



## **Bachelorarbeit**

zum Thema

# **Analyse und Vergleich von Methoden zur Prädiktion der Auswirkungen von Infrastrukturmaßnahmen auf den Modal Split**

vorgelegt von: Sabine Krause  
Matrikel-Nr.: 3598286  
Studiengang: Verkehrswirtschaft  
geboren am: 17.03.1990 in Schwedt/Oder  
Verantwortlicher  
Hochschullehrer: Dr.-Ing. Stefan Lämmer

Betreut durch

Dipl.-Ing. Dipl.-Wirtsch.-Ing. Florian Brinkmann

Dipl.-Ing.-Inf. Thomas Böhm



Deutsches Zentrum  
DLR für Luft- und Raumfahrt

Institut für Verkehrs-  
systemtechnik

Dresden, den 31.08.2012

---

## Aufgabenstellung für eine Bachelorarbeit

Bearbeiter: cand. B. Sc. Sabine Krause  
Matrikelnummer: 3598286  
Fachbereich: Fakultät Verkehrswissenschaften  
Betreuer: Dipl.-Ing. Dipl.-Wirtsch.-Ing. Florian Brinkmann  
Dipl.-Ing.-Inf. Thomas Böhm  
Bearbeitungsdauer: 4 Monate  
Studiengang: Verkehrswirtschaft

Lilienthalplatz 7  
38108 Braunschweig

Telefon 0531 295-3514  
Telefax 0531 295-3402  
E-Mail Florian.Brinkmann@dlr.de

09. Februar 2012

### Thema: Analyse und Vergleich von Methoden zur Prädiktion der Auswirkungen von Infrastrukturmaßnahmen auf den Modal Split

Das Gesamtsystem Bahn soll wirtschaftlich, sicher und verfügbar sein, um einen pünktlichen, sicheren und wirtschaftlichen Bahnbetrieb zu gewährleisten. Gleichzeitig muss es auch gesetzlichen Vorgaben und gesellschaftlichen Interessen genügen. Ein maßgeblicher Aspekt ist hier die Gestaltung der Schienenverkehrsinfrastruktur. Die Informationsgrundlage, auf der entsprechende Entscheidungen beruhen, spiegelt oftmals nur den kostenfokussierten Standpunkt des Infrastrukturbetreibers wider oder beruht auf Bewertungsmethoden wie beispielsweise der Bewertung im Bundesverkehrswegeplan. Für eine objektivierte Entscheidung ist ein Bewertungsinstrument erforderlich, welches neben dem betriebswirtschaftlichen (technisch-monetären) auch den gesellschaftlichen Nutzen über den vollständigen Lebenszyklus abbildet.

Das DLR Institut für Verkehrssystemtechnik (ITS) entwickelt ein solches Instrument für die Evaluation des Life Cycle Benefit. Dabei sollen verschiedenen Auswirkungen auf den Modalsplit (Verkehrsentlastung, Stauverringerng), die Gesellschaft (Lärmwahrnehmung, Reisegeschwindigkeit, Komfort, Sicherheit), die Volkswirtschaft (Wettbewerbsfähigkeit, Effizienz, Kosten) und die Umwelt (Ressourcenverbrauch, CO<sub>2</sub>-Emission, Lärmemission, Flächenverbrauch) einbezogen werden.

Inhalt der Bachelorarbeit soll es sein, in diesem Kontext den Teilaspekt der Prädiktion von Auswirkungen von Infrastrukturmaßnahmen auf den Modal Split zu untersuchen. Infrastrukturmaßnahmen eines Verkehrsträgers haben immer auch eine Wechselbeziehung zu anderen Verkehrsträgern. Bisher wurden diese Wechselwirkungen nur vereinzelt in eine Betrachtung eingebunden, obwohl hier bereits Modelle z.B. für die Verkehrsentlastung oder die Stauverringerng existieren.

Die Aufgabe von Frau Krause umfasst zunächst eine Literaturrecherche zu verfügbaren Modellen und Methoden zur Vorhersage von Änderungen im Modal Split. Der diesbezügliche Stand der Forschung ist zu dokumentieren.

Die jeweils relevanten Kriterien und Parameter, die Einfluss auf die Entwicklung des Modal Split haben, sollen identifiziert und beschrieben werden.

Die verfügbaren Methoden und Modelle sind hinsichtlich Ihrer Stärken und Schwächen zu analysieren. Dabei soll auch Bezug auf die Eignung für die am Institut für Verkehrssystemtechnik laufende



Forschung genommen werden. Je nach Ergebnis sollten dann Vorschläge zur Modifizierung oder Verbesserung der Modelle abgeleitet werden.

Eine durch Betreuung des DLR Instituts für Verkehrssystemtechnik angefertigte studentische Arbeit darf nur nach Rücksprache mit dem Institut vom Studenten an Dritte weitergegeben oder veröffentlicht werden. Dies gilt insbesondere auch für so genannte Internetbörsen für studentische Arbeiten.

Über die Ergebnisse der Arbeit darf nur nach Rücksprache mit dem DLR Institut für Verkehrssystemtechnik verfügt werden. Die Arbeit bleibt Eigentum des Instituts.

Datum der Ausgabe: 11.05.2012

Datum der Abgabe: 31.08.2012

Thomas Böhm

Florian Brinkmann



## Inhaltsverzeichnis

<b>ABBILDUNGSVERZEICHNIS .....</b>	<b>III</b>
<b>TABELLENVERZEICHNIS .....</b>	<b>III</b>
<b>ABKÜRZUNGSVERZEICHNIS.....</b>	<b>IV</b>
<b>SYMBOLVERZEICHNIS .....</b>	<b>V</b>
<b>1 EINLEITUNG .....</b>	<b>1</b>
1.1 MOTIVATION.....	1
1.2 PROBLEMSTELLUNG.....	1
1.3 AUFBAU DER ARBEIT .....	1
<b>2 ALLGEMEINES ZUR VERKEHRSMITTELWAHL .....</b>	<b>3</b>
2.1 DAS VIER-STUFEN-MODELL DER VERKEHRSPLANUNG .....	3
2.2 AUSGEWÄHLTE MODELLANSÄTZE DER VERKEHRSMITTELWAHL.....	5
2.3 DETERMINANTEN DER VERKEHRSMITTELWAHL .....	7
2.4 WAHLTHEORIE .....	8
2.5 DIE NUTZENFUNKTION .....	8
<b>3 MODELLE DISKRETER ENTSCHEIDUNGEN.....</b>	<b>12</b>
3.1 LP-MODELL .....	13
3.2 PROBIT-MODELL .....	14
3.3 LOGIT-MODELL.....	15
3.4 WEITERE MODELLE .....	17
<b>4 DIE LOGIT-ANALYSE .....</b>	<b>19</b>
4.1 AUFSTELLEN DER NUTZENFUNKTION .....	19
4.2 DATENERHEBUNG.....	20
4.3 SCHÄTZUNG DER PARAMETER .....	22
4.4 ÜBERPRÜFUNG.....	27
4.5 BESTIMMUNG DES MODAL SPLIT .....	31
<b>5 PROGNOSE VON ÄNDERUNGEN IM MODAL SPLIT .....</b>	<b>32</b>
5.1 INTERPRETATION DER PARAMETER .....	32



5.2	ABLEITUNGEN .....	33
5.3	ELASTIZITÄTEN.....	34
5.4	VERGLEICH DER BERECHNETEN ANTEILE .....	34
<b>6</b>	<b>BEISPIELHAFTE ANWENDUNG DER LOGIT-ANALYSE.....</b>	<b>35</b>
6.1	AUSWAHL DER EINFLUSSFAKTOREN .....	35
6.2	NUTZENFUNKTION .....	36
6.3	EXEMPLARISCHE DATENSÄTZE .....	37
6.4	MODELLSCHÄTZUNG .....	38
6.5	ÜBERPRÜFUNG.....	39
6.6	INTERPRETATION DER PARAMETER .....	40
6.7	BERECHNUNG DES MODAL SPLITS EINER RELATION .....	43
6.8	ÄNDERUNGEN IM MODAL SPLIT DURCH VERKEHRSMÄßNAHMEN.....	44
6.9	DISKUSSION DER ERGEBNISSE .....	45
<b>7</b>	<b>ZUSAMMENFASSUNG UND AUSBLICK .....</b>	<b>47</b>
	<b>LITERATURVERZEICHNIS.....</b>	<b>VI</b>
	<b>ANHANG.....</b>	<b>X</b>
	<b>EHRENWÖRTLICHE ERKLÄRUNG .....</b>	<b>XX</b>



## Abbildungsverzeichnis

Abbildung 1: Das Vier-Stufen-Modell der Verkehrsplanung (aus Treiber, 2011)....	4
Abbildung 2: Verteilungsfunktion des binären LP-Modells (aus Ben-Akiva/Lerman, 1985) .....	13
Abbildung 3: Verteilungsfunktion des binären Probit-Modells (aus Ben-Akiva/Lerman, 1985) .....	15
Abbildung 4: Beispiel eines zweistufigen Nested Logit-Modells (aus Urban, 1993) .....	18
Abbildung 5: Übersicht über Instrumente der Modellevaluation (vgl. Tutz, 2000).	28
Abbildung 6: Änderung des Modal Split durch Infrastrukturmaßnahmen im Beispiel .....	45
Abbildung 7: Auswahlwahrscheinlichkeiten des Beispiels .....	XIX

## Tabellenverzeichnis

Tabelle 1: Eigenschaften der gebräuchlichsten Modelle diskreter Entscheidungen .....	17
Tabelle 2: Einflussgrößen der Verkehrsmittelwahl im Beispiel .....	36
Tabelle 3: fiktive Befragungsdaten (Auszug) .....	37
Tabelle 4: geschätzte Parameter der Nutzenfunktion.....	38
Tabelle 5: geschätzte Modellparameter bei Änderung der Einheiten der Reisezeit von Minuten auf Stunden .....	41
Tabelle 6: Werte der t-Statistik der geschätzten Parameter .....	41
Tabelle 7: Zeitwerte nach Einkommensklassen sortiert .....	42
Tabelle 8: Elastizitäten der Parameter des Modells .....	42
Tabelle 9: Eigenschaften der Wahlalternative Bahn einer beispielhaften Relation	43
Tabelle 10: Quantile der Chi-Quadrat-Verteilung nach ausgewählten Wahrscheinlichkeiten p und Freiheitsgraden FG (vgl. wikibooks, 2011 und Eckey, 2001) .....	XI
Tabelle 11: Tabelle der fiktiven Befragungsdaten .....	XII
Tabelle 12: Ergebnisse des Beispiels.....	XVIII



## Abkürzungsverzeichnis

DLR	Deutsches Zentrum für Luft- und Raumfahrt e.V.
Gl.	Gleichung
h	Stunden
i.i.d.	Independent and identically distributed (unabhängig und identisch verteilt)
IIA	Independence from Irrelevant Alternatives (Unabhängigkeit von irrelevanten Alternativen)
km	Kilometer
KONTIV	kontinuierliche Verkehrserhebung
LP	Linear Probability (lineare Wahrscheinlichkeit)
MiD	Mobilität in Deutschland
(M)IV	(Motorisierter) Individualverkehr
Min	Minuten
ML	Maximum Likelihood
MNL	Multinomial Logit
ÖV	öffentlicher Verkehr
PKW	Personenkraftwagen
PTV Group	Planung Transport Verkehr (AG mit Sitz in Karlsruhe)
SrV	System repräsentativer Verkehrsbefragung



## Symbolverzeichnis

e	eulerische Zahl (=2,718...)
$i=1, \dots, I$	Personenindex
$j=1, \dots, J$	Parameterindex
$k=1, \dots, K$	Alternativenindex
$U_{ki}$	Gesamtnutzen der Alternative k für Person i
$V_{ki}$	deterministischer Nutzenanteil der Alternative k für Person i
$\varepsilon_{ki}$	stochastischer Nutzenanteil der Alternative k für Person i
$P_{ki}$	Wahlwahrscheinlichkeit von Person i für Alternative k
$\beta$	Modellparameter (Gewichtungsfaktor für x)
x	exogene Modellvariable (Einflussfaktoren der Verkehrsmittelwahl)
$\lambda$	Skalierungsparameter der Gumbel-Verteilung UND: Testvariable des Likelihood-Ratio-Tests
$\rho^2$	Likelihood-Ratio-Index
$L(\beta)$	Likelihood-Funktion
$LL(\beta)$	Log-Likelihood-Funktion (die logarithmierte Likelihood-Funktion $\ln L(\beta)$ )
E	Elastizität
$y_{ki}$	Häufigkeit der Wahl von Alternative k in Personengruppe i
$\hat{Y}_{ki}$	geschätzte Häufigkeit der Wahl von Alternative k in Gruppe i
$n_i$	Anzahl an Entscheidungen in Personengruppe i
$F_j$	Schätzbedingung des Newton'schen Iterationsverfahrens $(F_j = X_j^{MNL} - X_j^{data})$
J	Jacobi-Matrix $J = \frac{\partial F}{\partial \beta}$
$B_i$	Multinomialkoeffizient $(B_{(y_i)} = \frac{(n_i)!}{\prod_k (y_{ki})!})$





# **1 Einleitung**

## **1.1 Motivation**

Mit dem wachsenden Verkehrsaufkommen und dem steigenden Interesse für die Auswirkungen des Verkehrs auf die Umwelt werden immer häufiger Maßnahmen zur Verkehrsvermeidung, Stauverringern und der Verlagerung des Verkehrs auf umweltfreundlichere Verkehrsmittel untersucht. Um die Wirksamkeit von Verkehrsmaßnahmen zur Verlagerung des Verkehrs berechnen zu können, werden in der Verkehrsplanung die Entscheidungen von Verkehrsteilnehmern nachgebildet. Ein wesentlicher Aspekt hierbei ist es, nachzuvollziehen, wie Verkehrsteilnehmer bei der Entscheidungsfindung für ein bestimmtes Verkehrsmittel vorgehen.

## **1.2 Problemstellung**

Am Institut für Verkehrssystemtechnik des Deutschen Zentrums für Luft- und Raumfahrt e.V. (DLR) wurde im Rahmen eines Forschungsprojektes ein Instrumentarium zur Bewertung verschiedener Szenarien von Maßnahmen in der Infrastruktur von Bahnanlagen entwickelt. Hierbei wurden auch deren Wirtschaftlichkeit und die resultierenden Auswirkungen auf den Bahnbetrieb untersucht. Im Folgeprojekt sollen nun weitere Faktoren wie Auswirkungen auf die Gesellschaft, die Umwelt, die Volkswirtschaft und den Modal Split ebenfalls in die Bewertung einbezogen werden. Ziel ist es, eine computergestützte, multikriterielle Bewertungsmethodik für Infrastrukturmaßnahmen unter Einbeziehung von allen relevanten Einflussfaktoren zu erschaffen. Aufgabe dieser Arbeit ist es, den Teilaspekt der Auswirkungen von Infrastrukturmaßnahmen auf den Modal Split zu untersuchen. Dazu sollen existierende Modelle zur Berechnung des Modal Split und dessen Änderungen analysiert und verglichen werden, um diese Ansätze später in die Bewertung mit einfließen zu lassen.

