_d)$  in die Zielfunktion eingesetzt werden kann und deren Struktur somit ungeeignet für die numerische Maximierung ist. Neben der großen Anzahl von Parametern ist dies der Hauptgrund für die hohe Komplexität dieses Schätzproblems.

Allerdings kann die Log-Likelihood-Funktion vereinfacht werden, wenn zusätzliche Informationen, im Kontext des EM-Verfahrens die unbekannten Daten  $\mathbf{z}$ , zu jeder Transaktion als verfügbar angenommen werden (vgl. Simsek und Topaloglu (2018, S. 751) für die folgenden Überlegungen). Dabei handelt es sich zum einen um das Produkt  $j \in \mathcal{J}$ , das der für eine Transaktion verantwortliche Kunde gegenüber allen anderen Produkten bevorzugt. Diese Information wird für alle Transaktionen  $d$  durch die Vektoren  $\mathbf{\Lambda}^d = (\Lambda_1^d, \dots, \Lambda_n^d)$  gespeichert, wobei  $\Lambda_j^d = 1$  für das global präferierte Produkt  $j$  und  $\Lambda_k^d = 0$  für alle anderen Produkte  $k \in \mathcal{J} \setminus \{j\}$  gilt. Zum anderen wird angenommen, dass bekannt ist, wie oft im Laufe des Entscheidungsprozesses des Kunden in einer Transaktion  $d$  ein Zustandsübergang von  $X_t = j$  nach  $X_{t+1} = k$  stattgefunden hat. Die Matrix  $\mathbf{P}^d = (p_{jk}^d)_{j,k \in \mathcal{J}}$  speichert diese Daten für jede Transaktion  $d$ . Die Werte  $(\mathbf{\Lambda}^d, \mathbf{P}^d)$  für alle  $d$  können deswegen nicht direkt dem Trainingsdatensatz entnommen werden, weil die Zustände  $X_t$  mit  $t = 1, \dots, T - 1$  nicht beobachtet werden können (vgl. Gupta und Hsu (2017, S. 3)). Aus den Transaktionsdaten gehen jedoch indirekt Informationen über die unbekannten Daten hervor, konkret der letzte Zustand der jeweiligen Markov-Kette  $X_T = j_d$  und die absorbierenden Zustände  $X_t = j \in S_d$ . Ist eine Parameterschätzung  $(\boldsymbol{\lambda}, \boldsymbol{\rho})$  gegeben, kann nun die Wahrscheinlichkeit dafür formuliert werden, dass der Entscheidungsprozess eines Kunden, der ausgehend von seinem global präferierten Produkt  $k$  zum Kauf eines Produkts  $j_d$  führt, einem bestimmten Pfad  $X_1 = k, X_2 = k_1, \dots, X_{T-1} = j, X_T = j_d$  in der Markov-Kette entspricht:

$$P(X_1 = k, X_2 = k_1, \dots, X_{T-1} = j, X_T = j_d) = \lambda_j \cdot \rho_{kk_1} \cdot \dots \cdot \rho_{jj_d} \quad (7.10)$$

Die Beziehung (7.10) kann genutzt werden, um die kombinierte Likelihood  $L^{EM}(\boldsymbol{\lambda}, \boldsymbol{\rho})$  der Kaufentscheidungen der Kunden in allen Transaktionen  $d$  unter Verwendung von  $(\mathbf{\Lambda}^d, \mathbf{P}^d)$  zu formulieren:

$$L^{EM}(\boldsymbol{\lambda}, \boldsymbol{\rho}) = \prod_{d=1}^D \left( \prod_{j \in \mathcal{J}} \lambda_j^{\Lambda_j^d} \cdot \prod_{j,k \in \mathcal{J}} \rho_{jk}^{P_{jk}^d} \right) \quad (7.11)$$

Wird die Funktion  $L^{EM}(\boldsymbol{\lambda}, \boldsymbol{\rho})$  logarithmiert, ergibt sich eine zu (7.4) äquivalente Log-Likelihood-Funktion  $LL^{EM}(\boldsymbol{\lambda}, \boldsymbol{\rho})$ , unter der Annahme, dass die Daten  $(\boldsymbol{\Lambda}^d, \mathbf{P}^d)$  für alle Transaktionen  $d$  bekannt sind:

$$LL^{EM}(\boldsymbol{\lambda}, \boldsymbol{\rho}) = \sum_{d=1}^D \sum_{j \in \mathcal{J}} \Lambda_j^d \ln(\lambda_j) + \sum_{d=1}^D \sum_{j,k \in \mathcal{J}} P_{jk}^d \ln(\rho_{jk}) \quad (7.12)$$

Diese transformierte Log-Likelihood-Funktion ist global konkav und liegt in geschlossener Form vor. Mit einer komplexen Log-Likelihood-Funktion, die sich durch die Einbeziehung nur indirekt beobachtbarer, unbekannter Daten  $\mathbf{z}$  vereinfachen lässt, erfüllt das Schätzproblem des Markov-Chain-Modells genau die in Abschnitt 3.6.3 beschriebenen Anwendungsvoraussetzungen für ein EM-Verfahren. Die Funktion  $LL^{EM}(\boldsymbol{\lambda}, \boldsymbol{\rho})$  bildet daher die Grundlage für das EM-Verfahren von Simsek und Topaloglu (2018). Der Ansatz, den das Verfahren verfolgt, weicht jedoch in einem entscheidenden Aspekt von dem in Abschnitt 3.6.3 vorgestellten Grundkonzept eines EM-Verfahrens ab: Anstelle der bedingten Dichtefunktionen der unbekannten Daten  $h(\Lambda_j^d | j_d, S_d, \boldsymbol{\lambda}^i, \boldsymbol{\rho}^i)$  bzw.  $h(P_{jk}^d | j_d, S_d, \boldsymbol{\lambda}^i, \boldsymbol{\rho}^i)$ , können für die Formulierung der erwarteten Log-Likelihood-Funktion  $ELL(\boldsymbol{\lambda}, \boldsymbol{\rho} | \boldsymbol{\lambda}^i, \boldsymbol{\rho}^i)$  auf Basis von (7.12) direkt die bedingten Erwartungswerte der unbekannten Daten  $E(\Lambda_j^d | j_d, S_d, \boldsymbol{\lambda}^i, \boldsymbol{\rho}^i)$  bzw.  $E(P_{jk}^d | j_d, S_d, \boldsymbol{\lambda}^i, \boldsymbol{\rho}^i)$  verwendet werden. Diese Erwartungswerte werden für  $\Lambda_j^d$  bzw.  $P_{jk}^d$  in (7.12) eingesetzt, da die tatsächlichen Werte der unbekannten Daten nicht beobachtet werden können.

Der Ablauf des Verfahrens wird nun zunächst kurz zusammengefasst, bevor der E- und der M-Schritt genauer erläutert werden (vgl. Simsek und Topaloglu (2018, S. 751f)).

Nach der Festlegung beliebiger Startwerte  $(\boldsymbol{\lambda}^1, \boldsymbol{\rho}^1)$  im Initialisierungsschritt werden im E-Schritt jeder Iteration  $i = 1, \dots, I$  die aktuellen bedingten Erwartungswerte der unbekannten Daten  $E(\Lambda^d | j_d, S_d, \boldsymbol{\lambda}^i, \boldsymbol{\rho}^i)$  und  $E(\mathbf{P}^d | j_d, S_d, \boldsymbol{\lambda}^i, \boldsymbol{\rho}^i)$  für alle Transaktionen  $d$  bestimmt. Der nachfolgende M-Schritt umfasst die Maximierung der erwarteten Log-Likelihood-Funktion  $ELL(\boldsymbol{\lambda}, \boldsymbol{\rho} | \boldsymbol{\lambda}^i, \boldsymbol{\rho}^i)$  in Abhängigkeit der zuvor bestimmten bedingten Erwartungswerte der unbekannten Daten. Daraus ergibt sich die verbessernde Parameterschätzung  $(\boldsymbol{\lambda}^{i+1}, \boldsymbol{\rho}^{i+1})$ . E-Schritt und M-Schritt werden so lange wiederholt, bis das Verfahren konvergiert.

Die Formel für die Berechnung von  $E(\Lambda^d | j_d, S_d, \boldsymbol{\lambda}^i, \boldsymbol{\rho}^i)$  kann aus dem allgemeinen



Ausdruck (3.23) hergeleitet werden, der in Abschnitt 3.6.3 beschrieben wurde (vgl. Simsek und Topaloglu (2018, S. 752f) für die folgenden Überlegungen). Übertragen auf das hier betrachtete Schätzproblem ergibt sich für alle Transaktionen  $d$  und Produkte  $j$  zunächst die bedingte Dichtefunktion  $h(\Lambda_j^d | j_d, S_d, \lambda^i, \rho^i)$  gemäß:

$$h(\Lambda_j^d | j_d, S_d, \lambda^i, \rho^i) = \frac{P(j_d | S_d, \Lambda_j^d, \lambda^i, \rho^i) f(\Lambda_j^d | \lambda^i, \rho^i)}{P(j_d | S_d, \lambda^i, \rho^i)} \quad (7.13)$$

Um daraus den Wert  $E(\Lambda_j^d | j_d, S_d, \lambda^i, \rho^i)$  herleiten zu können, wird ausgenutzt, dass die binären Parameter  $\Lambda_j^d$  Bernoulli-verteilt sind und deren bedingter Erwartungswert lediglich von  $h(1 | j_d, S_d, \lambda^i, \rho^i)$  abhängt:

$$\begin{aligned} E(\Lambda_j^d | j_d, S_d, \lambda^i, \rho^i) &= 1 \cdot h(1 | j_d, S_d, \lambda^i, \rho^i) + 0 \cdot h(0 | j_d, S_d, \lambda^i, \rho^i) \\ &= \frac{P(j_d | S_d, \Lambda_j^d = 1, \lambda^i, \rho^i) P(\Lambda_j^d = 1 | \lambda^i, \rho^i)}{P(j_d | S_d, \lambda^i, \rho^i)} \end{aligned} \quad (7.14)$$

Der im Zähler dieses Ausdrucks enthaltene Wert  $P(\Lambda_j^d = 1 | \lambda^i, \rho^i)$  gibt an, wie hoch die Wahrscheinlichkeit dafür ist, dass ein ankommender Kunde, bei gegebener aktueller Parameterschätzung  $(\lambda^i, \rho^i)$ , das Produkt  $j$  global bevorzugt. Daher ist  $P(\Lambda_j^d = 1 | \lambda^i, \rho^i) = \lambda_j^i$ . Die Wahrscheinlichkeit  $P(j_d | S_d, \lambda^i, \rho^i)$  im Nenner von (7.14) entspricht der unbedingten Kaufwahrscheinlichkeit von Produkt  $j_d$  in Transaktion  $d$  und Iteration  $i$  und ist demzufolge durch den Parameter  $\Theta_{j_d}^i(S_d)$  gegeben. Diese Parameter werden durch das Lösen des Gleichungssystems (7.2), unter Verwendung der aktuellen Modellparameterschätzung  $(\lambda^i, \rho^i)$  ermittelt.

Der Parameter  $P(j_d | S_d, \Lambda_j^d = 1, \lambda^i, \rho^i) = \Psi_{j_d}^i(j, S_d)$  repräsentiert die bedingte Kaufwahrscheinlichkeit des in Transaktion  $d$  gekauften Produkts  $j_d$  in Iteration  $i$ , unter der Voraussetzung, dass der für Transaktion  $d$  verantwortliche Kunde Produkt  $j$  gegenüber allen anderen Produkten bevorzugt.<sup>30</sup> Für alle  $j \in S_d$  ist deren Berechnung mithilfe einer Indikatorfunktion  $\mathbf{I}$  möglich. Wenn  $j$  und  $j_d$  identisch sind, findet der Kauf von  $j_d$  sicher statt, da  $j_d$  von dem in Transaktion  $d$  ankommenden Kunden global bevorzugt wird. Wenn  $j \neq j_d$  ist, findet der Kauf keinesfalls statt, da sich der Kunde in Transaktion  $d$  nicht für  $j_d$  sondern

---

<sup>30</sup> Um die folgenden Berechnungen kompakter darstellen zu können, werden die Parameter  $\Psi_{j_d}^i(j, S_d)$  anstelle von  $P(j_d | S_d, \Lambda_j^d = 1, \lambda^i, \rho^i)$  verwendet.

das ebenfalls angebotene und global bevorzugte Produkt  $j$  entscheiden würde. Daher gilt in jeder Iteration  $i$  für alle Produkte  $j \in S_d$  und Transaktionen  $d$ :

$$\Psi_{j_d}^i(j, S_d) = \mathbf{I}(j = j_d) \quad (7.15)$$

Für die nicht angebotenen Produkte  $j \in \mathcal{J} \setminus S_d$  kann die bedingte Kaufwahrscheinlichkeit nur durch Lösen des folgenden Gleichungssystems berechnet werden:

$$\Psi_{j_d}^i(j, S_d) = \rho_{jj_d}^i + \sum_{k \in \mathcal{J} \setminus S_d} \rho_{jk}^i \Psi_{j_d}^i(k, S_d), \forall j \in \mathcal{J} \setminus S_d \quad (7.16)$$

Die rechte Seite dieser Gleichungen bildet die möglichen Pfade durch die Markov-Kette ab, über die ein Kunde von seinem initialen Zustand  $X_1 = j$ , der dem global bevorzugten Produkt  $j$  entspricht, zum absorbierenden Zustand  $X_T = j_d$  gelangen kann. Einerseits kann ein direkter Übergang von  $X_1 = j$  zu  $X_2 = j_d$  mit  $T = 2$  stattfinden. Die Wahrscheinlichkeit hierfür ist durch  $\rho_{jj_d}^i$  gegeben. Andererseits besteht die Möglichkeit, dass zuvor, mit der Wahrscheinlichkeit  $\rho_{jk}^i$  ein Übergang zu einem beliebigen, nicht absorbierenden dritten Zustand  $k$  erfolgt. Aufgrund der Gedächtnislosigkeit des Markov-Prozesses unterliegt der weitere Verlauf des Pfades, ausgehend von  $k$ , den gleichen Wahrscheinlichkeiten, wie ein Pfad, der in diesem Zustand beginnt. Der Kunde verhält sich ab diesem Zeitpunkt folglich wie ein Kunde, der  $k$  gegenüber allen anderen Produkten bevorzugt und die Kaufwahrscheinlichkeit für  $j_d$  wird durch  $\Psi_{j_d}^i(k, S_d)$  beschrieben. Um in jeder Iteration  $i$  sämtliche Wahrscheinlichkeiten  $\Psi_{j_d}^i(j, S_d)$  für alle  $j \in \mathcal{J}$  und alle Transaktionen  $d$  ermitteln zu können, ist somit die Lösung der Indikatorfunktionen (7.15) für alle  $j \in S_d$  sowie der Gleichungssysteme nötig, die sich aus der Formulierung der Gleichung (7.16) für alle  $j \in \mathcal{J} \setminus S_d$  ergeben.

Die Erwartungswerte der unbekannten Daten  $\Lambda_j^d$  in Iteration  $i$  können somit durch folgenden Ausdruck für alle Produkte  $j$  und Transaktionen  $d$  berechnet werden:

$$E(\Lambda_j^d | j_d, S_d, \boldsymbol{\lambda}^i, \boldsymbol{\rho}^i) = \frac{P(j_d | S_d, \Lambda_j^d = 1, \boldsymbol{\lambda}^i, \boldsymbol{\rho}^i) P(\Lambda_j^d = 1 | \boldsymbol{\lambda}^i, \boldsymbol{\rho}^i)}{P(j_d | S_d, \boldsymbol{\lambda}^i, \boldsymbol{\rho}^i)} = \frac{\Psi_{j_d}^i(j, S_d) \lambda_j^i}{\Theta_{j_d}^i(S_d)} \quad (7.17)$$

Die Herleitung des Ausdrucks für  $E(P_{jk}^d | j_d, S_d, \boldsymbol{\lambda}^i, \boldsymbol{\rho}^i)$ , d. h. die erwartete Anzahl von Zustandsübergängen von  $X_t = j$  nach  $X_{t+1} = k$ , für alle Produkte  $j, k \in S_d$  und Transaktionen  $d$  ist anhand ähnlicher, wenngleich umfangreicherer Überlegungen möglich (vgl. Simsek

und Topaloglu (2018, S. 753f) für die folgenden Überlegungen).<sup>31</sup> Ausgehend von einem Zustand  $X_t = j$  mit  $j \in S_d$  findet zu keinem Zeitpunkt  $t$  ein Zustandsübergang statt, da dieser Zustand per Definition absorbierend ist. Somit ist  $E(P_{jk}^d | j_d, S_d, \lambda^i, \rho^i) = 0$  für alle Produkte  $j \in S_d, k \in \mathcal{J}$  und Transaktionen  $d$ . Für die nicht absorbierenden Produkte  $j \in \mathcal{J} \setminus S_d$  gilt folgender Ausdruck für alle Produkte  $k \in \mathcal{J}$ , Transaktionen  $d$  und Iterationen  $i$ :

$$E(P_{jk}^d | j_d, S_d, \lambda^i, \rho^i) = \frac{\Psi_{jd}^i(k, S_d) \rho_{jk}^i}{\Theta_{jd}^i(S_d)} \Theta_j^i(S_d) \quad (7.18)$$