## **1.3 Aufbau der Arbeit**

Als verkehrsplanerisches Thema wird die Verkehrsmittelwahl zunächst als ein Planungsschritt in dieses Gebiet eingebettet. Verschiedene Modellansätze zur



Berechnung des Modal Split werden hierbei mit ihren jeweiligen Vor- und Nachteilen vorgestellt. Die gebräuchlichsten Modelle der diskreten Wahltheorie werden genauer erläutert und diskutiert. Auf das am häufigsten verwendete Modell - das Logit-Modell - wird im Besonderen eingegangen. Hierzu wird in einem eigenen Kapitel schrittweise beschrieben, wie eine Logit-Analyse zur Prognose des Modal Split durchgeführt wird. Auf die Vorhersage von Änderungen im Modal Split wird gesondert eingegangen. Abschließend wird die Logit-Analyse in einem fiktiven Beispiel für den binomialen Entscheidungsfall angewandt. Die vorab erläuterten Konzepte finden hierbei Anwendung. Es folgt ebenso eine Interpretation der errechneten Parameter und auch exemplarische Infrastrukturmaßnahmen werden untersucht.



## 2 Allgemeines zur Verkehrsmittelwahl

Zur Klärung der Frage, welche Auswirkungen Infrastrukturmaßnahmen auf den Modal Split haben, muss zunächst die Bedeutung des Modal Split und dessen Bestimmung definiert werden. „*Der Modal Split ist die Aufteilung der Verkehrsnachfrage auf die verschiedenen Verkehrsmittel*“ (Kirchhoff, 2002). Er entsteht dadurch, dass sich Personen für ihre zurückzulegenden Wege für einen bestimmten Verkehrsmodus entscheiden.

Die Verkehrsplanung beschäftigt sich mit dem Verkehrsgeschehen im öffentlichen Verkehrsraum (Schnabel/Lohse, 2011). Dies beinhaltet die Aufgabe der Analyse und Prognose von verkehrlichen Auswirkungen, unter anderem als Folge von infrastrukturellen Änderungen (Steierwald, 2003). Im Folgenden wird daher zunächst die Verkehrsmittelwahl als ein Teilaspekt in die Verkehrsplanung eingebettet. Im Anschluss werden die Modelle zur Verkehrsmittelwahl genauer erläutert.

### 2.1 Das Vier-Stufen-Modell der Verkehrsplanung

In der Verkehrsplanung ist der Schritt *Verkehrsmittelwahl* einer von vier sequentiell ablaufenden Planungsstufen, welche im Folgenden dargestellt werden. Hierbei ist zu beachten, dass es in der Reihenfolge der Schritte zwei Herangehensweisen gibt, wobei bei der hier dargestellten vom *Trip-Interchange* Ansatz gesprochen wird (Schnabel/Lohse, 2011). Die jeweiligen Planungsschritte bilden zum einen die Reihenfolge der Entscheidungen nach, welche der Verkehrsteilnehmer trifft, zum anderen die Berechnungen, welche der Planer daraus ableiten kann.

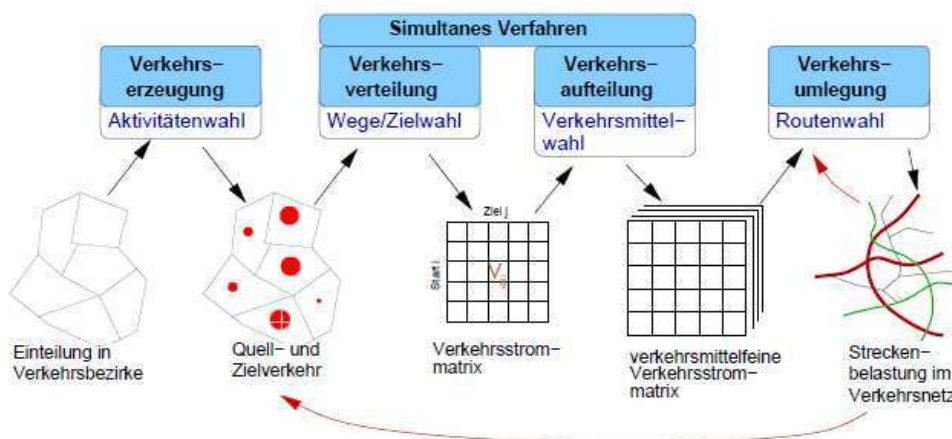
Zunächst treffen die Verkehrsteilnehmer eine Entscheidung über die durchzuführende Aktivität (Treiber, 2011). Im ersten Planungsschritt - der *Verkehrserzeugung* - wird die Anzahl der Fahrten, die pro Person, Zeiteinheit und Verkehrszelle durchgeführt werden, bestimmt. Hier werden sowohl ausgehende, als auch eingehende Verkehrsströme einer Zelle gemessen. Dabei werden zunächst nur Gesamtzahlen berechnet, ohne zu beachten, woher der jeweilige Verkehrsstrom kommt bzw. welches Ziel er hat (Eckey, 2000).

Dies folgt nun im zweiten Schritt; in der Zielwahl der Entscheidungsträger wählen diese ihren Zielort, die zugehörige *Verkehrsverteilung* ermittelt die konkrete Anzahl an Wegen von jeder Quelle zu jedem Ziel (Eckey, 2000).

In der Stufe der *Verkehrsaufteilung* geht es nun um die Verkehrsmittelwahl der Verkehrsteilnehmer; die oben genannten Verkehrsströme werden hier auf die einzelnen Verkehrsmodi aufgeteilt, der Modal Split wird bestimmt (Eckey, 2000).

Schließlich folgt die *Verkehrsumlegung*, in der die einzelnen Verkehrsströme auf konkrete Routen im Netz umgelegt werden. Die Verkehrsteilnehmer treffen die Routenwahl als letztes, nachdem sie sich für Ziel und Verkehrsmittel entschieden haben (Eckey, 2000).

Im Gegensatz zu diesem Ansatz, kehrt der *Trip-End* Ansatz die Schritte Verteilung und Aufteilung um. Der Verkehrsteilnehmer entscheidet sich also für ein Verkehrsmittel, bevor er das Ziel wählt (Schnabel/Lohse, 2011). Nicht nur die Reihenfolge der Schritte, auch der sequentielle Ablauf der Planungsschritte an sich ist ein viel diskutiertes Thema (Steierwald, 2003), da durchaus Wechselwirkungen zwischen den einzelnen Stufen bestehen (Krämer, 1992). Es existieren auch Simultanmodelle, bei denen mehrere Schritte gleichzeitig ablaufen (Sammer, 1982). Diese berechnen alle Entscheidungswahrscheinlichkeiten für sämtliche Kombinationen an Zielen, Verkehrsmitteln und Wegen. Sie beachten somit jedoch auch Lösungen, die evtl. unzulässig sind, da beispielsweise auf einigen Relationen ein bestimmtes Verkehrsmittel nicht zu Verfügung steht (Lohse/Schneider, 1997).



**Abbildung 1:** Das Vier-Stufen-Modell der Verkehrsplanung (aus Treiber, 2011)



Wie in der Abbildung erkennbar, werden im Vier-Stufen-Modell die Ergebnisse einer Modellstufe als Eingangsdaten für die nächste Stufe verwendet. Auch müssen Rückkopplungen zwischen den einzelnen Modellen berücksichtigt werden (Schnabel/Lohse, 2011).

Der Vollständigkeit halber sollte erwähnt werden, dass heutzutage in der Verkehrsplanung die verschiedenen Modellschritte des Vier-Stufen-Modells softwaregestützt modelliert und berechnet werden. In Deutschland marktführend im Bereich Software zur Verkehrsanalyse und -prognose ist das Programm *PTV Visum* von der PTV Group (PTV, 2012a). Solche Programme benötigen als Eingangsdaten ein Netzmodell des zu untersuchenden Verkehrsnetzes, sowie personenbezogene und strukturelle Daten des Untersuchungsgebietes aus Verkehrserhebungen. Daraus können dann Modal Split und die Wirkungen von Verkehrsmaßnahmen berechnet werden. (Kern, 2004)

Entscheidend für die Fragestellung dieser Arbeit ist die Modellstufe der Verkehrsmittelwahl, welche im Folgenden genauer betrachtet wird.

## **2.2 Ausgewählte Modellansätze der Verkehrsmittelwahl**

Zur Bestimmung des Modal Split gibt es verschiedene Herangehensweisen. Die Verkehrsteilnehmer entscheiden sich für ihre zurückzulegenden Wege für einen Verkehrsmodus. Der Modal Split wird als Aggregat aller Einzelentscheidungen bestimmt. Der Entscheidungsprozess kann durch verschiedene Modelle nachgebildet werden und ist von mehreren Determinanten abhängig.

Unterschieden werden können die Modelle nach dem Aggregationsniveau der Datenauswertung. In aggregierten Modellen wird das Verhalten eines repräsentativen Entscheidungsträgers abgebildet (Maier/Weiss, 1990), dies lässt jedoch keine Aussagen darüber zu, wie sich bestimmte Maßnahmen auf das Verhalten von Individuen auswirken (Gorr, 1997). In disaggregierten Modellen hingegen kann die Heterogenität der Entscheidungsträger berücksichtigt werden (Maier/Weiss, 1990). Hierbei stehen die Entscheidungen der Einzelpersonen im Vordergrund, wobei die Einflüsse der Umgebung auf die Entscheidung abgebildet werden (Lohse/Schneider, 1997).



Außerdem werden die Modelle nach der Abbildung des Verhaltens der Entscheidungsträger differenziert. Zunächst wurde davon ausgegangen, dass sich das Verhalten durch äußere, deterministische Merkmale des Verkehrsangebotes erklären lässt (Gorr, 1997). Diese *nicht-verhaltenstheoretischen Modelle* bilden lediglich ausgewählte Charakteristika des Verkehrsangebotes ab und nehmen an, dass die Entscheidung ausschließlich darauf basiert (Keuchel, 1994). Sie beschreiben also nicht komplett das Verhalten einzelner Entscheidungsträger und können deshalb lediglich Momentandarstellungen sein (Bühler, 2006).

In *verhaltensorientierten Verkehrsmodellen* wird ein ökonomisches Verhalten bei der Entscheidungsfindung unterstellt. Es wird demnach das Verkehrsmittel gewählt, welches für den Verkehrsteilnehmer den höchsten Nutzen aufweist (Bühler, 2006). Zunächst wurden als Einflussfaktoren jedoch nur messbare Merkmale, wie Reisekosten und Reisezeit betrachtet, was sich als unzureichend herausstellte. Eine Erweiterung wurde mit den *einstellungsorientierten Verkehrsmodellen* entwickelt. Diese enthalten auch sog. weiche Faktoren, wie beispielsweise den Komfort, welche nicht objektiv messbar sind. Auch wird hier beachtet, dass die Entscheidungsträger nicht allein wegen der genannten Faktoren eine Entscheidung treffen, sondern auch Einstellungen von Personen zu den jeweiligen Verkehrsmitteln ausschlaggebend sind. (Gorr, 1997)

Als Erweiterung dieses Modells wird ein weiterer Aspekt in Betracht gezogen, denn in den bisherigen Modellansätzen wurde nicht überprüft, inwiefern die Entscheidungsträger überhaupt wahlfrei sind. Im Ansatz der *abgestuften Wahlmöglichkeiten* wird beachtet, dass einige Verkehrsteilnehmer in ihren Entscheidungsmöglichkeiten eingeschränkt sind, weil sie beispielsweise keinen Zugang zu öffentlichen Verkehrsmitteln haben oder keinen Führerschein besitzen (Gorr, 1997). Meist wird unterschieden zwischen *captive riders*, diejenigen Personen, die in ihrer Entscheidung abhängig vom öffentlichen Verkehr (ÖV) sind, da ihnen beispielsweise kein Personenkraftwagen (PKW) zur Verfügung steht und *captive drivers*, die hingegen keinen Zugang zum ÖV haben und den (motorisierten) Individualverkehr (im Folgenden mit (M)IV abgekürzt) nutzen müssen. Wahlfreie Personen werden *choice riders* genannt. (Lange, 2009)



### **2.3 Determinanten der Verkehrsmittelwahl**

Die Entscheidung für ein bestimmtes Verkehrsmittel hängt von verschiedenen Faktoren ab. Diese werden im Modell als exogene Variable in einer Nutzenfunktion berücksichtigt. Unterscheiden kann man zwischen verschiedenen Klassen von Variablen.

Die generischen Variablen hängen im Allgemeinen vom Verkehrsmittel ab und variieren nur teilweise von Person zu Person. Sie werden also durch das Verkehrsangebot bestimmt und sind meist direkt messbar. Zu solchen Variablen gehören die Reisezeit, Reisekosten, Entfernung und Zuverlässigkeit. (Urban, 1993)

Als sozioökonomisch werden die Variablen bezeichnet, welche von der Person abhängen. Dies sind beispielsweise das Alter, das Geschlecht oder das Einkommen. Sie bilden die Eigenschaft einer Person ab und somit auch gewisse Einschränkungen in der Mobilität. Durch Einbeziehung solcher Faktoren werden ihre Einflüsse auf das Verkehrsmittelwahlverhalten im Modell untersucht. (Knapp, 1998)

Außerdem gibt es zu jeder zur Wahl stehenden Option eine alternativenspezifische Konstante, welche dazu dient, die globale Präferenz einer Person für ein Verkehrsmittel gegenüber den anderen Alternativen und unabhängig von den Eigenschaften, die im Modell erfasst sind, auszudrücken. (Maier/Weiss, 1990)

Betrachtet man die Verkehrsmittelwahl im Güterverkehr, ist darauf zu achten, dass der Entscheidungsträger hier der Spediteur oder der Verloader ist und als Determinanten vor allem Preis und Qualität des Transportes, aber auch der Service des Spediteurs von Bedeutung sind. Die Wahl hängt vor allem vom zu transportierenden Gut ab. (Bühler, 2006)

Die Einflussfaktoren, welche in das Modell aufgenommen werden, sollen die Erklärungsqualität verbessern. Ein mathematisches Modell trifft quantitative Aussagen, es können daher nur quantifizierbare, prognostizierbare und statistisch signifikante Einflussfaktoren aufgenommen werden. (Sammer, 1982)



Zu beachten ist, dass ein Modell nie alle Einflüsse genauestens abbilden kann. Oftmals existieren Charakteristika, die unbeobachtet bleiben. Außerdem kann es durch Messfehler zu Abweichungen im Modell kommen. Diese Abweichungen werden als Zufallskomponente im Modell dargestellt. (Maier/Weiss, 1990)