Auch die bedingten Erwartungswerte der unbekannten Daten des Typs  $P_{jk}^d$  sind somit lediglich abhängig von den bedingten Kaufwahrscheinlichkeiten  $\Psi_{jd}^i(k, S_d)$  und den unbedingten Kaufwahrscheinlichkeiten  $\Theta_j^i(S_d)$  bzw. erwarteten Besuchen  $\Theta_j^i(S_d)$  des mit Produkt  $j$  korrespondierenden Zustands. Diese Inputparameter werden bereits für die Ermittlung der bedingten Erwartungswerte  $E(\Lambda_k^d | j_d, S_d, \lambda^i, \rho^i)$  benötigt und müssen daher nicht erneut berechnet werden. Auch darüber hinaus sind keine weiteren Berechnungen erforderlich, denn die Werte von  $\rho_{jk}^i$  können der aktuellen Modellparameterschätzung  $(\lambda^i, \rho^i)$  entnommen werden. Zusammenfassend werden im E-Schritt jeder Iteration  $i$  des EM-Verfahrens von Simsek und Topaloglu (2018) die folgenden Rechenschritte für jede Transaktion  $d$  durchgeführt:

- Aufstellen und Lösen des Gleichungssystems (7.2), um die unbedingten Kaufwahrscheinlichkeiten bzw. erwarteten Besuche  $\Theta_j^i(S_d)$  für alle Produkte  $j \in \mathcal{J}$  zu berechnen.
- Lösen von (7.15), um die bedingten Kaufwahrscheinlichkeiten  $\Psi_{jd}^i(j, S_d)$  für die im Offerset  $S_d$  enthaltenen Produkte  $j \in S_d$  zu berechnen.
- Aufstellen und Lösen des Gleichungssystems (7.16), um die bedingten Kaufwahrscheinlichkeiten  $\Psi_{jd}^i(j, S_d)$  für die nicht im Offerset  $S_d$  enthaltenen Produkte  $j \in \mathcal{J} \setminus S_d$  zu berechnen.
- Berechnen der bedingten Erwartungswerte  $E(\Lambda_j^d | j_d, S_d, \lambda^i, \rho^i)$  der unbekannten Daten des Typs  $\Lambda_j^d$  gemäß (7.17) für alle Produkte  $j \in \mathcal{J}$ .

---

<sup>31</sup> Vgl. Simsek und Topaloglu (2018, S. 753f) für eine ausführliche Herleitung, auf deren Darstellung im Rahmen dieser Arbeit verzichtet wird.

- Berechnen der bedingten Erwartungswerte  $E(P_{jk}^d | j_d, S_d, \lambda^i, \rho^i)$  der unbekannten Daten des Typs  $P_{jk}^d$  gemäß (7.18) ausgehend von allen nicht im Offersset  $S_d$  enthaltenen Produkten  $j \in \mathcal{J} \setminus S_d$  für alle Produkte  $k \in \mathcal{J}$ .
- Setzen der bedingten Erwartungswerte  $E(P_{jk}^d | j_d, S_d, \lambda^i, \rho^i)$  der unbekannten Daten des Typs  $P_{jk}^d$  ausgehend von allen im Offersset  $S_d$  enthaltenen Produkten  $j \in S_d$  für alle Produkte  $k \in \mathcal{J}$  auf den Wert 0.

Im M-Schritt jeder Iteration  $i$  des Verfahrens wird die transformierte Log-Likelihood-Funktion (7.10) betrachtet, wobei die unbekannten Daten  $\Lambda_j^d$  und  $P_{jk}^d$  durch ihre bedingten Erwartungswerte  $E(\Lambda_j^d | j_d, S_d, \lambda^i, \rho^i)$  und  $E(P_{jk}^d | j_d, S_d, \lambda^i, \rho^i)$  ersetzt werden (vgl. Simsek und Topaloglu (2018, S. 754)). Dadurch ergibt sich die erwartete Log-Likelihood-Funktion:

$$\begin{aligned}
 ELL(\lambda, \rho | \lambda^i, \rho^i) = & \sum_{d=1}^D \sum_{k \in \mathcal{J}} E(\Lambda_k^d | j_d, S_d, \lambda^i, \rho^i) \ln(\lambda_k) \\
 & + \sum_{d=1}^D \sum_{j, k \in \mathcal{J}} E(P_{jk}^d | j_d, S_d, \lambda^i, \rho^i) \ln(\rho_{jk})
 \end{aligned} \tag{7.19}$$

Um die aktualisierte Parameterschätzung  $(\lambda^{i+1}, \rho^{i+1})$  zu bestimmen, wird diese Funktion maximiert, wobei die Nebenbedingungen (7.5), (7.6), (7.8) und (7.9) zu berücksichtigen sind. Das damit verbundene Optimierungsproblem ist separabel und kann in ein Teilproblem, aus dessen optimaler Lösung  $\lambda^{i+1}$  resultiert, sowie  $|\mathcal{J}|$  weitere Teilprobleme, die zur Bestimmung der Zeilen  $j = 0, \dots, n$  der Matrix  $\rho^{i+1}$  gelöst werden müssen, zerlegt werden. Das erstgenannte Teilproblem kann wie folgt formuliert werden:

$$\text{Max} \sum_{d=1}^D \sum_{j \in \mathcal{J}} E(\Lambda_j^d | j_d, S_d, \lambda^i, \rho^i) \ln(\lambda_j) \tag{7.20}$$

unter den Nebenbedingungen

$$\sum_{j \in \mathcal{J}} \lambda_j = 1 \tag{7.21}$$

$$\lambda_j \geq 0 \quad \forall j \in \mathcal{J} \tag{7.22}$$

Die optimale Lösung dieses Problems kann mithilfe eines Ausdrucks in geschlossener Form bestimmt werden:

$$\lambda_j^{i+1} = \frac{\sum_{d=1}^D E(\Lambda_j^d | j_d, S_d, \lambda^i, \rho^i)}{\sum_{d=1}^D \sum_{k \in \mathcal{J}} E(\Lambda_k^d | j_d, S_d, \lambda^i, \rho^i)} \quad \forall j \in \mathcal{J} \quad (7.23)$$

Um die neue Schätzung der Übergangsmatrix  $\rho^{i+1}$  zu erhalten, muss die optimale Lösung des folgenden Optimierungsproblems für alle Produkte  $j \in \mathcal{J}$  berechnet werden:

$$\text{Max} \sum_{d=1}^D \sum_{k \in \mathcal{J}} E(P_{jk}^d | j_d, S_d, \lambda^i, \rho^i) \ln(\rho_{jk}) \quad (7.24)$$

unter den Nebenbedingungen

$$\sum_{k \in \mathcal{J}} \rho_{jk} = 1 \quad (7.25)$$

$$\rho_{jk} \geq 0 \quad \forall k \in \mathcal{J} \quad (7.26)$$

Da diese Teilprobleme und das bereits vorgestellte Teilproblem für die Bestimmung von  $\lambda^{i+1}$  strukturell identisch sind, ergibt sich die optimale Lösung für das mit Produkt  $j \in \mathcal{J}$  korrespondierende Teilproblem gemäß:

$$\rho_{jk}^{i+1} = \frac{\sum_{d=1}^D E(P_{jk}^d | j_d, S_d, \lambda^i, \rho^i)}{\sum_{d=1}^D \sum_{j' \in \mathcal{J}} E(P_{jj'}^d | j_d, S_d, \lambda^i, \rho^i)} \quad \forall k \in \mathcal{J} \quad (7.27)$$

Simsek und Topaloglu (2018) zeigen, dass  $LL(\lambda^{i+1}, \rho^{i+1}) \geq LL(\lambda^i, \rho^i)$  nur dann für alle Iterationen  $i$  gilt, wenn bei der Lösung aller Teilprobleme die Nebenbedingungen (7.22) durch  $\lambda_j \geq \epsilon, \forall j \in \mathcal{J}$  und die Nebenbedingungen (7.26) durch  $\rho_{jk} \geq \epsilon, \forall k \in \mathcal{J}$ , jeweils mit  $\epsilon > 0$ , ersetzt werden.<sup>32</sup> Ohne diese Nebenbedingungen kann der Fall auftreten, dass  $(\lambda^{i+1}, \rho^{i+1})$  verglichen mit  $(\lambda^i, \rho^i)$  eine verschlechternde Parameterschätzung darstellt. Generell kann nicht garantiert werden, dass das Verfahren in einem globalen Optimum von  $LL(\lambda, \rho)$  konvergiert. Je nach gewähltem initialen Parameterset  $(\lambda^1, \rho^1)$  können somit unterschiedliche lokale Optima erreicht werden. Abschließend ist der gesamte Verfahrensablauf in Abbildung 15 dargestellt.