## 2.4 Wahltheorie

Aus ökonomischer Betrachtungsweise hat sich zur Nachbildung des Entscheidungsprozesses für ein Verkehrsmittel aus einer Menge von Alternativen die Anwendung diskreter stochastischer Entscheidungsmodelle durchgesetzt. Diese verfolgen das Ziel der Nutzenmaximierung der einzelnen Entscheidungsträger und berücksichtigen dabei die relevanten Einflussfaktoren (Lohse/Schneider, 1997). Es wird davon ausgegangen, dass einer Person mehrere sich ausschließende Alternativen zur Verfügung stehen, jeweils mit unterschiedlichen Eigenschaften. Ziel des Entscheidungsträgers, welcher sich als homo oeconomicus rational und unter Abwägung von Vor- und Nachteilen entscheidet, ist es, jede Alternative zu bewerten und diejenige auszuwählen, welche den höchsten Nutzen für ihn aufweist. (Ben-Akiva/Lerman, 1985)

$$k_{\text{gewählt}} = \arg(\max_k (U_k)) \quad (2.1)$$

Mit  $k_{\text{gewählt}}$ , dem gewählten Verkehrsmittel  $k$   
 $U_k$ , dem Nutzen der Alternative  $k$ , welchen es zu maximieren gilt.  
(Treiber, 2011)

Als probabilistische oder stochastische Modelle können sie außerdem Einflussfaktoren abbilden, die bei der Entscheidung für ein Verkehrsmittel von Bedeutung sind, jedoch empirisch nicht erfassbar sind. (Lohse/Schneider, 1997) Als Ergebnis liefern sie die Wahrscheinlichkeit, dass betrachtetes Verkehrsmittel den höchsten Nutzen aufweist und die Person dieses wählt. (Ben-Akiva/Lerman, 1985)

## 2.5 Die Nutzenfunktion

Der zu maximierende Nutzen soll für jede Person und jedes Verkehrsmittel möglichst genau nachgebildet werden. Dabei sollen die Determinanten, welche die Verkehrsmittelwahl beeinflussen in einer Funktion ausgedrückt werden, welche





sich aus einem deterministischen und einem stochastischen Teil zusammensetzt (Maier/Weiss, 1990). Zum deterministischen Teil der Nutzenfunktion gehören die Größen, welche empirisch bestimmt oder gemessen werden können. Andere Komponenten werden dem stochastischen Anteil zugeordnet (Ben-Akiva/Lerman, 1985).

Der Gesamtnutzen ergibt sich dann beispielsweise als Summe<sup>1</sup> des deterministischen und stochastischen Nutzenanteils nach Gleichung (2.2). (Train, 2003)

$$U_{ki} = V_{ki} + \varepsilon_{ki} \quad (2.2)$$

Mit  $U_{ki}$ , dem Gesamtnutzen für die Alternative k und Individuum i  
 $V_{ki}$ , dem deterministischen Nutzen für die Alternative k und Individuum i  
 $\varepsilon_{ki}$ , dem stochastischen Nutzen für die Alternative k und Individuum i

In den deterministischen Teil gehen die oben genannten Einflussfaktoren als exogene Variable ein. Da jeder Einflussfaktor eine unterschiedlich hohe Bedeutung hat und somit unterschiedlich stark in die Nutzenfunktion eingehen soll, werden die Variablen mit Parametern gewichtet.

$$V_{ki} = \beta_0 + \sum_i \beta_{kj} * X_{ki} \quad (2.3)$$

$\beta$ , der Parameter der Gewichtung des Faktors X,  
 X, der Einflussfaktor (exogene Variable)

Die Parameter  $\beta$  bewerten die zugehörige Variable X für die spezifische Person und das Verkehrsmittel. Sie sind mit Hilfe von statistischen Methoden zu schätzen (Train, 2003). Der Parameter  $\beta_0$  ist die alternativenspezifische Konstante und beschreibt die globale Bevorzugung des betrachteten Verkehrsmittels gegenüber allen anderen oder beinhaltet die Einflüsse aller Faktoren, die nicht in der Funktion beachtet wurden.

Die Variablen X können verschiedene Skalierungen aufweisen. Kardinalskalierte

---

<sup>1</sup> Auch andere Arten der Verknüpfung außer die additive sind möglich, dies kann individuell entschieden werden.



Werte wie Reisezeit [in min] können als solche übernommen werden, der Gewichtungsfaktor  $\beta$  dieser Variable trägt dann den Kehrwert der Einheit [ $\text{min}^{-1}$  bzw.  $1/\text{min}$ ]. Die endogene Variable  $V_{ki}$  des Nutzens ist somit einheitenlos. (Treiber, 2011) Es können jedoch auch nominalskalierte Variable in die Nutzenfunktion aufgenommen werden, welche mit Hilfe von sog. Dummyvariablen modelliert werden. Diese formen qualitative Werte zu rechenbaren Größen um. Beispielsweise wird die sozioökonomische Variable Geschlecht mit den dichotomen Werten 0 für männlich und 1 für weiblich ausgedrückt. (Maier/Weiss, 1990)

Soll nur eine Nutzenfunktion für alle Alternativen formuliert werden, ist es üblich Selektoren einzuführen, da die Parameter im Entscheidungsmodell lediglich über Eigenschaften, nicht aber über die Alternativen variieren. Das Merkmal Umsteigehäufigkeit ist beispielsweise nur für die Alternative ÖV relevant und geht nur im Fall, dass diese Alternative betrachtet wird, in die Nutzenfunktion ein. Ein Selektor ist dann eine dichotome Variable, die mit dem Term  $\beta_{\text{Umsteigehäufigkeit}} \cdot X_{\text{Umsteigehäufigkeit}}$  multipliziert wird. Sie weist nur im Fall der Alternative  $k=\text{ÖV}$  den Wert Eins auf. Ist die betrachtete Alternative eine andere, hat sie den Wert Null; somit entfällt der ganze Term aus der Funktion. Dies wird folgendermaßen dargestellt:

$$Y_k = \beta_1 X_1 \delta_{k1} \text{ mit } \delta_{kk'} = \begin{cases} 1 & k = k' \\ 0 & k \neq k' \end{cases} \quad (2.4)$$

Generell sind es nur die generischen Variablen, welche direkt in die Nutzenfunktion eingehen können. Sozioökonomische Variable werden stets an eine Alternative oder eine generische Variable gekoppelt. Eine Aussage über die Veränderung des Nutzens mit der Änderung des Geschlechts ist schließlich nur sinnvoll, wenn sie auf eine bestimmte Alternative bezogen ist. Auch Einflüsse wie die höhere Bewertung der Zeit durch ein höheres Einkommen können durch eine Kopplung der sozioökonomischen Variablen Einkommen an die generische Variable Reisezeit mit in Betracht gezogen werden. (Treiber, 2011)

Welche Variablen konkret in die Nutzenfunktion aufgenommen werden, kann individuell entschieden werden. Als nicht relevant betrachtete Einflussgrößen entfallen; andere können hinzugefügt werden, sodass die Nutzenfunktion je nach Anwendungsfall eine unterschiedliche Komplexität und Genauigkeit aufweist.



Für die Formulierung der Nutzenfunktion gibt es verschiedene Ansätze. Die einfachste Form ist die lineare Verknüpfung, bei der die Anzahl der Parameter der Anzahl der Variablen entspricht. Von Nachteil ist hierbei, dass beispielsweise die sozioökonomischen Variablen, welche nicht zwischen Alternativen unterscheiden, nicht beachtet werden. Da im Nutzenvergleich lediglich Differenzen von Bedeutung sind, hat ein Summand, der jeder Funktion hinzugefügt wird, keinen Einfluss. Die oben beschriebene Form mit Einsatz von Selektoren ist nicht linear, da der Einflussfaktor und auch der Selektor exogene Variable sind, welche zusammen mit nur einem Parameter verknüpft sind. Die Zahl der Parameter entspricht hier also nicht der Zahl der exogenen Variablen. In der quasilinearen Form sind die Modellparameter linear, die exogenen Variablen jedoch nicht. Somit kann das oben genannte Problem, welches in der linearen Form auftritt, verhindert werden. Dies macht die Funktion häufig aussagekräftiger. (Treiber, 2011)

Der Wert, welcher sich für die endogene Variable  $V_{ki}$  ergibt, ist zunächst nicht aussagekräftig, wichtig sind lediglich die Nutzendifferenzen. Der Entscheidungsträger wählt stets den größten Wert, unabhängig davon, welche Ausprägung dieser aufweist. (Ben-Akiva/Lerman, 1985) Somit kann eine Vereinfachung bei der Formulierung der Nutzenfunktionen vorgenommen werden, indem die Nutzenfunktion eines Verkehrsmittels a priori auf Null gesetzt wird und die Funktionen für die anderen Verkehrsmittel in Relation zu dieser Referenzalternative formuliert werden (Treiber, 2011).

Die Nutzenfunktion beinhaltet außerdem die stochastische Komponente  $\varepsilon$ . Sie ist eine Zufallskomponente, die auch Störterm genannt wird. Zufällig ist sie einerseits deshalb, weil nicht alle Einzelheiten beobachtet werden können (Maier/Weiss, 1990), andererseits, weil einzelne Aspekte einer Entscheidung nicht empirisch erfassbar sind (Lohse/Schneider, 1994).

Es kann also lediglich festgelegt werden, welche Verteilung diesen Zufallsanteilen zugrunde liegen. Diesen gibt dem Entscheidungsmodell seine Form. Das Ergebnis wird dann als Zufallsvariable interpretiert und als Wahrscheinlichkeit ausgedrückt. (Maier/Weiss, 1990)



### 3 Modelle diskreter Entscheidungen

Die im Folgenden beschriebenen Modelle bilden diskrete Entscheidungen, also das Wahlverhalten mit einer endlichen Anzahl an unterscheidbaren Optionen, ab. Dabei wird sowohl beachtet, dass ein Teil der Entscheidungen auf die ökonomische Theorie der Nutzenmaximierung zurückzuführen ist, als auch, dass ein Teil der Entscheidungen zufällig, bzw. nicht vollständig zu nachzubilden ist. Die vorgestellten Modelle unterscheiden sich dadurch, dass sie der zufälligen Nutzenkomponente, bzw. den Differenzen aus diesen, unterschiedliche Verteilungen unterstellen. Am einfachsten ist es, wenn unterstellt wird, dass alle Zufallsanteile eines Modells der gleichen Verteilung folgen und unabhängig voneinander sind, diese Eigenschaft wird als i.i.d. (aus dem englischen „independent and identically distributed“) bezeichnet. (Lohse/Schneider, 1994)

Letztlich folgt aus der Annahme, dass das Entscheidungsverhalten nicht einhundertprozentig nachgebildet werden kann, dass lediglich Wahrscheinlichkeitsaussagen über die Wahl einer Option getroffen werden. Als Ergebnis liefern die Modelle also Anteilswerte. (Maier/Weiss, 1990) Somit entsteht die Forderung nach bestimmten Eigenschaften, die das Modell abbilden können sollte.

Wichtig ist, dass die Ergebnisse  $P_{ni}$  stets zwischen Null und Eins liegen, um diese als Wahrscheinlichkeit interpretieren zu können. Die Werte Null und Eins selbst sollten jedoch fast nie erreicht werden, Eins höchstens dann, wenn das Choice Set, also der Satz aller möglicher Alternativen, nur aus einer Alternative besteht. Das ist jedoch sehr unüblich, da für einen solchen Fall kein Wahlmodell benötigt wird. Ein Wert von Null ist nicht sinnvoll, denn sollte die Wahrscheinlichkeit, dass ein Verkehrsteilnehmer ein Verkehrsmittel nutzt bei 0% liegen, ist es meist ein Anzeichen dafür, dass dieses Verkehrsmittel für die entsprechende Person nicht zur Verfügung steht und kann im Vorhinein aus dem Choice Set entfernt werden. Weiterhin sollte zu beobachten sein, dass die Auswahlwahrscheinlichkeit  $P_{ni}$  steigt, wenn der deterministische Nutzenanteil steigt. Ebenfalls wichtig ist, dass sich die Auswahlwahrscheinlichkeiten aller Alternativen zu Eins addieren. (Train, 2003) Des Weiteren sollte es mit einem guten Modell möglich sein, auch mehr als zwei unterschiedliche Alternativen darstellen zu können, wobei jedoch der Rechenaufwand möglichst gering gehalten werden soll.

Im Folgenden werden die meist genutzten Modelle beschrieben und auf diese Eigenschaften hin überprüft.

### 3.1 LP-Modell

Das lineare Wahrscheinlichkeitsmodell (LP-Modell von Linear Probability) ist das wohl einfachste Entscheidungsmodell. Hier gilt die Annahme, dass die Differenz der stochastischen Nutzenanteile gleichverteilt ist. Bei einer Anwendung auf mehr als zwei Alternativen müsste dies jedoch auf alle Differenzen zwischen den einzelnen Zufallskomponenten zutreffen. Dies ist jedoch nicht möglich, da es keine Verteilung gibt, die den stochastischen Termen zugrunde gelegt werden kann, sodass sich für deren Differenz immer eine gleichverteilte Zufallsvariable ergibt. Dadurch ist dieses Modell nur für den binären Fall, also die Wahl zwischen zwei Alternativen, plausibel anwendbar. Um zu verhindern, dass die Ergebniswerte, also die Auswahlwahrscheinlichkeiten, kleiner als Null, bzw. höher als Eins werden, wird es auf ein bestimmtes Intervall zwischen  $-L$  und  $+L$  beschränkt. Der Parameter  $L$  ist ein Skalierungsparameter, der im Vorhinein festgelegt wird. Die Auswahlwahrscheinlichkeiten können in einer Verteilungsfunktion abgebildet werden, welche in niedrigen Werten bis einschließlich  $-L$  den Wert Null hat und bei  $L$  und höher Eins beträgt. Dazwischen verläuft die die Funktion linear. (Maier/Weiss, 1990)

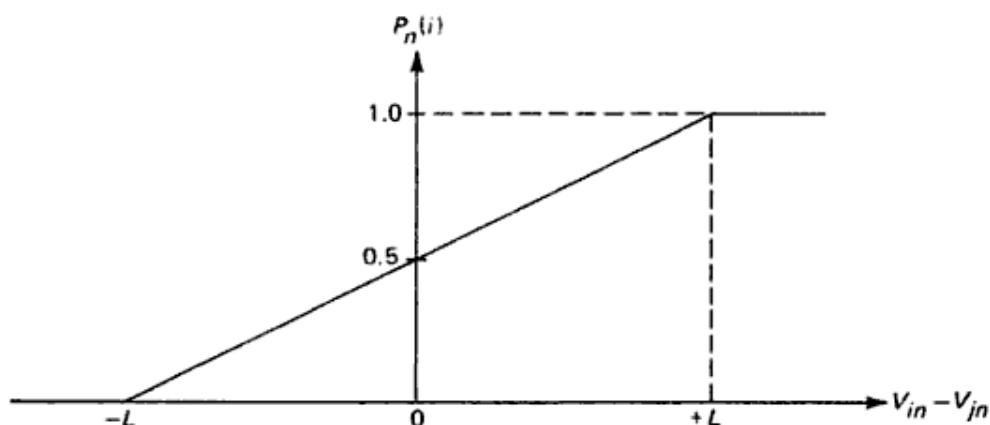


Abbildung 2: Verteilungsfunktion des binären LP-Modells (aus Ben-Akiva/Lerman, 1985)

Die dadurch entstehenden Knickstellen in der Funktion sind in der Realität jedoch nicht plausibel (Maier/Weiss, 1990). Einerseits, weil nicht ausgeschlossen werden



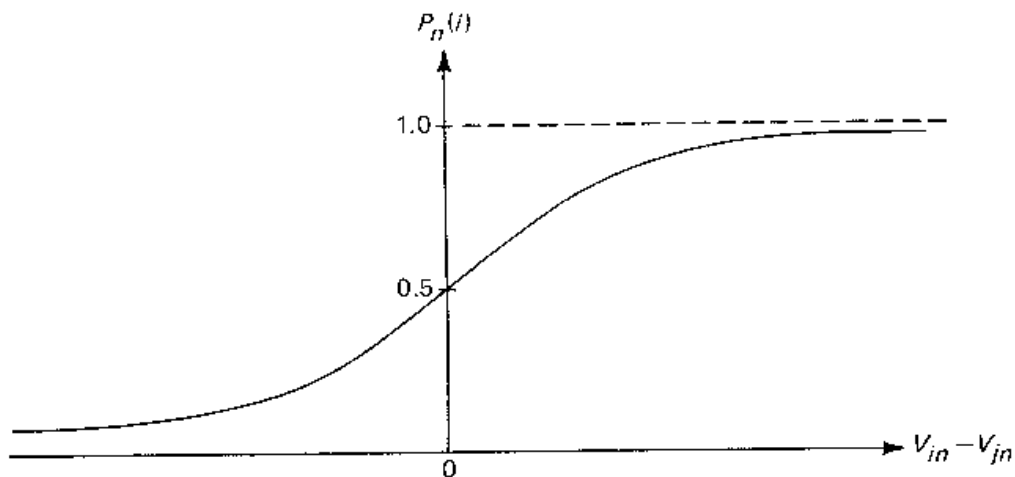
kann, dass einige Personen eine Option wählen, welche vom Modell mit einer Wahrscheinlichkeit von 0% vorausgesagt wird (Ben-Akiva/Lerman, 1985), andererseits, weil somit die erste Ableitung, welche im Verfahren zur Schätzung der Parameter berechnet werden muss, unstetig ist (Maier/Weiss, 1990).

Obwohl die Berechnung hier relativ einfach ist, findet das Modell auf Grund der nur bedingten Eignung für die Modellierung bei einer größeren Alternativenmenge, seiner problematischen theoretischen Fundierung und durch die Existenz von besseren Modellen in der Verkehrsplanung kaum Anwendung (Schnabel/Lohse, 2011).

### **3.2 Probit-Modell**

Im Probit-Modell (aus dem englischen „probability unit“, Einheit der Wahrscheinlichkeit) wird davon ausgegangen, dass die stochastischen Nutzenterme die Nutzenfunktion jeweils unabhängig voneinander und nur geringfügig verändern. Sie müssen außerdem nicht der gleichen Verteilung folgen. Die Gesamtheit der Störterme kann dann als Summe aller einzelnen Störterme nach dem Zentralen Grenzwertsatz durch eine Normalverteilung angenähert werden. Der Erwartungswert kann hier einfach auf Null gesetzt werden, denn sollte er einen Erwartungswert haben, müsste dieser in den deterministischen Teil übernommen werden. Da die i.i.d. Eigenschaft nicht eingehalten wird, erlaubt das Modell, dass die Störterme unterschiedliche Varianzen aufweisen und, dass eine Korrelation unter den einzelnen Zufallsnutzen besteht. (Maier/Weiss, 1990)

Die Verteilungsfunktion weist eine sigmoide Form auf, wobei die Auswahlwahrscheinlichkeit sich der Eins annähert, wenn die Nutzenwerte gegen  $+\infty$  gehen und der Null, wenn sie gegen  $-\infty$  gehen. Da im Wahlmodell nur Differenzen entscheidend sind, wird die Differenz zwischen den Nutzenwerten der Wahlalternativen betrachtet. Ist die Differenz zwischen den Nutzenwerten zweier Alternativen gerade Null, so liegt die Wahlwahrscheinlichkeit jeweils bei 50%. An dieser Stelle befindet sich der Wendepunkt der Funktion.



**Abbildung 3:** Verteilungsfunktion des binären Probit-Modells (aus Ben-Akiva/Lerman, 1985)

Die Normalverteilung gilt mit diesen Eigenschaften als realitätstreueste Darstellung des stochastischen Nutzens. Allerdings ist auch dieses Modell bei mehr als zwei Alternativen nicht mehr praktikabel, da dies analytisch nicht mehr berechnet werden kann. Somit ist es in der Verkehrsplanung nicht sehr verbreitet. (Schnabel/Lohse, 2011) Außerdem kann die flexible Form dieses Modells auch Probleme mit sich bringen, da eventuell systematische Einflüsse nicht als solche erkannt werden und im stochastischen Teil statt im deterministischen berücksichtigt werden. Restriktivere Modelle wie das im Folgenden beschriebene Logit-Modell, bieten hier auch die Möglichkeit der Qualitätsüberprüfung des Modells. (Mayer/Weiss, 1990)

### 3.3 Logit-Modell

Die logistische Verteilung der Störterme  $\varepsilon$  wird beim Logit-Modell unterstellt. Sie hat einen ähnlichen Verlauf wie die Normalverteilung, ist jedoch wesentlich einfacher zu berechnen. Bei der Betrachtung des multinomialen Falls, der Wahl zwischen mehr als zwei Alternativen, wird hier unterstellt, dass allen Störtermen die gleiche Verteilung unterliegt und diese gleiche Mittelwerte und Varianzen aufweisen. Sie sind also im Gegensatz zur Normalverteilung unabhängig und identisch verteilt (i.i.d.). Dies wird durch die Anwendung der Gumbel-Verteilung, eine Näherung an die Normalverteilung, dargestellt. Somit ist dieses Modell auch für den multinomialen Fall analytisch berechenbar.



Dadurch, dass alle Störterme im Logit-Modell unabhängig voneinander sind, dürfen sie nicht durch die Existenz anderer Alternativen beeinflusst werden, das Verhältnis darf nur von den eigenen Attributen der Alternativen bestimmt sein. Diese Eigenschaft wird *Independence from Irrelevant Alternatives* (im Folgenden IIA bezeichnet) genannt. Dies bedeutet, dass das Verhältnis der Auswahlwahrscheinlichkeiten zweier Alternativen unabhängig von der Verfügbarkeit einer Weiteren ist. Dies ist einerseits eine wünschenswerte Eigenschaft eines Entscheidungsmodells, restringiert dieses jedoch andererseits, da es auf manche Fragestellungen mit dieser Eigenschaft nicht mehr anwendbar ist. Ein bekanntes Beispiel hierfür ist das „red bus- blue bus“ Paradoxon<sup>2</sup>. (Maier/weiss, 1990)

Auf Grund seiner Eigenschaften, welche die Berechnung für mehr als zwei Alternativen vereinfachen, ist dies das am häufigsten genutzte Modell für die Verkehrsmittelwahl in der Verkehrsplanung (Schnabel/Lohse, 2011). In den folgenden Kapiteln wird daher genauer darauf eingegangen, wie mit diesem Modell gerechnet werden kann.

In der folgenden Tabelle sind die Vor- und Nachteile der verschiedenen Verkehrsmittelwahlmodelle noch einmal zusammenfassend dargestellt.

---

<sup>2</sup> Das „red bus-blue bus“ Paradoxon beschreibt eine Situation, in der die IIA-Eigenschaft des Logit-Modells unpraktikabel wäre. Bei einer Auswahlwahrscheinlichkeit von  $P(\text{Auto}) = P(\text{Bus}) = 1/2$ , würde man annehmen, dass das Hinzufügen einer weiteren Alternative, beispielsweise eines weiteren Busses, mit, abgesehen von der Farbe, identischen Eigenschaften wie der bereits existierenden, zu den Wahrscheinlichkeiten  $P(\text{Auto}) = 1/2$  und  $P(\text{blauer Bus}) = P(\text{roter Bus}) = 1/4$  führen würde. Das Logit-Modell berechnet als Folge der IIA-Eigenschaft hier jedoch Auswahlwahrscheinlichkeiten von  $P(\text{Auto}) = P(\text{blauer Bus}) = P(\text{roter Bus}) = 1/3$ . (Train, 2003)





	LP	Probit	Logit
Verteilung	linear	normal	logistisch
Eignung $k=2$	ja	ja	ja
Eignung $k>2$	nein	nein	ja
Vorteil	Einfache Berechnung	Realistische Verteilung	Auch für multinomialen Fall berechenbar
Nachteil	Nur im binomialen Fall anwendbar	Nur für binomialen Fall berechenbar	Restriktion durch IIA-Eigenschaft

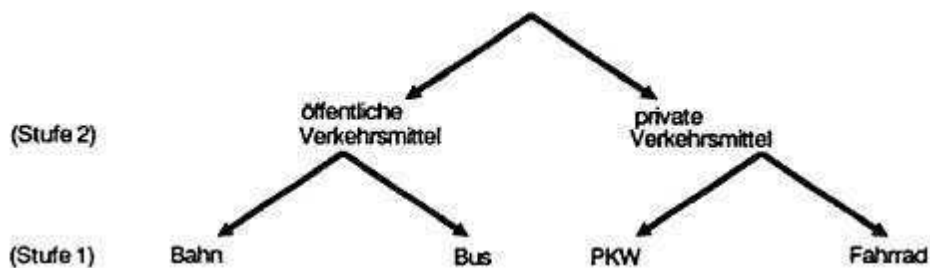
**Tabelle 1:** Eigenschaften der gebräuchlichsten Modelle diskreter Entscheidungen

### 3.4 Weitere Modelle

Als weitere Modelle mit nützlichen Eigenschaften können einige Erweiterungen und Abwandlungen des Logit-Modells genannt werden, die vor allem zum Zweck der Umgehung der IIA-Eigenschaft entwickelt wurden.

Das *Dogit-Modell* (aus dem englischen *to dodge* = ausweichen) von Gaudry und Dagenais (1979) versucht bei einigen Alternativen die IIA-Eigenschaften des Logit-Modells zu umgehen, indem ein weiterer Parameter eingefügt wird. Ist dieser von Null verschieden, wird das Einführen einer weiteren Alternative zu veränderten relativen Auswahlwahrscheinlichkeiten der anderen Alternativen führen. (Gaudry/Dagenais, 1979) Bei einem Wert des Parameters von Null hat dieses die Form des multinomialen Logit-Modells (im Folgenden auch MNL-Modell genannt), also des Logit-Modells mit mehr als zwei Wahlalternativen. Außerdem drückt der Parameter die Loyalität aus, die einige Entscheidungsträger zu einem bestimmten Verkehrsmittel unabhängig von dessen Eigenschaften haben. Je höher der Parameter ausfällt, desto höher auch die Wahrscheinlichkeit, dass der betrachtete Nutzer zum betrachteten Verkehrsmittel eine höhere Loyalität hat. Die Parameterschätzung ist damit immer noch einfacher als im Probit-Modell. (Hensher/Button, 2000)

Des Weiteren ist das *Nested Logit-Modell* eine nützliche Erweiterung. Es trägt seinen Namen, da es die Wahlalternativen in sogenannte Nester einteilt. Die Alternativen, die in einem Nest zusammengefasst werden, sind unabhängig voneinander. Somit gilt die IIA-Eigenschaft innerhalb der Nester. Zwischen den Nestern kann das Verhältnis der Auswahlwahrscheinlichkeiten von der Existenz oder den Eigenschaften anderer Alternativen abhängig sein. Hier wird die IIA-Eigenschaft nicht zwangsläufig eingehalten. Der Entscheidungsprozess kann hier in einem Entscheidungsbaum dargestellt werden.



**Abbildung 4:** Beispiel eines zweistufigen Nested Logit-Modells (aus Urban, 1993)

Letztlich ist das Nested Logit-Modell eine sequentielle Anwendung des MNL-Modells. Die Entscheidung wird hier in mehreren Stufen getroffen. Mit diesem Modell ist beispielsweise auch die Einschränkung darstellbar, dass eine Person, welche für den Hinweg den öffentlichen Verkehr nutzt, den Rückweg meist nicht mit eigenem Pkw zurücklegen kann. (Train, 2003)

Es existieren eine Reihe weiterer Modelle, die versuchen bestimmte Eigenschaften zu umgehen oder zu verknüpfen. Letztlich ist die Auswahl eines Modells vom Anwendungsfall abhängig.

Für die Anwendung im Rahmen des Projektes des DLR empfiehlt sich die Nutzung eines diskreten Wahlmodells. Vor allem bei der Betrachtung von mehr als zwei Alternativen ist das Logit-Modell vorteilhaft. Je nach Anwendungsfall sollte auch eine Abwandlung des Logit-Modells in Betracht gezogen werden.

## 4 Die Logit-Analyse

Im Folgenden wird das am häufigsten genutzte Modell der Verkehrsmittelwahl, das Logit-Modell, schrittweise beschrieben. Hierbei wird zunächst, wie in allen anderen diskreten Wahlmodellen auch, eine Nutzenfunktion bestimmt, welche alle relevanten Variablen beinhaltet. Um den Modal Split, bzw. die Wahrscheinlichkeiten der Wahl eines bestimmten Verkehrsmittels berechnen zu können, wird das Logit-Modell mit Gumbel-verteiltern Zufallsnutzentermen angewandt. Dieses Modell hat die folgende Form. (Maier/Weiss, 1990)

$$P_k = \frac{e^{V_k}}{\sum_{k'=1}^K e^{V_{k'}}} = \frac{1}{\sum_{k' \neq k} e^{V_{k'} - V_k}} \quad (4.1)$$

Hierzu müssen also alle Bewertungen der Einflussfaktoren, die in der Nutzenfunktion vorkommen, bekannt sein. Zunächst wird bestimmt, welche exogenen Variablen in die Nutzenfunktion aufgenommen werden. Im zweiten Schritt werden die benötigten Daten erhoben, um im dritten Schritt mit Hilfe von statistischen Methoden die Parameter der Nutzenfunktion schätzen zu können. Diese geschätzten Parameter werden in Schritt Vier überprüft. Somit kann eine Aussage darüber getroffen werden, ob das Modell die Realität gut abbilden kann. Der Modal Split wird im letzten Schritt mit dem kalibrierten Modell berechnet.

### 4.1 Aufstellen der Nutzenfunktion

Aufgabe im ersten Schritt ist es, die entscheidenden Determinanten der Verkehrsmittelwahl zu identifizieren und diese sinnvoll zu einer Nutzenfunktion zu verknüpfen.