---

<sup>32</sup> Auf die Erläuterung des Lösungsverfahrens für diese modifizierten Teilprobleme wird in dieser Arbeit verzichtet. Eine ausführliche Betrachtung findet sich in Simsek und Topaloglu (2018, Online Appendix A8).

```

i := 1, konv := False
for j ∈ J:
     $\lambda_j^1 := \frac{1}{|J|}$ 
     $\rho_{jj}^1 := 0$ 
    for k ∈ J \ {j}:
         $\rho_{jk}^1 := \frac{1}{|J|-1}$ 
while konv = False:
    for d = 1, ..., D:
        Berechne  $\Theta_j^i(S_d)$  für alle j ∈ J durch Lösen des Gleichungssystems (7.2)
        for j ∈ Sd:
             $\Psi_{j_d}^i(j, S_d) := I(j = j_d)$ 
            Berechne  $\Psi_{j_d}^i(j, S_d)$  für alle j ∈ J \ Sd durch Lösen des
            Gleichungssystems (7.16)
        for j ∈ J:
            Berechne  $E(\Lambda_j^d | j_d, S_d, \lambda^i, \rho^i)$  gemäß (7.17)
            for k ∈ J:
                if j ∉ Sd:
                    Berechne  $E(P_{jk}^d | j_d, S_d, \lambda^i, \rho^i)$  gemäß (7.18)
                else:
                     $E(P_{jk}^d | j_d, S_d, \lambda^i, \rho^i) := 0$ 
        for j ∈ J:
            Berechne  $\lambda_j^{i+1}$  gemäß (7.23)
        for k ∈ J:
            Berechne  $\rho_{jk}^{i+1}$  gemäß (7.27)
        if  $|\lambda_j^{i+1} - \lambda_j^i| < K \wedge |\rho_{jk}^{i+1} - \rho_{jk}^i| < K \forall j, k \in J$ :
            konv := True
        i = i + 1

```

Abbildung 15: Ablauf des EM-Verfahrens von Simsek und Topaloglu (2018)

Als Startwerte der Modellparameter  $(\lambda^1, \rho^1)$  können beliebige nichtnegative Werte gewählt werden, sofern die Bedingungen  $\sum_{j \in \mathcal{J}} \lambda_j^1 = 1$  und  $\sum_{k \in \mathcal{J}} \rho_{jk}^1 = 1, \forall j \in \mathcal{J}$  erfüllt sind. In Abbildung 15 ist daher lediglich eine mögliche Strategie für die Definition zulässiger Startparameter  $(\lambda^1, \rho^1)$  dargestellt. Die erste for-Schleife innerhalb jeder Iteration beinhaltet die Berechnungen des E-Schritts. Auf diese folgt eine zweite for-Schleife, die den M-Schritt repräsentiert. Da Simsek und Topaloglu (2018) kein konkretes Konvergenzkriterium angeben, wird, wie bereits im EM-Verfahren für das LC-MNL-Modell, die Veränderung der Modellparameterschätzung zur Prüfung der Konvergenz verwendet. Liegt der Betrag der Veränderung aller Modellparameter in einer Iteration  $i + 1$  verglichen mit der Schätzung der vorherigen Iteration  $i$  unter einer festgelegten Schranke  $K$ , wird das Verfahren abgebrochen.

### 7.3 Exemplarische Schätzung des Markov-Chain-Modells

Die Beispielinstantz, für die ein Markov-Chain-Modell geschätzt werden soll, umfasst  $|\mathcal{J}| = 5$  Produkte. Da dieser Wert das Parameterset des Modells determiniert und keine Spezifikation durch die Airline oder während des Schätzverfahrens erfolgt, enthält das Modell folgende Parameter:

$$\lambda = \begin{pmatrix} \lambda_0 \\ \vdots \\ \lambda_4 \end{pmatrix} \text{ und } \rho = \begin{pmatrix} \rho_{00} & \cdots & \rho_{04} \\ \vdots & \ddots & \vdots \\ \rho_{40} & \cdots & \rho_{44} \end{pmatrix} \quad (7.28)$$

Dem in Abbildung 15 beschriebenen Ansatz folgend, werden in der Initialisierung des EM-Verfahrens die Startwerte  $\lambda_j^1 = 0,2$  für alle  $j = 0, \dots, 4$  sowie  $\rho_{jk}^1 = 0,25$  für alle  $j, k = 0, \dots, 4$  mit  $j \neq k$  und  $\rho_{jj}^1 = 0$  für alle  $j = 0, \dots, 4$  gewählt.

Die Berechnungen des E-Schrittes der ersten Iteration werden nun für die Transaktionen  $d = 1, 2$  ausführlich veranschaulicht. Für die übrigen Transaktionen  $d = 3, \dots, 10$  werden lediglich die Ergebnisse angegeben (vgl. Tabelle 16), da diese in gleicher Weise ermittelt werden. Bevor die bedingten Erwartungswerte der unbekannten Daten für Transaktion  $d = 1$  berechnet werden können, müssen die Parameter  $\Psi_1^1(j, S_1)$  und  $\Theta_j^1(S_1)$  auf Basis von  $\lambda^1$  und  $\rho^1$  ermittelt werden. Das Gleichungssystem zur Berechnung der Parameter  $\Theta_j^1(S_1)$  umfasst eine Gleichung für jedes Produkt  $j$  und daher insgesamt  $|\mathcal{J}| = 5$  Gleichungen. Da  $\mathcal{J} \setminus S_1 = \emptyset$  ist, fällt der zweite Summand in allen Gleichungen weg und es ergibt sich das Gleichungssystem:

$$\theta_0^1(S_1) = \lambda_0^1 \quad (7.29)$$

$$\theta_1^1(S_1) = \lambda_1^1 \quad (7.30)$$

$$\theta_2^1(S_1) = \lambda_2^1 \quad (7.31)$$

$$\theta_3^1(S_1) = \lambda_3^1 \quad (7.32)$$

$$\theta_4^1(S_1) = \lambda_4^1 \quad (7.33)$$

Die Lösung dieses Gleichungssystems lautet trivialerweise  $\theta_j^1(S_1) = \lambda_j^1 = 0,2$  für alle  $j = 0, \dots, 4$ . Auch die Berechnung der bedingten Kaufwahrscheinlichkeiten ist für die erste Transaktion sehr einfach möglich. Für alle Produkte  $j = 0, \dots, 4$  kann der Ausdruck  $\Psi_1^1(j, S_1) := \mathbf{I}(j = 1)$  verwendet werden, da  $S_1 = \mathcal{J}$  gilt. Es ergibt sich  $\Psi_1^1(1, S_1) = 1$  und  $\Psi_1^1(j, S_1) = 0$  für die Produkte  $j = 0, 2, 3, 4$ .

Nun werden die bedingten Erwartungswerte  $E(\Lambda_j^1 | 1, S_1, \boldsymbol{\lambda}^1, \boldsymbol{\rho}^1)$  bzw.  $E(P_{jk}^1 | 1, S_1, \boldsymbol{\lambda}^1, \boldsymbol{\rho}^1)$  der unbekannten Daten des Typs  $\Lambda_j^1$  bzw.  $P_{jk}^1$  berechnet. So ergibt sich beispielsweise für  $j = 1$  der Wert  $E(\Lambda_1^1 | 1, S_1, \boldsymbol{\lambda}^1, \boldsymbol{\rho}^1) = \frac{\Psi_1^1(1, S_1) \lambda_1^1}{\theta_1^1(S_1)} = \frac{1 \cdot 0,2}{0,2} = 1$ . Für die übrigen Werte ergibt sich aus analogen Berechnungen  $E(\Lambda_j^1 | 1, S_1, \boldsymbol{\lambda}^1, \boldsymbol{\rho}^1) = 0$  für  $j = 0, 2, 3, 4$ . Da alle Produkte in  $S_1$  enthalten sind, gilt  $E(P_{jk}^1 | 1, S_1, \boldsymbol{\lambda}^1, \boldsymbol{\rho}^1) = 0$  für alle  $j, k = 0, \dots, 4$ . Dieses Ergebnis ist intuitiv plausibel, da die Kunden aus einem Offerset, welches alle Produkte enthält, stets das Produkt wählen, welches sie global präferieren. Somit wird in dem zugrunde liegenden Entscheidungsprozess bereits zum Zeitpunkt  $t = 1$  in jedem Fall ein absorbierender Zustand erreicht und es finden keine Zustandsübergänge statt, da der Prozess in  $t = 1$  bereits endet. Es besteht daher keine Unsicherheit darüber, dass der für Transaktion  $d = 1$  verantwortliche Kunde das Produkt  $j = 1$  global bevorzugt.

Das Offerset  $S_2 = \{0, 3, 4\}$  enthält nur eine Teilmenge der Produkte, weswegen die Berechnungen für Transaktion  $d = 2$  im E-Schritt des EM-Verfahrens umfangreicher sind. Die rechte Seite jeder Gleichung des Gleichungssystems (7.2) zur Berechnung der unbedingten Kaufwahrscheinlichkeiten bzw. erwarteten Besuche  $\theta_j^1(S_2)$  besteht in diesem Fall aus drei Summanden, da  $\mathcal{J} \setminus S_2 = \{1, 2\}$  gilt:

$$\theta_0^1(S_2) = \lambda_0^1 + \rho_{10}^1 \theta_1^1(S_2) + \rho_{20}^1 \theta_2^1(S_2) \quad (7.34)$$



$$\Theta_1^1(S_2) = \lambda_1^1 + \rho_{11}^1 \Theta_1^1(S_2) + \rho_{21}^1 \Theta_2^1(S_2) \quad (7.35)$$

$$\Theta_2^1(S_2) = \lambda_2^1 + \rho_{12}^1 \Theta_1^1(S_2) + \rho_{22}^1 \Theta_2^1(S_2) \quad (7.36)$$

$$\Theta_3^1(S_2) = \lambda_3^1 + \rho_{13}^1 \Theta_1^1(S_2) + \rho_{23}^1 \Theta_2^1(S_2) \quad (7.37)$$

$$\Theta_4^1(S_2) = \lambda_4^1 + \rho_{14}^1 \Theta_1^1(S_2) + \rho_{24}^1 \Theta_2^1(S_2) \quad (7.38)$$

Durch Einsetzen der aktuellen Modellparameter und Lösen des Gleichungssystems ergeben sich die erwarteten Besuche  $\Theta_1^1(S_2) = \Theta_2^1(S_2) = 0,267$  sowie die unbedingten Kaufwahrscheinlichkeiten  $\Theta_0^1(S_2) = \Theta_3^1(S_2) = \Theta_4^1(S_2) = 0,333$ . Die mit den Produkten  $j = 1,2$  korrespondierenden, nicht absorbierenden Zustände werden im Erwartungswert jeweils 0,267 Mal besucht. Für die übrigen Parameter  $\Theta_j^1(S_2)$ , mit  $j = 0,3,4$ , liegt dieser Wert bei  $\frac{1}{3}$  und kann als unbedingte Kaufwahrscheinlichkeit interpretiert werden, da die Parameter mit den absorbierenden Zuständen der Markov-Kette für  $d = 2$  korrespondieren. Von den bedingten Kaufwahrscheinlichkeiten müssen für diese Markov-Kette die Werte  $\Psi_4^1(j, S_2)$  mit  $j = 1,2$  mithilfe eines Gleichungssystems berechnet werden.

$$\Psi_4^1(1, S_2) = \rho_{14}^1 + \rho_{11}^1 \Psi_4^1(1, S_2) + \rho_{12}^1 \Psi_4^1(2, S_2) \quad (7.39)$$

$$\Psi_4^1(2, S_2) = \rho_{24}^1 + \rho_{21}^1 \Psi_4^1(1, S_2) + \rho_{22}^1 \Psi_4^1(2, S_2) \quad (7.40)$$

Das Lösen des Gleichungssystems, nachdem die aktuelle Modellparameterschätzung eingesetzt wurde, ergibt in diesem Fall  $\Psi_4^1(1, S_2) = \Psi_4^1(2, S_2) = 0,333$ . Mit Ausdruck (7.15) können die restlichen Parameter dieses Typs  $\Psi_4^1(0, S_2) = \Psi_4^1(3, S_2) = 0$  und  $\Psi_4^1(4, S_2) = 1$  berechnet werden.

Wie bereits für  $d = 1$  gezeigt, besteht der nächste Verfahrensschritt in der Berechnung der bedingten Erwartungswerte  $E(\Lambda_j^2 | 4, S_2, \lambda^1, \rho^1)$  und  $E(P_{jk}^2 | 4, S_2, \lambda^1, \rho^1)$ . Die Werte für  $E(\Lambda_j^2 | 4, S_2, \lambda^1, \rho^1)$ , die über den Ausdruck (7.17) berechnet werden, sind in Tabelle 11 aufgelistet.

$j$	0	1	2	3	4
$E(\Lambda_j^2   4, S_2, \lambda^1, \rho^1)$	0	0,2	0,2	0	0,6

**Tabelle 11: Bedingte Erwartungswerte der unbekannten Daten des Typs  $\Lambda_j^2$  für Transaktion  $d = 2$  und Produkte  $j = 0, \dots, 4$**

Im Unterschied zu Transaktion  $d = 1$  können bzgl. der Markov-Kette zu Transaktion  $d = 2$  nicht alle bedingten Erwartungswerte  $E(P_{jk}^2 | 4, S_2, \lambda^1, \rho^1)$  auf den Wert 0 gesetzt werden. Möglich ist dies lediglich für Übergänge von den in  $S_2$  enthaltenen Produkten  $j = 0, 3, 4$  zu einem beliebigen Produkt  $k = 0, \dots, 4$ . Für  $j = 1, 2$  und  $k = 0, \dots, 4$  ist die Berechnung der bedingten Erwartungswerte gemäß  $E(P_{jk}^2 | 4, S_2, \lambda^1, \rho^1) = \frac{\Psi_4^1(j, S_2) \rho_{jk}^1}{\Theta_4^1(S_2)} \Theta_j^1(S_2)$  erforderlich. Beispielhaft ergibt sich für die erwartete Anzahl an Übergängen von Produkt  $j = 1$  zu  $k = 2$  der Wert  $E(P_{12}^2 | 4, S_2, \lambda^1, \rho^1) = \frac{0,333 \cdot 0,25}{0,333} \cdot 0,267 = 0,067$ . Die Ergebnisse der übrigen Berechnungen sind in Tabelle 12 dargestellt.

$k$	0	1	2	3	4
$E(P_{1k}^2   4, S_2, \lambda^1, \rho^1)$	0	0	0,067	0	0,2
$E(P_{2k}^2   4, S_2, \lambda^1, \rho^1)$	0	0,067	0	0	0,2

**Tabelle 12:** Bedingte Erwartungswerte der unbekannten Daten des Typs  $P_{jk}^2$  für Transaktion  $d = 2$  und Produkte  $j = 1, 2$  sowie  $k = 0, \dots, 4$

Im weiteren Verlauf des E-Schrittes werden die Rechenoperationen, die zu den bedingten Erwartungswerten der unbekannten Daten führen, für die Transaktionen  $d = 3, \dots, 10$  durchgeführt. Die Ergebnisse sind in Tabelle 16 aufgelistet.<sup>33</sup>

Anschließend beginnt der M-Schritt, in dem anhand dieser Werte die aktualisierte Modellparameterschätzung  $(\lambda^2, \rho^2)$  ermittelt wird. Die Elemente  $\lambda_j^2$  des Parametervektors  $\lambda^2$  werden mittels der Formel  $\lambda_j^2 = \frac{\sum_{d=1}^{10} E(\Lambda_j^d | j_d, S_d, \lambda^1, \rho^1)}{\sum_{d=1}^{10} \sum_{k \in J} E(\Lambda_k^d | j_d, S_d, \lambda^1, \rho^1)}$  für alle Produkte  $j = 0, \dots, 4$  berechnet. Im Zähler werden die bedingten Erwartungswerte bzgl. des Produkts  $j$  über alle Transaktionen  $d = 1, \dots, 10$  summiert. Den Nenner bildet die Summe aller bedingten Erwartungswerte des Typs  $E(\Lambda_k^d | j_d, S_d, \lambda^1, \rho^1)$ . Diese nimmt den Wert  $D = 10$  an, da  $\sum_{k \in J} E(\Lambda_k^d | j_d, S_d, \lambda^1, \rho^1) = 1$  für alle Transaktionen  $d$  gilt. Auf diese Weise ergibt sich  $\lambda^2 = (0,1; 0,3; 0,22; 0,1; 0,28)^T$ . Bereits diese erste Aktualisierung der Parameterschätzung lässt sich vor dem Hintergrund der tatsächlichen Kundenpräferenzen, auf Basis derer die Beispielinstantz generiert wurde, interpretieren. In der tatsächlichen Kundenpopulation bevorzugen 40% der Kunden das Produkt  $j = 1$  gegenüber allen anderen Produkten und 60%

<sup>33</sup> Der Übersichtlichkeit wegen werden in Tabelle 16 für jede Transaktion  $d$  die Parameter  $E(P_{jk}^d | j_d, S_d, \lambda^1, \rho^1)$  für Produkte  $j \in S_d$  nicht angegeben, da diese den Wert 0 annehmen.

das Produkt  $j = 4$ . Entsprechend liegt der Wert der Parameter  $\lambda_1^2$  und  $\lambda_4^2$  gegenüber den initialen Schätzungen  $\lambda_1^1$  und  $\lambda_4^1$  wesentlich höher. Der Schätzwert der Parameter  $\lambda_0$  und  $\lambda_3$  verringert sich dagegen deutlich, was angesichts dessen, dass sich die Produkte  $j = 0$  und  $j = 3$  am Ende der Präferenzlisten der beiden Kundensegmente befinden, ebenfalls der Erwartung entspricht.