Wie im vorherigen Kapitel erwähnt, gibt es verschiedene Determinanten die die Verkehrsmittelwahl von Personen oder Unternehmen beeinflussen. Diese gilt es zunächst herauszufinden. Trivial ist es, die Reisezeit und die Reisekosten in die

---

<sup>3</sup> Korrekterweise gehört zur Gumbel-Verteilung der Skalierungsparameter  $\lambda$ , welcher die Nutzenwerte  $V_k$  skaliert. Da das Logit-Modell jedoch skaleninvariant ist, kann dieser Parameter a priori auf 1 gesetzt werden und entfällt somit in der Gleichung, da er multiplikativ mit dem Nutzen verknüpft ist. (Maier/Weiss, 1990)



Funktion aufzunehmen. Zusätzlich können andere Charakteristika einbezogen werden. Auch kann eine Aufteilung der komplexen Reisezeit sinnvoll sein, wenn im späteren Verlauf beispielsweise Einflüsse einzelner Zeiten, wie zum Beispiel der Zugangs-, Abgangs- oder Wartezeit auf das Wahlverhalten identifiziert werden sollen. Zusätzlich zu den Charakteristika des Verkehrsangebots sollten immer auch sozioökonomische Variable und alternativenspezifische Konstanten in die Funktion aufgenommen werden.

Der Einfluss einer Variablen auf den Gesamtnutzen ist an den jeweiligen Vorzeichen zu erkennen. Sinkt der Nutzen beispielsweise mit steigender Reisezeit, so trägt der Gewichtungsfaktor zur Reisezeit ein negatives Vorzeichen. (Urban, 1993) Auf die Schätzung der Gewichtungsfaktoren sowie die Interpretation der Parameter wird im Späteren genauer eingegangen.

Die Verknüpfung der Variablen mit den Gewichtungsparametern kann unterschiedlicher Gestalt sein. Am häufigsten wird die multiplikative Verknüpfung zwischen Variable und Parameter ( $\beta_1 \cdot X_1$ ) gewählt. Andere Arten, wie zum Beispiel die exponentielle Verknüpfung ( $X_1^{\beta_1}$ ), sind jedoch ebenfalls denkbar. Genauso muss über die Verknüpfung zwischen Teilnutzen und Gesamtnutzen entschieden werden. Hier wird meist die additive Form gewählt. Auch über die Linearität der Nutzenfunktion, welche in Kapitel 2 diskutiert wurde, wird in diesem Schritt entschieden. (Knapp, 1998)

## **4.2 Datenerhebung**

Im zweiten Schritt werden die relevanten Daten durch Befragungen und Messungen erfasst. Mit Hilfe von statistischen Methoden werden die Daten dann aufbereitet.

Die benötigten Daten können aus verschiedenen Quellen stammen. Während sozioökonomische Charakteristika aus dem Melderegister entnommen oder erhoben werden, können Eigenschaften der Infrastruktur meist direkt gemessen werden bzw. aus Abbildungen des Straßen- und Schienennetzes ermittelt werden. Um im späteren Verlauf die Parameter zur Bewertung einzelner Faktoren berechnen zu können, sind Befragungen notwendig. In Deutschland sind zwei



Erhebungsmethoden nennenswert, zum einen das *System repräsentativer Verkehrsbefragungen* und außerdem die *kontinuierliche Verkehrserhebung*.

Das *System repräsentativer Verkehrsbefragungen (SrV)* wurde auf dem Gebiet der neuen Bundesländer in regelmäßigen Abständen von fünf Jahren durchgeführt. (Lohse/Schneider, 1997) Die Befragung trägt auch den Namen „Mobilität in Städten“, da ausschließlich größere Städte untersucht werden. Sie erfolgte bis 1990 nur auf dem Gebiet der ehemaligen DDR, seitdem in der gesamten Bundesrepublik. Die Teilnehmer erhalten jeweils drei Fragebögen. Im Personenfragebogen werden verschiedene sozioökonomische Merkmale identifiziert, besonders von Interesse sind der Stand im Lebenszyklus, eventuelle Mobilitätseinschränkungen sowie Gewohnheiten im Mobilitätsverhalten. Der Haushaltsfragebogen ordnet den Haushalt nach verkehrsrelevanten Gesichtspunkten ein und erhebt Faktoren wie die Jahresfahrleistung, die Lagegunst des Haushaltes und die Verfügbarkeit von motorisierten Verkehrsmitteln oder Zeitkarten für den öffentlichen Verkehr. Im Wegefragebogen sollen die einzeln getätigten Wege eines Stichtages genau beschrieben werden, dazu gehören unter anderem Verkehrsmittel, Dauer und Wegezweck. (Ließke, 2011)

Als *KONTIV* wurde die „*kontinuierliche Verkehrserhebung*“ auf dem Gebiet der Bundesrepublik bezeichnet, nach der Überarbeitung wird sie heute *Mobilität in Deutschland* (kurz *MiD*) genannt und wird im Auftrag des Bundesministeriums für Verkehr, Bau und Stadtentwicklung durchgeführt (Lohse/Schneider, 1997). In größeren Zeitabständen als *SrV* werden ebenfalls Merkmale der Haushalte, der Personen und der Wege erhoben, sowie weitere Daten zu Reisen und zum PKW des Haushaltes (infas GmbH, 2010).

Für die Messung der Präferenzen gibt es verschiedene Befragungsarten. Die oben beschriebenen Erhebungsarten ermitteln *revealed preferences* (aus dem englischen „offenbarte Präferenzen“). Hierbei wird tatsächliches Verhalten analysiert. Aus den *Mobilitätstagebüchern*, welche die Testpersonen in den oben genannten Erhebungen führen, werden die real getätigten Entscheidungen abgefragt und analysiert. (PTV AG, 2012) Diese Befragungen sind im Allgemeinen sehr realitätstreu, jedoch auch sehr aufwendig. Der Nachteil besteht außerdem darin, dass manche Entscheidungen schwer auf bestimmte Einflussfaktoren



zurückzuführen sind. Auch die Akzeptanz neuer Optionen, welche in der Realität noch nicht existieren, kann nicht geprüft werden. (Ortúzar, 1994)

Andere Erhebungen werden mit der *stated preference* Methode (aus dem englischen „bekundete Präferenz“) durchgeführt. Hier werden die Versuchspersonen in eine hypothetische Entscheidungssituation versetzt und nach ihren Präferenzen gefragt. Dabei ist jedoch nicht auszuschließen, dass Personen die eine bestimmte Präferenz bekunden, sich in der Realität anders entscheiden würden. (Louviere/Street, 2000) In solchen Befragungen sollen die Auskunftspersonen verschiedene Auswahlmöglichkeiten mit unterschiedlichen Eigenschaften, sog. Choice Sets, in eine Rangfolge bringen, diese auf einer Skala bewerten oder schlicht die Frage beantworten, welche der vorgeschlagenen Alternativen sie wählen würden<sup>4</sup>. Die Choice Sets sollten dabei möglichst so gewählt werden, dass der Einfluss der einzelnen Eigenschaften eines Verkehrsmittels auf die Entscheidung der Person mit großer Genauigkeit untersucht werden kann.

Zur Analyse dieser Art von Befragung wird die *Conjoint-Analyse* (aus dem englischen „considered jointly“, ganzheitliche betrachtet) herangezogen. Sie dient dazu, den Beitrag einzelner Komponenten zu einem Gesamtnutzen zu ermitteln (Backhaus, 1996) und entspricht den Schritten der Formulierung, Schätzung und Diskussion eines Logit-Modells.

### **4.3 Schätzung der Parameter**

Aus den erhobenen Daten müssen nun die Gewichtungparameter der Nutzenfunktion geschätzt werden. Ziel ist es, aus den Daten der Stichprobe eine möglichst genaue Abbildung der Grundgesamtheit zu gewinnen.

Für die Schätzung der Parameter müssen einige Voraussetzungen erfüllt sein, damit das Ergebnis möglichst realitätsnah ausfällt. Die Stichprobe sollte groß

---

<sup>4</sup> Streng genommen wird bei der *stated preference* Methode die Alternative auf einer Skala bewertet. Bei der Auswahl einer Alternative von mehreren spricht man von *stated choice* und bei der Ordnung in eine Rangfolge von *stated ranking*. Diese beiden Typen sind Untertypen der *stated preference* Methode und werden deshalb hier zusammengefasst. Meist wird die *stated choice* Methode angewandt. (Axhausen, 2003)



genug sein, damit Verzerrungen vermieden werden. Generell gilt, je größer eine Stichprobe ist, desto weniger weicht sie von der Realität ab. Diese Voraussetzung wird Konsistenz der Modell-Schätzung genannt. (Urban, 1993) Außerdem sollte sie erwartungstreu und effizient sein. Erwartungstreu ist ein Schätzwert für einen Parameter dann, wenn sein Erwartungswert dem wahren Wert des Parameters entspricht. Effizient ist die Modell-Schätzung, wenn der Schätzer für den Parameter eine geringere Varianz aufweist, als ein Schätzer, der mit einem anderen Verfahren geschätzt wurde. Bei jeder Schätzung muss also geprüft werden, ob das angewandte Verfahren die wirklich besten Ergebnisse liefert. (Maier/Weiss, 1990) Im Folgenden werden zwei Verfahren zur Schätzung von Parametern vorgestellt, die Methode der kleinsten Quadrate und die Maximum-Likelihood-Methode.

Die Methode der kleinsten Quadrate nutzt als Schätzkriterium die Summe der quadrierten Differenzen zwischen den beobachteten und den vorhergesagten Werten der endogenen Variablen. Für die Modellfunktion, welche im Vorhinein modelliert wurde, werden die Parameter so gewählt, dass sie so wenig wie möglich von den tatsächlichen Werten abweichen. Als gutes Schätzergebnis muss also die Differenz zwischen Beobachtung und geschätzter Funktion minimiert werden. Mit den Quadraten der Abweichungen, der sog. Fehlerquadratsumme, wird gerechnet, weil somit größere Abweichungen stärker bestraft werden als kleine. Für die Nutzung dieses Schätzverfahrens müssen jedoch einige Voraussetzungen erfüllt sein. So sollen die Abweichungen im Durchschnitt Null sein, eine gleichbleibende Varianz haben und stochastisch unabhängig von anderen Abweichungen sein. Dies setzt eine Normalverteilung der Fehler voraus. Außerdem ist die Methode bei linear spezifizierten Modellen mit qualitativen, abhängigen Variablen, wie es bei der Logit-Analyse zur Bestimmung des Modal Split der Fall ist, nicht anwendbar (Urban, 1993). Die Anwendung der Methode der kleinsten Quadrate wäre bei einem linearen oder quasilinearen Probit-Modell mit i.i.d. gaußverteilten Störtermen (Treiber, 2011) praktikabel. Sie ist in einem Logit-Modell jedoch nicht sinnvoll, hier wäre lediglich eine gewichtete Kleinste-Quadrate-Schätzung möglich. (Urban, 1993)

Eine bessere Methode zur Schätzung der Parameter des Logit-Modells, welches sowohl aus qualitativen als auch aus quantitativen Variablen besteht (Eckey,



2001), ist die Maximum-Likelihood-Schätzung (aus dem englischen „größte Wahrscheinlichkeit“, im Folgenden auch mit ML abgekürzt). Diese wählt die Koeffizienten als Schätzwerte aus, welche am wahrscheinlichsten aus der Stichprobe resultieren würden, wären sie identisch mit den Koeffizienten der Grundgesamtheit. (Urban, 1993) Die Wahrscheinlichkeit dafür, dass das Modell mit den Messungen übereinstimmt soll möglichst groß sein. Die Bedingung für die Schätzung der Parameter  $\beta$  lautet also: „Die kombinierte Wahrscheinlichkeit dafür, dass alle prognostizierten absoluten Häufigkeiten gleich den beobachteten Häufigkeiten sind, soll maximal sein.“ (Treiber, 2011)

$$\hat{\beta} = \arg \max_{\beta} [P_{\text{komb}}(\hat{Y}_{ki} = y_{ki}; \beta)] \quad (4.2)$$

Mit  $\hat{\beta}$ , dem geschätzten Parameter  $\beta$   
 $\hat{Y}_{ki}$ , den prognostizierten absoluten Häufigkeiten der Wahl von Alternative k in Personengruppe i  
 $y_{ki}$ , den beobachteten absoluten Häufigkeiten der Wahl von Alternative k in Personengruppe i, welche den Befragungsdaten zu entnehmen sind

Aus den Befragungsdaten werden also im Modell Wahrscheinlichkeitsaussagen über die tatsächlichen Wahlhäufigkeiten abgeleitet. Hierbei ergibt sich der Zufallsvektor  $\hat{Y}_{ki}$  der prognostizierten Wahlhäufigkeiten aus der Multinomialverteilung. Diese ist eine Verallgemeinerung der Binomialverteilung und findet Anwendung, wenn bei der Entscheidung mehr als zwei Alternativen zur Wahl stehen.

Da die Annahme gilt, dass die Individuen unabhängig voneinander entscheiden und somit die einzelnen Zufallsnutzen unabhängig sind, können die einzelnen Multinomialverteilungen multipliziert werden. Daraus ergibt sich die Likelihoodfunktion. (Maier/Weiss, 1990)

$$L(\beta) = \prod_i B_i \prod_k (P_{ki})^{y_{ki}} \quad (4.3)$$

Mit  $B_i$ , dem Multinomialkoeffizienten ( $B_{(y_i)} = \frac{(n_i)!}{\prod_k (y_{ki})!}$ )





Es gilt nun die Parameter  $\beta$  zu suchen, welche die Funktion  $L(\beta)$  maximieren. Da sich mathematisch besser mit Summen als mit Produkten rechnen lässt, wird die obige Gleichung zunächst logarithmiert, sodass sich die Log-Likelihood-Funktion ergibt. Dies ist unproblematisch, da das Maximum der logarithmierten Funktion an der gleichen Stelle liegt wie bei der nicht logarithmierten. (Urban, 1993)

$$LL(\beta) = \sum_i \sum_k y_{ki} \ln P_{ki}(\beta) \quad (4.4)$$

Da der Multinomialkoeffizient  $B_i$  nur eine Konstante ist, welche nicht von den Parametern abhängt, wird er hier weggelassen (Treiber, 2011). Nun kann durch differenzieren und Null setzen der Gleichung der Extremwert errechnet werden. Da die Funktion konkav ist, ergibt die Ableitung das Maximum. Im Zweifelsfall kann dies mit Hilfe der zweiten Ableitung überprüft werden. (Maier/Weiss, 1990)

$$\frac{\partial LL(\beta)}{\partial \beta} = 0 \quad (4.5)$$

Die Parameter  $\beta$  werden aus dem sich ergebenden Gleichungssystem bestimmt. (Eckey, 2001)

$$\sum_k \sum_i (y_{ki} - P_{ki}) \cdot x_{ki} = 0 \quad (4.6)$$

Es ergeben sich also so viele Gleichungen wie es Parameter  $\beta$  gibt. Da diese Gleichungen jedoch nicht linear sind, können die Parameter  $\beta$  nicht analytisch bestimmt werden. Es ist ein numerisches Verfahren nötig, das sich dem ML-Schätzwert annähert (Eckey, 2001). Das am häufigsten verwendete iterative Näherungsverfahren ist der sog. *Newton-Raphson-Algorithmus*. (Urban, 1993) Dieser versucht, ausgehend von einem Startwert, sich in wenigen Schritten an den ML-Schätzwert anzunähern. (Maier/Weiss, 1990)

Dieses Newton'sche Iterationsverfahren geht von der Schätzbedingung aus, dass die Merkmalssummen  $X$  des Modells gleich der Daten  $X$  der Stichprobe sind. Die Differenz mit dem richtigen Schätzer für den Parameter  $\beta$  soll also Null sein.

$$F_j(\beta) = X_j^{Modell} - X_j^{Daten} \quad (4.7)$$



$$F(\hat{\beta}) = 0 \quad (4.8)$$

Konkret werden die Funktionen  $F_j$  für die Logit-Modelle wie folgt berechnet:

$$F_j(\beta) = -\frac{\partial LL(\beta)}{\partial \beta_j} = \sum_{ki} (n_i P_{ki}(\beta) - y_{ki}) X_{jki}. \quad (4.9)$$

Wobei  $n_i$  die Anzahl der Wahlentscheidungen in Personengruppe  $i$  ist,  
 $P_{ik}$  die Wahlwahrscheinlichkeit der Alternative  $k$  in Personengruppe  $i$ ,  
 welche sich im Logit-Modell aus  $\frac{e^{V_{ik}}}{\sum_k e^{V_{ki}}}$  ergibt  
 $y_{ki}$  die beobachtete Häufigkeit der Wahl von Alternative  $k$  in der Personengruppe  $i$  der Stichprobe  
 $X_{jki}$  der betrachtete Einflussfaktor der Wahlentscheidung für Alternative  $k$  der Personengruppe  $i$

Schon im ersten Schritt müssen Werte für die Parameter  $\beta$  in die Formel eingesetzt werden, da die Wahlwahrscheinlichkeit  $P_{ki}$  von diesem abhängt. Hier wird zunächst ein beliebiger Startwert angenommen. In den folgenden Iterationsschritten berechnen sich die Funktionen  $F$  dann aus

$$F(\beta) \approx F(\hat{\beta}) + J \cdot (\beta - \hat{\beta}) = J \cdot (\beta - \hat{\beta}). \quad (4.10)$$

Hierbei ist  $J$  eine symmetrische Funktionalmatrix, die sich aus den partiellen Ableitungen der Funktionen  $F$  nach den Parametern  $\beta$  ergibt. Sie trägt den Namen  $J$ , weil sie gemeinhin auch als Jacobi-Matrix bekannt ist. Die einzelnen Elemente ergeben sich aus

$$J_{jm} = \sum_{ik} n_i P_{ki} x_{jki} (x_{mki} - \sum_{k'} x_{mk'i} P_{k'i}). \quad (4.11)$$

Im binomialen Fall, in dem die Alternativenmenge  $K$  nur aus zwei Alternativen besteht vereinfacht sich die Formel zu

$$J_{jm} = \sum_i n_i x_{j1i} x_{m1i} P_{1i} P_{2i}. \quad (4.12)$$

Um den Parameter  $\beta$  des nächsten Iterationsschrittes berechnen zu können wird Gl. 4.10 nach diesem aufgelöst und es muss mit dem Inversen der Matrix gerechnet werden:



$$\beta^{(i+1)} = \beta^{(i)} - (J(\beta^{(i)}))^{-1} \cdot F(\beta^{(i)}). \quad (4.13)$$

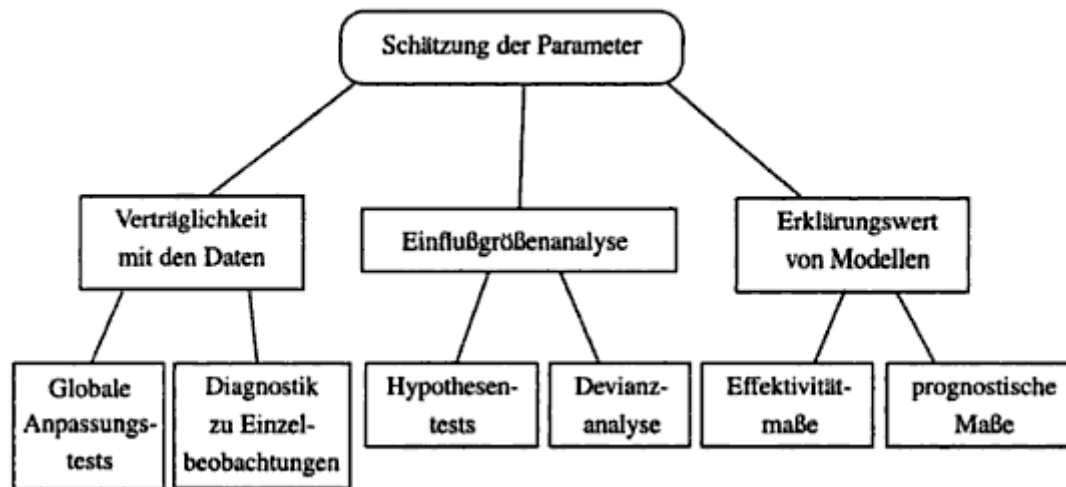
- Hierbei ist  $\beta^{(i)}$  ein Vektor mit allen Parametern  $\beta$  der i-ten Iteration,  
 $J$  besitzt  $X \cdot X$  Elemente, wobei  $X$  die Anzahl an Einflussfaktoren, also exogenen Variablen des Modells ist.  
 $F$  wird ebenfalls als Vektor der Funktionen der Schätzbedingungen geschrieben.

Mit beliebigen Startwerten für alle  $\beta$  wird also zunächst eine Startlösung kreiert. In den folgenden Schritten werden dann neue  $\beta$  berechnet, welche sich immer näher dem ML-Schätzwert annähern. Zur Überprüfung, wie nah der Schätzwert dem ML-Wert ist, wird die Gleichung (4.7) berechnet. Diese nähert sich schon nach wenigen Iterationen der Null an. Ist sie ausreichend nahe der Zahl Null, ist das Abbruchkriterium erreicht, das Verfahren kann beendet werden und die Parameter sind gefunden. (Treiber, 2011)

#### 4.4 Überprüfung

Im nun folgenden Schritt wird die Güte des Modells bewertet. Dazu können verschiedene Aspekte evaluiert werden. Zum einen soll überprüft werden, ob das Modell die Daten gut repräsentiert. Hierzu werden Anpassungstests durchgeführt. Des Weiteren ist es möglich, Aussagen über die Relevanz einzelner im Modell enthaltener Einflussgrößen zu treffen. Mit Hilfe von Hypothesentests werden Nullhypothesen formuliert und es wird mit Teststatistiken überprüft, ob bestimmte Einflussgrößen relevant sind oder unterschiedlich stark wirken. Auch die Effektivität bzw. der Erklärungswert dient zur Bewertung von Modellen. Die hierzu ermittelten Effektivitätsmaße prüfen, ob die untersuchten Einflussfaktoren Änderungen der endogenen Variablen erklären. Prognostische Maße verdeutlichen, wie gut die prognostische Aussagekraft eines Modells ist. (Tutz, 2000)

Die verschiedenen Instrumente sind in Abbildung 5 noch einmal in einer Übersicht dargestellt. Von jeder der drei Arten die Modellschätzung zu überprüfen, wird im Folgenden ein Instrument vorgestellt.



**Abbildung 5:** Übersicht über Instrumente der Modellevaluation (vgl. Tutz, 2000)

Zur Kategorie „Verträglichkeit mit den Daten“, bei der zu testen ist, ob das geschätzte Modell eine gute Anpassung besitzt, kann das intuitive Vorgehen der *Bewertung des Prognoserfolgs* verwendet werden. Dieses überprüft, ob das Modell die Wahl der Testpersonen richtig prognostizieren kann. Für jede Person wird also die Wahrscheinlichkeit vorausgesagt, mit der diese ein bestimmtes Verkehrsmittel wählen würde. Im binomialen Fall gilt die Option als gewählt, sobald die Wahrscheinlichkeit größer als 50% ist. Sollten drei Alternativen zur Verfügung stehen, sinkt der Erwartungswert auf 33% und so weiter. Zu vergleichen ist die Vorhersage mit dem Tatsächlichen Wahlergebnis. Je höher der Prozentsatz der korrekt prognostizierten Ergebnisse, desto besser das Modell. (Urban, 1993)

Für die Einflussgrößenanalyse ist der *Likelihood-Ratio-Test* ein häufig genutztes Instrument. Dieses dient zur Überprüfung, ob das Modell auch in restringierter Form, also mit weniger Parametern, noch aussagekräftig ist. (Eckey, 2001) In diesem Verfahren, welches speziell auf die Maximum-Likelihood-Methode zugeschnitten ist (Maier/Weiss, 1990), wird die Nullhypothese getestet, welche postuliert, dass die zusätzlichen Parameter des unrestringierten Modells irrelevant sind (Tutz, 2000).

$$\beta_2 = \beta_3 = \dots = \beta_j = 0 \quad (4.14)$$



Die Nullhypothese lautet also „einige Koeffizienten der exogenen Variablen sind gleich Null“. Um diese Hypothese zu testen, wird das Modell nun zweimal geschätzt. Eine Schätzung erfolgt mit den exogenen Variablen, die andere ohne diese, da sie, wenn die Koeffizienten Null wären, entfielen. Der Wert der Test-Statistik lautet

$$\text{Test-Variable} = 2 * (\max(LL(\beta_{\text{alle}})) - \max(LL(\beta_{\text{restringiert}}))). \quad (4.15)$$

Wobei  $\max(LL(\beta_{\text{alle}}))$  das Maximum der Log-Likelihood-Funktion mit allen Parametern  $\beta$  ist, also der unrestringierte Fall  
 und  $\max(LL(\beta_{\text{restringiert}}))$  das Maximum der Funktion, wobei einige Terme, dessen Parameter in 4.14 auf Null gesetzt wurden, weggelassen werden.

Zu vergleichen ist dieser Wert mit dem kritischen  $\chi^2$ -Wert, welcher der Tabelle der  $\chi^2$ -Verteilung (siehe Anhang) zu entnehmen ist. Dabei ist  $\alpha$  die Fehlerwahrscheinlichkeit und die Anzahl an Freiheitsgraden ergibt sich aus der Anzahl der exogenen Variablen, die bei der zweiten Schätzung entfallen sind. (Tutz, 2000)  
 Bei

$$\chi^2_{1-\alpha, J-J_r} < \text{Test-Variable} \quad (4.16)$$

wird die Nullhypothese abgelehnt, was bedeutet, dass das unrestringierte Modell die Daten signifikant besser beschreibt als das Restringierte (Treiber, 2011). Anderenfalls würde die Nullhypothese angenommen und das Modell kann durch Weglassen der Parameter vereinfacht werden.

Der Erklärungswert des Modells lässt sich mit Hilfe von Indices quantifizieren. Ein Index der Modellqualität ist der *Likelihood-Ratio-Index*  $\rho^2$ <sup>5</sup>. Dieser sagt aus, wie gut das Modell die Daten repräsentiert (Train, 2003). Der Index berechnet sich aus

$$\rho^2 = 1 - \frac{LL(\hat{\beta})}{LL(0)}, \quad (4.17)$$

---

<sup>5</sup> Der Index wird als  $\rho^2$  bezeichnet, da er Ähnlichkeiten mit dem Bestimmtheitsmaß  $R^2$  aus dem linearen Regressionsmodell aufweist. (Maier/Weiss, 1990) Häufig wird er als Pseudo- $R^2$  bezeichnet (Urban, 1993). Die Quadrierung hat hier keine Bedeutung, in mancher Fachliteratur wird sie deshalb weggelassen. (vgl. Train, 2003)



Mit  $LL(\hat{\beta})$ , der Log-Likelihood-Funktion mit geschätzten Parametern  $\beta$   
 und  $LL(0)$ , der Log-Likelihood-Funktion bei der alle Parameter gleich Null  
 gesetzt sind.

Sollte der Quotient

$$\frac{LL(\hat{\beta})}{LL(0)} = 1 \quad (4.18)$$

ergeben, bedeutet dies, dass

$$LL(\hat{\beta}) = LL(0) \quad (4.19)$$

und somit

$$\rho^2 = 0. \quad (4.20)$$

Das ist der niedrigste Wert den der Index  $\rho^2$  annehmen kann. Mit diesem Ergebnis sagt er aus, dass das geschätzte Modell nicht besser ist als kein Modell zu haben<sup>6</sup>. Dies wäre also das schlechteste Ergebnis. Im Gegensatz dazu kann es der Fall sein, dass durch das geschätzte Modell alle Entscheidungen genauestens abgebildet werden können. Somit wäre die Likelihood-Funktion der geschätzten Parameter gleich Eins und die Log-Likelihood-Funktion gleich Null. Daraus ergäbe sich für  $\rho^2=1$ , dessen bestmöglicher Wert. (Train, 2003) Üblicherweise liegen jedoch die Werte des Pseudo- $R^2$  deutlich niedriger als die des Bestimmtheitsmaßes  $R^2$ . Hier kann schon bei Größen von 0,2 bis 0,4 von einer guten Modellschätzung gesprochen werden (Urban, 1993). Ähnlich wie bei dem Bestimmtheitsmaß existiert auch hier das Problem, dass der Index mit steigender Zahl an Variablen nicht fällt. Somit entsteht eine falsche Genauigkeit, da die Anzahl an Freiheitsgraden nicht einbezogen wird. Korrigiert werden kann dies, indem der Index in folgender Form berechnet wird:

$$\rho_{korr}^2 = 1 - \frac{LL(\beta) - \text{Koeffizientenzahl}}{LL(0)}. \quad (4.21)$$

---

<sup>6</sup> Da ein Modell ohne Parameter genauso gut ist wie kein Modell (Train, 2003)



Somit kann der Index nur dann steigen, wenn der Beitrag der Variablen den Verlust an Freiheitsgrade ausgleicht. (Tutz, 2000)

#### **4.5 Bestimmung des Modal Split**

Im letzten Schritt kann nun mit dem geschätzten und validierten Modell auf die Grundgesamtheit geschlossen und der Modal Split berechnet oder vorhergesagt werden. Dazu werden die vorher berechneten Parameter verwendet und als Variablenwerte diejenigen, welche auf die betrachtete Relation zutreffen, eingesetzt. Daraus ergeben sich die Wahrscheinlichkeiten für die Wahl eines bestimmten Verkehrsmittels. Diese Anteilswerte können direkt als Modal Split interpretiert werden, da der Modal Split als Anteil der Verkehrsmittel am gesamten Verkehrsaufkommen definiert wird.