Für die Berechnung der Elemente der  $j$ -ten Zeile der Parametermatrix  $\boldsymbol{\rho}^2$  wird ein strukturell gleiches Maximierungsproblem gelöst, wie für die Berechnung der Elemente von  $\boldsymbol{\lambda}^2$ . Der

Ausdruck für deren Berechnung  $\rho_{jk}^2 = \frac{\sum_{d=1}^{10} E(P_{jk}^d | j_d, S_d, \boldsymbol{\lambda}^1, \boldsymbol{\rho}^1)}{\sum_{d=1}^{10} \sum_{j' \in \mathcal{J}} E(P_{jj'}^d | j_d, S_d, \boldsymbol{\lambda}^1, \boldsymbol{\rho}^1)}$  für alle Produkte  $j =$

$0, \dots, 4$  ist daher ähnlich aufgebaut und wird anhand des Parameters  $\rho_{21}^2$  veranschaulicht. Der

Nenner  $\sum_{d=1}^{10} \sum_{j' \in \mathcal{J}} E(P_{2j'}^d | j_d, S_d, \boldsymbol{\lambda}^1, \boldsymbol{\rho}^1) = 1,467$  wird für diesen Parameter aus der

Summe der erwarteten Zustandsübergänge ausgehend von Produkt  $j = 2$  zu einem beliebigen Produkt  $j' = 0, \dots, 4$  in allen Transaktionen  $d$  errechnet. Um den Zähler

$\sum_{d=1}^{10} E(P_{21}^d | j_d, S_d, \boldsymbol{\lambda}^1, \boldsymbol{\rho}^1) = 0,267$  zu berechnen, müssen die erwarteten Zustandsübergänge von  $j = 2$  zu  $k = 1$  über alle Transaktionen  $d$  aufsummiert werden. Es ergibt sich

folglich die Schätzung  $\rho_{21}^2 = \frac{0,267}{1,467} = 0,182$ . Die vollständige Parametermatrix  $\boldsymbol{\rho}^2$  ist in Ta-

belle 13 dargestellt.<sup>34</sup>

---

<sup>34</sup> Da die Nichtkaufalternative in jedem Offersset  $S_d$  enthalten ist, gilt  $E(P_{0k}^d | j_d, S_d, \boldsymbol{\lambda}^i, \boldsymbol{\rho}^i) = 0$  für alle  $k = 0, \dots, 4$ . Dies führt dazu, dass der Nenner  $\sum_{d=1}^{10} \sum_{j' \in \mathcal{J}} E(P_{0j'}^d | j_d, S_d, \boldsymbol{\lambda}^1, \boldsymbol{\rho}^1) = 0$  beträgt und so bei der Berechnung der Parameter  $\rho_{0k}^2$  eine Division durch den Wert 0 stattfindet. Simsek und Topaloglu (2018) erläutern nicht, wie mit diesem Problem umzugehen ist. Daher wird  $\rho_{0k}^2 = 0$  für  $k = 1, \dots, 4$  und  $p_{00}^2 = 1$  gesetzt, um die Restriktion  $\sum_{k \in \mathcal{J}} \rho_{0k}^2 = 1$  zu erfüllen. Dies ist zulässig, weil auch in der Anwendung des Modells die Nichtkaufalternative in jedem Offersset enthalten sein muss, sodass keine Zustandsübergänge, ausgehend von der Nichtkaufalternative, stattfinden können und die Werte der 0-ten Zeile der Übergangsmatrix damit irrelevant für die Berechnung der Kaufwahrscheinlichkeiten sind.

$k$	0	1	2	3	4
$\rho_{0k}^2$	1	0	0	0	0
$\rho_{1k}^2$	0	0	0,583	0,083	0,333
$\rho_{2k}^2$	0,273	0,182	0	0,159	0,386
$\rho_{3k}^2$	0,143	0,333	0,310	0	0,214
$\rho_{4k}^2$	0,429	0,071	0,393	0,107	0

Tabelle 13: Aktualisierte Parameterschätzung  $\rho^2$ 

Auch bei der Betrachtung der aktualisierten Schätzwerte dieses Parametertyps zeigen sich die erwarteten Veränderungen bzgl. der initialen Werte. In der Beispielinstanz wählen beispielsweise die Privatreisenden, aufgrund ihrer Präferenzliste  $\sigma^1 = (4,2)$ , bei Nichtverfügbarkeit ihres global präferierten Produkts  $j = 4$  entweder das Produkt  $j = 2$  oder die Nichtkaufalternative  $j = 0$ . In der Übergangsmatrix  $\rho^2$  spiegelt sich dieser Zusammenhang in den hohen Werten der Übergangswahrscheinlichkeiten  $\rho_{40}^2$  und  $\rho_{42}^2$  wider. Dass es sich bei den Werten  $\lambda^2$  und  $\rho^2$  um eine verbessernde Parameterschätzung handelt, zeigt die Berechnung der Log-Likelihood-Werte  $LL(\lambda^1, \rho^1)$  und  $LL(\lambda^2, \rho^2)$ :

$$\begin{aligned}
 LL(\lambda^1, \rho^1) &= \sum_{d=1}^{10} \ln(\theta_{j_d}^1(S_d)) \\
 &= \ln(0,2) + \ln(0,333) + \ln(0,333) + \ln(0,333) + \ln(0,2) + \ln(0,333) \\
 &\quad + \ln(0,333) + \ln(0,333) + \ln(0,25) + \ln(0,5) = -11,89 \text{ und} \\
 LL(\lambda^2, \rho^2) &= -8,93
 \end{aligned} \tag{7.41}$$

Wird das Verfahren beispielsweise nach 50 Iterationen abgebrochen, ergibt sich folgende Lösung des Schätzproblems mit dem Zielfunktionswert  $LL(\lambda^{50}, \rho^{50}) = -5,59$ .

$k$	0	1	2	3	4
$\lambda_k^{50}$	0	0,493	0,028	0	0,479
$\rho_{0k}^{50}$	1	0	0	0	0
$\rho_{1k}^{50}$	0	0	1	0	0
$\rho_{2k}^{50}$	0,003	0	0	0	0,997
$\rho_{3k}^{50}$	0	0,464	0,536	0	0
$\rho_{4k}^{50}$	0,087	0	0,913	0	0

Tabelle 14: Parameterschätzung  $(\lambda^{50}, \rho^{50})$  in Iteration  $i = 50$ 

Verglichen mit den anderen in dieser Arbeit betrachteten Discrete-Choice-Modellen erzielt das Markov-Chain-Modell die höchste Abbildungsgenauigkeit, beinhaltet jedoch auch die größte Anzahl an Modellparametern. Bei der Betrachtung der Lösung fällt auf, dass die Übergangswahrscheinlichkeiten  $\rho_{jk}$  nur dann deutlich von 0 verschieden sind, wenn in mindestens einer der beiden den Trainingsdaten zugrunde liegenden Präferenzlisten  $\sigma_l$  mit  $l = 1, 2$  gilt:  $\sigma_l(j) = \sigma_l(k) - 1$ . Nicht auf diese Weise interpretierbar ist die Zeile  $j = 3$ , da sich dieses Produkt im irrelevanten Teil beider Präferenzlisten befindet und daher im tatsächlichen Entscheidungsprozess der Kunden keine Rolle spielt. Bemerkenswert ist darüber hinaus, dass gemäß der Schätzung jeweils ungefähr 50 % der Kunden die Produkte  $j = 1$  und  $j = 4$  bevorzugen. Diese Abweichung von der tatsächlichen Zusammensetzung der Kundenpopulation tritt auch bei der Schätzung des LC-MNL-Modells und des Rank-based-Modells auf. Daher ist anzunehmen, dass die verwendeten Trainingsdaten nicht umfangreich genug sind, um die Anteile der beiden Kundensegmente präzise schätzen zu können.

## 8 Diskussion und Ausblick

In dieser Arbeit wurde am Beispiel des Airline-RM gezeigt, dass die Integration des Kundenwahlverhaltens in die klassischen Optimierungsprobleme des RM deren Exaktheit wesentlich erhöhen kann. Dieses Ziel kann durch den Einsatz verschiedener Discrete-Choice-Modelle erreicht werden, die mit zunehmender Komplexität das Substitutionsverhalten der Nachfrager exakter abbilden können. So erhöht sich die Genauigkeit der berechneten Kaufwahrscheinlichkeiten, die als Inputdaten z. B. in das vorgestellte Assortment-Optimierungsproblem einfließen. Somit führt ein komplexeres Discrete-Choice-Modell letztendlich zu einer höheren Qualität der Lösung dieses RM-Problems. Die Lösung integrierter RM-Probleme wird dadurch jedoch aus zwei Gründen erschwert: Einerseits erhöht sich die Komplexität des nichtlinearen Maximum-Likelihood-Schätzproblems. Dies ist vor allem auf eine nicht global konkave Log-Likelihood-Funktion, zu beachtende Restriktionen und eine hohe Zahl von Parametern zurückzuführen. Andererseits beeinflusst die Wahl des Discrete-Choice-Modells auch Struktur und Komplexität des Assortment-Optimierungsproblems. Wird das MNL-Modell eingesetzt, so weist das Assortment-Optimierungsproblem die geringste Komplexität auf, denn für beliebige Werte des DAR enthält das optimale Offerset  $S$  neben der Nichtkaufalternative stets die  $k$  teuersten Produkte, wobei  $k = 1, \dots, n$  gilt. Um das optimale Offerset zu berechnen, genügt es folglich, die Zielfunktionswerte von  $n$  Offersets zu vergleichen (vgl. Talluri und van Ryzin (2004b)). Für das LC-MNL-Modell mit überlappenden Considerationsets ergibt sich dagegen ein NP-schweres Optimierungsproblem, das beispielsweise mithilfe eines exakten bzw. heuristischen Spaltengenerierungsverfahrens oder einer Greedy-Heuristik gelöst werden kann (vgl. Miranda Bront et al. (2009)). Das Assortment-Optimierungsproblem für das Rank-based-Modell ist ebenfalls NP-schwer. Eine exakte Lösung ist u. a. mithilfe des In-Out-Algorithmus möglich, darüber hinaus können verschiedene Heuristiken verwendet werden (vgl. Honhon et al. (2015)). Weniger komplex ist das Assortment-Optimierungsproblem, das sich bei der Anwendung des Markov-Chain-Modells ergibt. Dieses kann als lineares Programm formuliert werden, das in polynomieller Zeit lösbar ist (vgl. Feldman und Topaloglu (2017)).