Hierbei muss nun wieder aggregiert werden, da die personenspezifischen Variablen immer für eine bestimmte Personengruppe gelten. Die Personen werden also zu verhaltenshomogenen Gruppen zusammengefasst und für jede dieser Gruppen  $i$  entsteht eine Nutzenfunktion mit den vorab geschätzten Parametern.

In diesem Schritt der Logit-Analyse werden also die Auswahlwahrscheinlichkeiten der verschiedenen Verkehrsmittel mit dem im vorherigen geschätzten Logit-Modell berechnet. Dabei sind sowohl die Nutzenfunktion als auch die korrekt geschätzten Parameter wichtig. Beim Aggregieren, um den Modal Split auf einer Relation vorhersagen zu können, werden außerdem Daten über die Reisenden auf dieser Relation benötigt.



## 5 Prognose von Änderungen im Modal Split

Mit Hilfe der im vorherigen Kapitel berechneten und kalibrierten Modellgleichung sollen nun prognostische Aussagen getroffen werden. Um beispielsweise Aussagen über die Wirkung von Verkehrsmaßnahmen treffen zu können, ist es von großer Bedeutung die Ergebnisse der Modellkalibrierung korrekt zu interpretieren. Hierzu ist es nützlich die berechneten Parameter genauer zu untersuchen. Aus ihnen können durch Berechnung von Elastizitäten oder durch Bildung der Ableitungen auch Aussagen über das Ausmaß der Änderung in Bezug auf den Modal Split getroffen werden. Die folgenden Abschnitte geben einen Einblick über die Vorgehensweise.

Da es in der vorliegenden Arbeit um Veränderungen im Modal Split durch Infrastrukturmaßnahmen geht, werden ausschließlich Änderungen im Verkehrsangebot, welche sich aus solchen Maßnahmen ergeben, betrachtet. Gegenstand aktueller Forschung sind häufig sog. *soft policy* Maßnahmen, welche das Ziel verfolgen, die Einstellungen der Personen bezüglich bestimmter Verkehrsmittel oder deren Attribute zu ändern (Brög, 1997). Dies würde sich dann auf die Parameter des Modells auswirken. Im vorliegenden Fall sind jedoch Änderungen in den exogenen Variablen von Bedeutung. Ebenfalls muss beachtet werden, dass nicht nur die Änderungen im Verkehrsangebot zu Änderungen der Nachfrage führen, sondern auch sozi-demographische und wirtschaftliche Faktoren einen Einfluss auf die Verkehrsmittelwahl haben. Zunächst sollte daher für jeden Verkehrsträger eine Prognose über das angebotsunabhängige Nachfragewachstum gemacht werden. Somit bilden die im Folgenden beschriebenen Berechnungen nur noch Nachfrageveränderungen ab, welche aus der Änderung des Angebots resultieren. (Vrtic/Axhausen, 2003)

### 5.1 Interpretation der Parameter

Die aus der Modell-Schätzung errechneten Parameter geben die Einflussstärken der zugehörigen unabhängigen Variablen auf die abhängige Variable wieder. Wächst also die Variable  $X_1$  um eine empirische Einheit, so steigt der Wert der endogenen Variable in Logitform um den Betrag den der Parameter  $\beta_1$  hat, welcher die Variable  $X_1$  gewichtet. Weist der Parameter ein negatives Vorzeichen





auf, so sinkt die endogene Variable um den entsprechenden Betrag. Zu beachten ist jedoch die Skalierung der jeweiligen Variablen. Ein bloßer Vergleich der Parameter, um die Aussage treffen zu können, welche Variable einen größeren Einfluss auf die endogene Variable hat, ist zunächst nicht möglich. Ändert man die Messskala einer Variablen (z.B. von km auf m) in einem Modell, so kann es vorkommen, dass die Parameter unterschiedlich groß geschätzt werden. Vergleichbar werden die Parameter erst nachdem man mit Hilfe von statistischen Methoden, wie beispielsweise der t-Statistik, solche Skalenverzerrungen entfernt hat. (Urban, 1993)

$$t - \text{Statistik} = \beta_j / s_j \quad (5.1)$$

mit  $s_j$  der Standardabweichung des Logit-Koeffizienten  $\beta$

Ein Vorteil der Modellparameter ist außerdem, dass sie als Einheiten den Kehrwert der Einheiten der zugehörigen Variablen tragen. Somit können aus ihnen direkt einige Aussagen getroffen werden. Beispielsweise kann der Zeitwert der

Personengruppe mit dem Quotient aus  $\frac{\beta_{\text{Zeit}}}{\beta_{\text{Kosten}}} = \frac{\frac{1}{h}}{\frac{1}{\epsilon}} = \frac{\epsilon}{h}$  berechnet werden.

## 5.2 Ableitungen

Als mathematischer Ausdruck der relativen Änderung der abhängigen Variablen durch die Änderung der unabhängigen Variablen sind die Ableitungen bekannt. Die partielle Ableitung nach einer beliebigen X-Variable gibt also den Einfluss dieser auf die endogene Variable, welche im vorliegenden Fall die Auswahlwahrscheinlichkeit einer Person für bestimmtes Verkehrsmittel  $P_{ki}$  ist, wieder.

$$\Delta P_{ji} = \frac{\partial P_{ki}}{\partial x_{ji}} = \beta_j P_i (1 - P_i) \quad (5.2)$$

Da die Ableitung jedoch die mittlere, prozentuale Veränderungsrate der endogenen Variablen als eine Reaktion auf Skalensprünge der exogenen Variablen von einer empirischen Einheit beschreibt, ist sie von der jeweiligen Skala der exogenen Variablen abhängig. Das heißt, dass die Ergebnisse nicht direkt verglichen werden können, sondern zunächst interpretiert bzw. vereinheitlicht



werden müssen. Hierzu ist die im Folgenden beschriebene Elastizität ein gutes Maß. (Urban, 1993)

### 5.3 Elastizitäten

Die Elastizität gibt an, um wie viel Prozent sich die endogene Variable ändert, wenn sich eine Einflussvariable um ein Prozent ändert (Maier/Weiss, 1990). Elastizitäten werden in der Ökonomie bevorzugt verwendet, da sie nicht von der Skala der exogenen Variablen abhängig sind (Train, 2003).

$$E_{ji} = \frac{x_{ji}}{P_i} \frac{\partial P_i}{\partial x_{ji}} \quad (5.3)$$

Durch Umformen ergibt sich, sofern der deterministische Nutzen in den Variablen  $x_{ji}$  linear ist, mit den Parametern  $\beta_j$  die folgende Gleichung: (Train, 2003)

$$E_{ji} = \beta_j x_{ji} P_i (1 - P_i). \quad (5.4)$$

Ist der Betrag der Elastizität eines Faktors größer als 1, so ist die prozentuale Änderung der endogenen Variablen größer als die der exogenen, welche den Effekt auslöst. In einem solchen Fall spricht man von elastischen Variablen. (Urban, 1993)

### 5.4 Vergleich der berechneten Anteile

Soll nicht nur der Einfluss eines einzelnen Einflussfaktors auf die Änderung der endogenen Variable untersucht werden, sondern beispielsweise eine ganze Reihe an Faktoren, die sich durch eine umfangreiche Maßnahme ergeben, so ist es meist die einfachste Methode, das Modell mit den jeweils unterschiedlichen Ausprägungen in den Variablen mehrmals zu berechnen und die Ergebnisse zu vergleichen. Sind die Parameter des Modells einmal bekannt, so kann die Funktion dahingehend verändert werden, dass die exogenen Variablen das neue Verkehrsangebot darstellen und somit die neuen Auswahlwahrscheinlichkeiten berechnet werden.



## 6 Beispielhafte Anwendung der Logit-Analyse

Zur Veranschaulichung der Berechnungen, welche in den vorherigen Kapiteln erklärt sind, wird im Folgenden ein Beispiel dargestellt. Zur Vereinfachung wird hier ein binäres Logit-Modell mit nur zwei Auswahlalternativen betrachtet. Der Entscheidungsträger kann entweder die Option ÖV oder die Alternative MIV wählen.

### 6.1 Auswahl der Einflussfaktoren

Zunächst werden die Einflussfaktoren bestimmt, welche in den Befragungen erhoben werden sollen. Hierbei soll von jeder Kategorie an Variablen, welche in Kapitel 2 vorgestellt wurden, eine enthalten sein. Von Bedeutung sind vor allem die generischen Variablen Reisezeit und Kosten, welche direkt in Minuten bzw. € angegeben werden können.

Weiterhin sollen sozioökonomische Variable in die Nutzenfunktion aufgenommen werden. Zum einen eine alternativenspezifische und außerdem eine an eine generische Variable gekoppelte. Als spezifisch für die Alternative ÖV wird das Umweltbewusstsein gewählt. Dies wird aus den Angaben der befragten Personen nach der Einschätzung des persönlichen Umweltbewusstseins als Zahlenwert von 0 (=nicht umweltbewusst) bis 2 (=sehr umweltbewusst) angegeben. Die Variable Einkommen wird an die Reisezeit gekoppelt, da häufig zu beobachten ist, dass Personen mit einem höheren Einkommen die Zeit höher gewichten als Personen mit niedrigerem Einkommen. Hinzu kommt die alternativenspezifische Konstante für die Wahloption ÖV.



Einflussgröße	Bezeichnung	Einheit
Reisezeit	$T_k$	min
Kosten	$C_k$	€
Einkommen	$e_i$	Tsd. €/Monat
Umweltbewusstsein	$u_i$	einheitenlos, Angabe als Zahlenwert [0,1,2]

**Tabelle 2:** Einflussgrößen der Verkehrsmittelwahl im Beispiel

## 6.2 Nutzenfunktion

Die Nutzenfunktion soll in den Parametern eine lineare Gestalt haben. Durch die sozioökonomischen Variablen ist die Funktion somit quasilinear. Die einzelnen Einflussfaktoren werden additiv miteinander verknüpft. Der deterministische Teil lässt sich wie folgt darstellen;

$$V_{ki} = \beta_0 \delta_{k2} + \beta_1 T_{ki} + \beta_2 C_{ki} + \beta_4 e_i T_{ki} + \beta_5 u_i \delta_{k2}. \quad (6.1)$$

Der Parameter  $\beta_0$  wird nur bei der später zu schätzenden Alternative ÖV (=Alternative 2) benötigt, da diese in Referenz zum MIV berechnet wird. Die nächsten zwei Komponenten der Gleichung bewerten die Faktoren Zeit und Kosten mit den jeweiligen Parametern  $\beta$ . Das Umweltbewusstsein findet bei Alternative Zwei, dem ÖV Beachtung, der Parameter  $\beta_4$  entfällt also in der Nutzenfunktion für den MIV. Der letzte Summand beschreibt den (voraussichtlich negativen) Zusatznutzen der Reisezeit, welcher durch ein höheres Einkommen generiert wird.

Explizit ergibt sich daraus für die beiden Alternativen je eine Funktion. Da es lediglich auf Nutzendifferenzen ankommt wird der Alternative MIV der Wert Null zugewiesen, sie gilt als Referenz.

$$V_{MIV,i} = \beta_1 T_{MIV,i} + \beta_2 C_{MIV,i} + \beta_5 T_{MIV,i} e_i = 0 \quad (6.2)$$

$$V_{ÖV,i} = \beta_0 + \beta_1 T_{ÖV,i} + \beta_2 C_{ÖV,i} + \beta_4 u_i + \beta_5 T_{ÖV,i} e_i \quad (6.3)$$



### 6.3 Exemplarische Datensätze

Die benötigten Daten zur Ermittlung der Parameter können aus unterschiedlichen Quellen stammen. In umfangreichen Verkehrsanalysen werden die abhängigen Variablenwerte aus dem Zensus zum Mobilitätsverhalten wie MiD oder SrV bestimmt. Zur Bestimmung der unabhängigen Variablen, also der Einflussfaktoren, wird eine Abbildung des Straßen- und Schienennetzes des Ortes benötigt.

In einer fiktiven revealed preference Erhebung seien 32 Personen, die zu einer verhaltenshomogenen Gruppe gehören (z.B. täglich Einpendler in einen bestimmten Ort) befragt worden, welches Verkehrsmittel sie für die Fahrt genutzt haben<sup>7</sup>. Alle Einflussfaktoren, die als relevant identifiziert wurden und in die Nutzenfunktion aufgenommen werden, müssen hierbei erhoben werden. Einige beispielhafte Antworten sind in Tabelle 3 dargestellt.

Person	T <sub>1</sub> [min]	T <sub>2</sub> [min]	C <sub>1</sub> [€]	C <sub>2</sub> [€]	e <sub>i</sub> [Tsd. €/Monat]	u <sub>i</sub> [0=wenig, 1=mäßig, 2=sehr]	Wahl
1	40	50	15	15	4	2	ÖV
2	45	55	17	13	3	1	MIV
3	40	30	24	20	2	0	MIV
4	40	35	10	14	5	0	MIV
5	80	70	15	24	1	2	ÖV

**Tabelle 3:** fiktive Befragungsdaten (Auszug)

Für jede Person werden also die 4 Einflussfaktoren der Verkehrsmittelwahl dargestellt, bei Zeit und Kosten jeweils für beide zur Auswahl stehenden Alternativen. Die letzte Spalte gibt Auskunft darüber, für welche Alternative sich die Testperson entschieden hat. Person 1 entscheidet sich bei gleichen Kosten für die Alternative mit der geringeren Reisezeit. Dies ist der ÖV, auch das Umweltbe-

<sup>7</sup> Somit entfällt die sog. *no choice* Variante, in der eine befragte Person aussagen kann, sie wählt keines der zur Auswahl stehenden Verkehrsmittel, da die Daten a posteriori erhoben werden.



wusstsein ist bei dieser Person hoch, was auf eine Präferenz des ÖV hindeuten könnte. Testperson 2, welche ein geringes Einkommen hat, wählt trotz längerer Reisedauer die Alternative MIV, hier sind die Kosten deutlich geringer. In ähnlicher Weise können auch die Daten der anderen Personen interpretiert werden. Um genauere Aussagen über die Bewertung der einzelnen Faktoren der gesamten Gruppe treffen zu können, werden im folgenden Schritt die Parameter des Modells geschätzt.

### 6.4 Modellschätzung

Aus den erhobenen Daten können nun die Parameter für das Entscheidungsmodell geschätzt werden. Dazu wird die Maximum-Likelihood-Methode angewandt. Mit dem Näherungsverfahren von Newton wurden die Parameter in sieben Iterationsschritten angenähert. Die Rechnungsschritte finden sich in Form eines Matlab-Skriptes im Anhang. Die Schätzung der Parameter hatte die in Tabelle 4 dargestellten Ergebnisse.