Neben den Kriterien der Lösungsqualität und der Lösbarkeit wurden in dieser Arbeit weitere Aspekte behandelt, die bei der Modellauswahl zu berücksichtigen sind. Dazu gehören die Interpretierbarkeit der Modellparameter und die Möglichkeit die

Kaufwahrscheinlichkeiten neuer Produkte ohne erneute Parameterschätzung prognostizieren zu können. Beides ist nur bei der Verwendung eines RUM möglich. Ist der Anwender jedoch mangels Erfahrungen bzw. Vergangenheitsdaten nicht in der Lage eine fundierte Spezifikation durchzuführen, spricht dies für den Einsatz von RUT-kompatiblen Modellen wie dem Rank-based-Modell. Zudem muss die Menge der zur Verfügung stehenden Trainingsdaten berücksichtigt werden. Modelle mit einer großen Anzahl von Parametern, wie das Markov-Chain-Modell, sind aufgrund des Overfitting-Problems ungeeignet, wenn nur eine geringe Datenbasis für die Schätzung vorhanden ist. Empirische Studien zeigen, dass unter diesen Bedingungen u. a. das Rank-based-Modell gute Ergebnisse liefert (vgl. Berbeglia et al. (2018, S. 33)).

Die beschriebenen Kriterien stellen wichtige Leitlinien für die Modellauswahl dar. In der praktischen Anwendung sollte darüber hinaus eine empirische Validierung des gewählten Modells durchgeführt werden (vgl. Cascetta (2009, S. 530-536)). Mithilfe verschiedener Hypothesentests kann so geprüft werden, ob die Annahmen des Modells im konkreten Anwendungsfall problematisch sind oder die Parameterschätzung nicht signifikant ist. Simulationsstudien können ebenfalls eingesetzt werden, um die Qualität der Lösung des integrierten Assortment-Optimierungsproblems zu beurteilen, falls das Kundenwahlverhalten durch das gewählte Discrete-Choice-Modell beschrieben wird (vgl. Miranda Bront et al. (2009, S. 777f)).

Aktuell werden besonders die komplexen Discrete-Choice-Modelle weiterhin intensiv erforscht. So gibt es beispielsweise Bestrebungen die Parameter des Markov-Chain-Modells als Funktion von Produktattributen zu formulieren und dadurch die Gesamtzahl der zu schätzenden Parameter zu verringern (vgl. Simsek und Topaloglu (2018, S. 760)). Außerdem ist zu erwarten, dass einerseits immer effizientere Schätz- und Assortment-Optimierungsverfahren für die vergleichsweise etablierten und bereits gut erforschten Modelle, wie für das Rank-based-Modell, entwickelt werden. Andererseits existieren noch immer Modelle, wie etwa die sogenannten Multi-Stage-Choice-Modelle, deren Integration in die Optimierungsprobleme des RM bisher kaum erforscht ist, in Zukunft jedoch ebenfalls möglich sein könnte (vgl. Strauss et. al. (2018, S. 384)). Anwender werden daher künftig aus einer noch größeren Zahl von im RM effizient einsetzbaren Discrete-Choice-Modellen wählen können. Die Kenntnis der Annahmen, des Spezifikationsprozesses und der Schätzverfahren der verschiedenen Modelle wird daher umso wichtiger, um eine optimale Modellauswahl für einen konkreten Anwendungsfall treffen zu können.

## Literaturverzeichnis

- Aouad, A.; Feldman, J. und Segev, D. (2018): The exponential choice model: Assortment optimization and application to public transit choice prediction in San Francisco. Working Paper, London Business School.
- BDF (2019): Sicherheitsentgelte. <http://www.bdf.aero/themen/sicherheit/sicherheitsentgelte/>. Abgerufen am 03.07.2019.
- Belobaba, P. (2009a): Overview of airline economics, markets and demand. In: Belobaba, P.; Odoni, A. und Barnhart, C. (Hrsg.): The global airline industry. S. 47-72.
- Belobaba, P. (2009b): Fundamentals of pricing and revenue management. In: Belobaba, P.; Odoni, A. und Barnhart, C. (Hrsg.): The global airline industry. S. 73-112.
- Belobaba, P. (2009c): Airline operating costs and measures of productivity. In: Belobaba, P.; Odoni, A. und Barnhart, C. (Hrsg.): The global airline industry. S. 113-152.
- Belobaba, P.; Swelbar, W. und Barnhart, C. (2009): Information technology in airline operations, distribution and passenger processing. In: Belobaba, P.; Odoni, A. und Barnhart, C. (Hrsg.): The global airline industry. S. 441-466.
- Berbeglia, G.; Garassino, A. und Vulcano, G. (2018): A comparative empirical study of discrete choice models in retail operations. Working Paper, University of Melbourne.
- Blanchet, J.; Gallego, G. und Goyal, V. (2016): A markov chain approximation to choice modeling. *Operations Research* 64(4). S. 886-905.
- Borman, S. (2004): The expectation maximization algorithm – A short tutorial. Working Paper, University of Notre Dame.
- Bundesministerium der Finanzen (2011): Die Luftverkehrssteuer. Monatsbericht des BMF April 2011. S. 73-77.
- Cascetta, E. (2009): Transportation systems analysis – Models and applications. Springer, New York.
- Cooper, W.; Homem-de-Mello, T. und Kleywegt, A. (2006): Models of the spiral-down effect in revenue management. *Operations Research* 54(5). S. 968-987.
- Cramer, E. und Kamps, U. (2017): Grundlagen der Wahrscheinlichkeitsrechnung und Statistik. Springer, Berlin.



- Dempster, A.; Laird, N. und Rubin, D. (1977): Maximum likelihood from incomplete data via the EM Algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)* 39(1). S. 1-38.
- Desrosiers, J. und Lübbecke, M. (2005): A primer in column generation. In: Desaulniers, G.; Desrosiers, J. und Solomon, M. (Hrsg.): *Column Generation*. Springer, New York. S. 1-31.
- Farias, V.; Jagabathula, S. und Shah, D. (2013): A non-parametric approach to modeling choice with limited data. *Management Science* 59(2). S. 305-322.
- Feldman, J. und Topaloglu, H. (2017): Revenue management under the markov chain choice model. *Operations Research* 65(5). S. 1322-1342.
- Gopalan, R. und Talluri, K. (1998): Mathematical models in airline schedule planning: A survey. *Annals of Operations Research* 76. S. 155-185.
- Gupta, A. und Hsu, D. (2017): Parameter identification in markov chain choice models. Working paper, Columbia University.
- Honhon, D.; Pan, X. und Jonnalagedda, S. (2015): Heuristics for the assortment planning problem under ranking-based consumer choice models. Working Paper, University of Texas.
- Klein, R. und Steinhardt, C. (2008): *Revenue Management – Grundlagen und Mathematische Methoden*. Springer, Berlin – Heidelberg.
- Kök, G. und Fisher, M. (2007): Demand estimation and assortment optimization under substitution: Methodology and application. *Operations Research* 55(6). S. 1001-1021.
- Larsson, T.; Migdalas, A. und Patriksson, M. (2015): A generic column generation principle: Derivation and convergence analysis. *Operational Research* 15(2). S. 163-198.
- Lewandowsky, S. und Farrell, S. (2011): *Computational Modeling in Cognition: Principles and Practice*. SAGE Publications, Thousand Oaks.
- Manager Magazin (2009): Schrumpfende Budgets. <https://www.manager-magazin.de/lifestyle/reise/a-667653.html>. Abgerufen am: 03.07.2019.
- Miranda Bront, J.; Mendez-Diaz, I. und Vulcano, G. (2009): A column generation algorithm for choice-based network revenue management. *Operations Research* 57(3). S. 769-784.

- SciPy (2019): Scipy.optimize.minimize. <https://docs.scipy.org/doc/scipy/reference/generated/scipy.optimize.minimize.html>. Abgerufen am 04.07.2019.
- Shen, J. (2009): Latent class model or mixed logit model? A comparison by transport mode choice data. *Applied Economics* 41(22). S. 2915-2924.
- Sher, I.; Fox, J.; il Kim, K. und Bajari, P. (2011): Partial identification of heterogeneity in preference orderings over discrete choices. Working Paper, National Bureau of Economic Research.
- Simsek, S. und Topaloglu, H. (2018): Technical note – An expectation-maximization algorithm to estimate the parameters of the markov chain choice model. *Operations Research* 66(3). S. 748-760.
- Strauss, A.; Klein, R. und Steinhardt, C. (2018): A review of choice-based revenue management: Theory and methods. *European Journal of Operational Research* 271(2). S. 375-387.
- Talluri, K. und van Ryzin, G. (2004a): The theory and practice of revenue management. Springer, New York.
- Talluri, K. und van Ryzin, G. (2004b): Revenue management under a general discrete choice model of consumer behavior. *Management Science* 50(1). S. 15-33.
- Train, K. (2009): Discrete choice methods with simulation. Cambridge University Press, Cambridge.
- van Ryzin, G. und Vulcano, G. (2015): A market discovery algorithm to estimate a general class of nonparametric choice models. *Management Science* 61(2). S. 281-300.
- Vulcano, G.; van Ryzin, G. und Chaar, W. (2010): OM practice – Choice-based revenue management: An empirical study of estimation and optimization. *Manufacturing & Service Operations Management* 12(3). S. 371-392.

## 9 Anhang

$i$	$\beta_{11}$	$\beta_{21}$	$\beta_{12}$	$\beta_{22}$	$s_1$	$s_2$	$LL(\boldsymbol{\beta}, \mathbf{s})$
0	-0,001	0,017	0,004	0,004	0,200	0,800	-11,730
1	-0,004	0,038	0,003	0,001	0,204	0,796	-11,661
2	-0,008	0,065	0,004	-0,004	0,210	0,790	-11,529
3	-0,015	0,110	0,005	-0,012	0,223	0,777	-11,229
4	-0,030	0,199	0,007	-0,024	0,252	0,748	-10,627
5	-0,060	0,362	0,010	-0,043	0,307	0,693	-9,811
6	-0,095	0,553	0,015	-0,072	0,375	0,625	-9,037
7	-0,132	0,751	0,023	-0,118	0,435	0,565	-8,328
8	-0,178	0,993	0,038	-0,218	0,475	0,525	-7,721
9	-0,241	1,326	0,093	-0,606	0,494	0,506	-7,126
10	-0,336	1,826	0,246	-1,703	0,500	0,500	-6,934
11	-0,423	2,249	0,727	-4,628	0,500	0,500	-6,931
12	-0,447	2,375	0,740	-4,706	0,500	0,500	-6,931
13	-0,448	2,375	0,739	-4,703	0,500	0,500	-6,931
14	-0,447	2,375	0,739	-4,703	0,500	0,500	-6,931
15	-0,447	2,375	0,739	-4,703	0,500	0,500	-6,931

**Tabelle 15:** Daten zur Schätzung eines LC-MNL-Modells mittels des EM-Verfahrens von Train (2009) auf Basis der Trainingsdaten der Beispielinstantz für die Iterationen  $i = 0, \dots, 15$

$j$	0	1	2	3	4
$d = 1$					
$\Theta_j^1(S_1)$	0,2	0,2	0,2	0,2	0,2
$\Psi_1^1(j, S_1)$	0	1	0	0	0
$E(\Lambda_j^1   j_1, S_1, \lambda^1, \rho^1)$	0	1	0	0	0
$d = 2$					
$\Theta_j^1(S_2)$	0,333	0,267	0,267	0,333	0,333
$\Psi_4^1(j, S_2)$	0	0,333	0,333	0	1
$E(\Lambda_j^2   j_2, S_2, \lambda^1, \rho^1)$	0	0,2	0,2	0	0,6
$E(P_{1j}^2   j_2, S_2, \lambda^1, \rho^1)$	0	0	0,067	0	0,2
$E(P_{2j}^2   j_2, S_2, \lambda^1, \rho^1)$	0	0,067	0	0	0,2
$d = 3$					
$\Theta_j^1(S_3)$	0,333	0,267	0,333	0,267	0,333
$\Psi_2^1(j, S_3)$	0	0,333	1	0,333	0
$E(\Lambda_j^3   j_3, S_3, \lambda^1, \rho^1)$	0	0,2	0,6	0,2	0
$E(P_{1j}^3   j_3, S_3, \lambda^1, \rho^1)$	0	0	0,2	0,067	0
$E(P_{3j}^3   j_3, S_3, \lambda^1, \rho^1)$	0	0,067	0,2	0	0
$d = 4$					
$\Theta_j^1(S_4)$	0,333	0,333	0,267	0,333	0,267
$\Psi_0^1(j, S_4)$	1	0	0,333	0	0,333

$E(\Lambda_j^4   j_4, S_4, \lambda^1, \rho^1)$	0,6	0	0,2	0	0,2
$E(P_{2j}^4   j_4, S_4, \lambda^1, \rho^1)$	0,2	0	0	0	0,067
$E(P_{4j}^4   j_4, S_4, \lambda^1, \rho^1)$	0,2	0	0,067	0	0
$d = 5$					
$\Theta_j^1(S_5)$	0,2	0,2	0,2	0,2	0,2
$\Psi_4^1(j, S_5)$	0	0	0	0	1
$E(\Lambda_j^5   j_5, S_5, \lambda^1, \rho^1)$	0	0	0	0	1
$d = 6$					
$\Theta_j^1(S_6)$	0,333	0,333	0,267	0,267	0,333
$\Psi_1^1(j, S_6)$	0	1	0,333	0,333	0
$E(\Lambda_j^6   j_6, S_6, \lambda^1, \rho^1)$	0	0,6	0,2	0,2	0
$E(P_{2j}^6   j_6, S_6, \lambda^1, \rho^1)$	0	0,2	0	0,067	0
$E(P_{3j}^6   j_6, S_6, \lambda^1, \rho^1)$	0	0,2	0,067	0	0
$d = 7$					
$\Theta_j^1(S_7)$	0,333	0,267	0,333	0,333	0,267
$\Psi_2^1(j, S_7)$	0	0,333	1	0	0,333
$E(\Lambda_j^7   j_7, S_7, \lambda^1, \rho^1)$	0	0,2	0,6	0	0,2
$E(P_{1j}^7   j_7, S_7, \lambda^1, \rho^1)$	0	0	0,2	0	0,067
$E(P_{4j}^7   j_7, S_7, \lambda^1, \rho^1)$	0	0,067	0,2	0	0

$d = 8$					
$\Theta_j^1(S_8)$	0,333	0,333	0,267	0,267	0,333
$\Psi_4^1(j, S_8)$	0	0	0,333	0,333	1
$E(\Lambda_j^8   j_8, S_8, \lambda^1, \rho^1)$	0	0	0,2	0,2	0,6
$E(P_{2j}^8   j_8, S_8, \lambda^1, \rho^1)$	0	0	0	0,067	0,2
$E(P_{3j}^8   j_8, S_8, \lambda^1, \rho^1)$	0	0	0,067	0	0,2
$d = 9$					
$\Theta_j^1(S_9)$	0,25	0,25	0,25	0,2	0,25
$\Psi_1^1(j, S_9)$	0	1	0	0,25	0
$E(\Lambda_j^9   j_9, S_9, \lambda^1, \rho^1)$	0	0,8	0	0,2	0
$E(P_{3j}^9   j_9, S_9, \lambda^1, \rho^1)$	0	0,2	0	0	0
$d = 10$					
$\Theta_j^1(S_{10})$	0,5	0,5	0,4	0,4	0,4
$\Psi_0^1(j, S_{10})$	1	0	0,5	0,5	0,5
$E(\Lambda_j^{10}   j_{10}, S_{10}, \lambda^1, \rho^1)$	0,4	0	0,2	0,2	0,2
$E(P_{2j}^{10}   j_{10}, S_{10}, \lambda^1, \rho^1)$	0,2	0	0	0,1	0,1
$E(P_{3j}^{10}   j_{10}, S_{10}, \lambda^1, \rho^1)$	0,2	0	0,1	0	0,1
$E(P_{4j}^{10}   j_{10}, S_{10}, \lambda^1, \rho^1)$	0,2	0	0,1	0,1	0

**Tabelle 16: Daten zur Schätzung eines Markov-Chain-Modells mittels des EM-Verfahrens von Simsek und Topaloglu (2018) auf Basis der Trainingsdaten der Beispielinstanz für Iteration  $i = 1$**

## **Schriftliche Versicherung**

Hiermit versichere ich, dass die vorliegende Arbeit von mir selbständig und ohne fremde Hilfe verfasst wurde. Alle Stellen, die ich wörtlich oder sinngemäß aus veröffentlichten oder nicht veröffentlichten Quellen übernommen habe, habe ich als solche gekennzeichnet. Es wurden alle Quellen, auch Internetquellen, ordnungsgemäß angegeben.

Ort, Datum

David Fleckenstein