$\beta_0$	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$
4,1273	-0,0175	-0,0987	-0,0418	4,5443

**Tabelle 4:** geschätzte Parameter der Nutzenfunktion

Die Parameter geben die Sensitivität bezüglich des zugehörigen Einflussfaktors an. Da bei der Wahl zwischen zwei Alternativen stets eine von beiden als Referenz gelten kann und der deterministische Nutzenanteil von der Alternative MIV auf den Wert Null gesetzt wird, sind hier alle dargestellten Parameter die der Alternative ÖV und werden im Verhältnis zur Wahl des MIV ausgedrückt. Der Parameter  $\beta_0$  weist auf Einflüsse auf die endogene Variable  $V$  hin, die von anderen als den im Modell enthaltenen exogenen Variablen ausgehen (Urban, 1993) oder drückt die globale Präferenz für den ÖV gegenüber dem MIV aus.

Wie zu erwarten war, tragen die Parameter  $\beta_1$  bis  $\beta_3$  negative Vorzeichen. Mit steigender Reisezeit, bzw. Kosten sinkt also der Nutzen. Ein zusätzliches Sinken des Nutzens erfolgt bei einer hohen Reisezeit, wenn der Entscheidungsträger ein höheres Einkommen hat. Der Parameter zur Gewichtung des Umweltbewusstseins trägt positiv zum Nutzen des ÖV bei. Bei steigendem Umweltbewusstsein



steigt also der Nutzen für den ÖV. Die Höhe der Parameter selbst ist zunächst nicht eindeutig interpretierbar.

Weiteres zur Interpretation der Werte folgt nachdem das Modell evaluiert wurde, da zunächst sichergestellt werden soll, dass das geschätzte Modell die Daten gut repräsentiert, bevor konkrete Aussagen im Zusammenhang mit den geschätzten Parametern getroffen werden.

## 6.5 Überprüfung

Die *Goodness-of-fit* wird mit Hilfe des *Likelihood-Ratio-Indexes* als absolutes Maß der Modellgüte berechnet. Dieser ergibt sich aus

$$\rho^2 = 1 - \frac{LL(\beta)}{LL(0)} = 1 - \frac{-8,4117}{-22,1807} = 0,6208. \quad (6.4)$$

Dieser Wert ohne Berücksichtigung der Anzahl an Koeffizienten muss zunächst korrigiert werden, sodass sich nach Anwendung der Gleichung 4.21 der korrigierte Likelihood-Ratio-Index ergibt. Dessen Wert von 0,3953 liegt im Bereich dessen, wo das Modell als sehr gute Repräsentation der Daten gewertet wird.

Mit dem Likelihood-Ratio-Test wird die Relevanz der exogenen Variablen und deren zugehöriger Parameter getestet. Hierbei wird wie üblich bei Hypothesentests, in den folgenden vier Schritten vorgegangen:

1. Aufstellen der Nullhypothese:  $H_0$ : „Die zusätzlichen Parameter des unrestringierten Modells tragen nichts zur Erklärung des Sachverhaltes bei.“ Somit lautet die Gegenhypothese  $H_1$ : „Das unrestringierte Modell beschreibt die Daten besser als das restringierte, die zusätzlichen Parameter sind relevant.“
2. Test-Statistik: Wie in Gleichung 4.15 in Kapitel 4 beschrieben, wobei hier als restringiertes Modell das betrachtet wird, bei dem alle Modellparameter gleich Null sind.
3. Realisierung: Die Test-Variable  $\lambda$  gehorcht einer Chi-Quadrat-Verteilung mit 5 Freiheitsgraden, jeder zusätzliche Parameter des unrestringierten Modells bedeutet ein zusätzlicher Freiheitsgrad.  $\lambda \sim \chi^2(5)$



4. Testergebnis: Mit einer gewählten Fehlerwahrscheinlichkeit von  $\alpha=5\%$  also einer Wahrscheinlichkeit von  $p=1-\alpha=0,95$  und  $5-0=5$  Freiheitsgraden, kann aus der Tabelle der  $\chi^2$ -Verteilung ein Wert von 11,7 abgelesen werden. Da im vorliegenden Fall gilt:  $\lambda_{data} = 27,5381 > \chi_{0,95,6}^2$ , kann  $H_0$  abgelehnt werden. Das unrestringierte Modell beschreibt die Daten signifikant besser als das restringierte Modell.

Als dritte Methode zu Überprüfung der Modellschätzung wird der Prognoseerfolg berechnet. Für jede Person kann durch einsetzen der Nutzfunktion mit individuellen Variablen und den geschätzten Parametern in die Logit-Formel, die Wahrscheinlichkeit berechnet werden, mit der die betrachtete Person das Verkehrsmittel ÖV wählt. Da es hier nur zwei Alternativen gibt, gilt die Alternative als gewählt, wenn die Wahrscheinlichkeit größer ist als 50%. Im andern Fall wird die Wahl der Alternative MIV vorhergesagt. Nun kann mit den realen Daten der Befragung verglichen werden. Der Prozentsatz der richtig prognostizierten Entscheidungen liegt im vorliegenden Fall bei 84,38%.

## 6.6 Interpretation der Parameter

Da das Modell mit den geschätzten Parametern als gute Repräsentation der Datensätze validiert wurde, können nun die Parameter genauer betrachtet werden. Wie vorab erwähnt, sind zunächst die Vorzeichen relevant. Ein negatives Vorzeichen bedeutet, dass bei Anstieg des jeweiligen Faktors der deterministische Nutzen sinkt. Die Vorzeichen der geschätzten Parameter scheinen plausibel.

Weiterhin soll die Effektstärke der exogenen Variablen untersucht werden, um herauszufinden, welche Einflussfaktoren größere und weniger große Bedeutung in der Verkehrsmittelwahl der betrachteten Gruppe haben. Um zu verdeutlichen, dass die Parameter stark von der Skalierung der zugehörigen Variable abhängen, wird das Modell erneut geschätzt. Diesmal wird die Reisezeiten in Stunden statt Minuten angegeben. Dies wirkt sich sofort auf die Parameter aus, welche mit dem Faktor Zeit multipliziert werden. So ergeben sich für die Parameter  $\beta_1$  und  $\beta_3$  die folgenden neuen Werte.





$\beta_1$ bei Zeit in h	$\beta_3$ bei Zeit in h
-1,0516	-2,5078

**Tabelle 5:** geschätzte Modellparameter bei Änderung der Einheiten der Reisezeit von Minuten auf Stunden

Vergleichbar werden die Parameter erst, wenn man diese durch die Standardabweichungen dividiert und somit die Werte der t-Statistik erhält, welche in beiden Varianten die gleichen Werte aufweisen.

t ( $\beta_0$ )	t ( $\beta_1$ )	t ( $\beta_2$ )	t ( $\beta_3$ )	t ( $\beta_4$ )
1,6495	-0,4539	-0,9176	-2,9016	3,1176

**Tabelle 6:** Werte der t-Statistik der geschätzten Parameter

Diese Werte sind besser interpretierbar und es kann festgestellt werden, dass der Einfluss des Faktors Kosten ca. doppelt so stark ist wie der des Faktors Zeit. Allerdings stimmt auch dies hier nicht, da auch der Parameter  $\beta_3$  mit der Reisezeit zusammenhängt. Die beiden Parameter müssen also zusammen betrachtet werden, da  $\beta_3$  den Anteil des sinkenden Nutzens durch die steigende Reisezeit ausdrückt, welcher nicht in  $\beta_1$  enthalten ist. Ebenso sind auch in den Parametern  $\beta_0$  und  $\beta_4$  andere Einflüsse enthalten, die den Nutzen der Alternative ÖV steigen lassen. Je mehr Einflussfaktoren in das Modell integriert werden, desto genauer werden die Aussagen über die Einflussstärken der Einzelnen.

Da in diesem Fall nicht nur der Parameter  $\beta_1$ , sondern auch  $\beta_3$  mit der Reisezeit multipliziert wird, errechnet sich in diesem Fall der Zeitwert der Personen aus den

Quotienten  $\frac{\beta_1}{\beta_2} + \frac{\beta_1}{\beta_3} * \text{Einkommen}$ . Deutlich erkennbar ist hier, dass in diesem

Modell berücksichtigt wird, dass Menschen mit einem höheren Einkommen auch einen höheren Zeitwert haben. Zu beachten ist jedoch, dass die Daten, mit denen das Modell geschätzt wurde, eine Umfrage unter Pendlern, die den Weg Wohnen – Arbeiten bewerten, nachbildet. Auf eine andere Quelle-Ziel-Gruppe, wie beispielsweise Wohnen – Einkaufen kann dieses Modell nicht ohne Weiteres übertragen werden. Die Zeitwerte sind also abhängig vom Einkommen der Personen. Wie in der dargestellten Tabelle erkennbar, steigt der Wert der Zeit mit steigendem Einkommen.



Einkommen [€/Monat]	1.000	2.000	3.000	4.000	5.000
Zeitwert [€/Minute]	0,5969	1,0163	1,4356	1,8549	2,2743
Zeitwert [€/Std.]	35,8144	60,9751	86,1358	111,2965	136,4573

**Tabelle 7:** Zeitwerte nach Einkommensklassen sortiert

Mit diesem Wissen können auch Effekte von Infrastrukturmaßnahmen quantifiziert werden. Eine Reisezeitverbesserung um eine Minute kann mit einer Preissteigerung von oben berechnetem Zeitwert verbunden sein, wobei sich die Auswahlwahrscheinlichkeit nicht ändern würde.

Des Weiteren werden die Elastizitäten der Parameter berechnet um Auswirkungen von Änderungen in den exogenen Variablen auf die endogene Variable zu untersuchen.

$E_0$	$E_1$	$E_2$	$E_3$	$E_4$
0,2180	-0,0463	-0,0417	-0,2208	0,2400

**Tabelle 8:** Elastizitäten der Parameter des Modells

Eine einprozentige Erhöhung der Kosten würde hier nur zu einer 0,04%igen Senkung der Auswahlwahrscheinlichkeiten  $P(\ddot{O}V)$  führen, wobei eine einprozentige Erhöhung des Umweltbewusstseins einer Person<sup>8</sup> schon zu einer Erhöhung der Wahrscheinlichkeit um 0,24% führen würde. Alle Elastizitäten sind jedoch kleiner als  $|1|$ , was bedeutet, dass die Veränderung der Wahrscheinlichkeit  $P(\ddot{O}V)$  stets kleiner ist als die prozentuale Änderung der X-Variable, die den Effekt auslöst.

---

<sup>8</sup> abgesehen davon, dass dies in der Realität wahrscheinlich gar nicht möglich, in jedem Fall nicht messbar ist



## 6.7 Berechnung des Modal Splits einer Relation

Im konkreten Fall soll nun eine spezielle Relation untersucht werden, auf der täglich viele Pendler gemessen werden. Die aus der obigen Schätzung bekannten Parameter sollen hier Anwendung finden und das Verhalten der Pendler beschreiben. Für die Verlagerung des Verkehrs sollen dann verschiedene Infrastrukturmaßnahmen auf ihre Wirksamkeit überprüft werden.

Die Eigenschaften der Wahlalternative ÖV lassen sich wie folgt darstellen: Die betrachtete Relation von Ort A nach Ort B verbindet eine Bahntrasse, welche von einem Regionalzug bedient wird, dieser benötigt eine Zeit von 40 Minuten. Die Fahrkarte für die einfache Strecke kostet 20 €.

Die sozioökonomischen Eigenschaften hängen nicht von der Alternative ab. In dieser aggregierten Betrachtung wird mit Mittelwerten gerechnet. Das durchschnittliche Einkommen der Pendler auf dieser Relation beträgt 4000€ pro Monat und das mittlere Umweltbewusstsein wird mit einem Index von 1 geschätzt.

Einflussgröße	ÖV
Reisezeit $T_k$	40 min
Kosten $C_k$	20 €
Mittleres Einkommen	4000€/Monat
Geschätztes mittleres Umweltbewusstsein	1

**Tabelle 9:** Eigenschaften der Wahlalternative Bahn einer beispielhaften Relation

Unter der Annahme, dass die untersuchte Personengruppe verhaltenshomogen mit denen der obigen Schätzung ist, können nun Nutzenfunktion und Parameter aus dieser übernommen werden. Die Nutzenfunktion der Alternative ÖV lautet dann:

$$V_{\text{ÖV}} = \beta_0 + \beta_1 * 40 \text{ min} + \beta_2 * 20\text{€} + \beta_3 * 4000\text{€} / \text{Monat} * 40 \text{ min} + \beta_4 * 1. \quad (6.5)$$



Nun können die errechneten Parameter in die Gleichung eingesetzt werden. Die Auswahlwahrscheinlichkeit für die Alternative ÖV ergibt sich aus

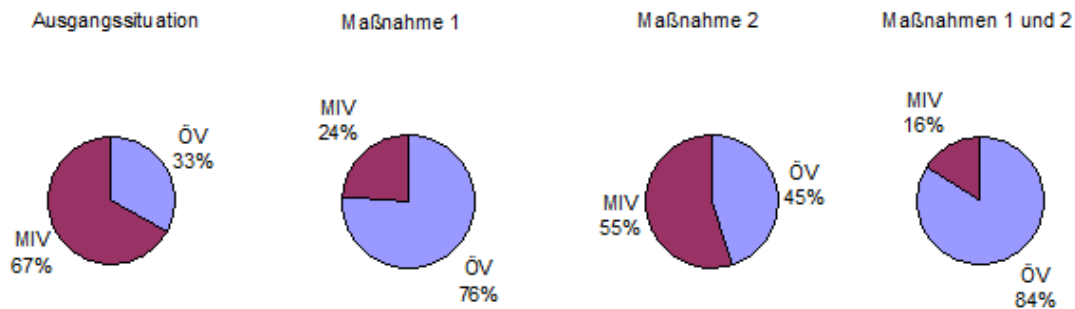
$$P_{\text{ÖV}} = \frac{1}{1 + e^{-V_{\text{ÖV}}}} = 33\% \text{ und somit } P_{\text{MIV}} = 1 - P_{\text{ÖV}} = 67\% .$$

Stehen also auf dieser Relation die beiden Alternativen ÖV und MIV zur Wahl und man betrachtet eine Gruppe, welche so wie die Testgruppe aus obigem Beispiel entscheidet, und vernachlässigt erneut, dass auch die Variante *no choice* besteht, unterstellt also, dass alle Entscheidungsträger eine der beiden Alternativen wählen müssen und nicht die Möglichkeit besteht den Weg schlicht gar nicht anzutreten, so kann man von einem Modal Split von ÖV=33% und MIV=67% ausgehen.

### **6.8 Änderungen im Modal Split durch Verkehrsmaßnahmen**

Im Rahmen einer Politikmaßnahme soll nun versucht werden, so viel des Pendlerverkehrs wie möglich auf den Verkehrsträger Schiene umzulegen. Dazu werden zwei Konzepte geprüft. Zum einen besteht die Möglichkeit, die Reisezeit der Bahn mit Hilfe eines Ausbaus der Strecke zu senken. Die Reisezeit würde sich dann von 40 auf 30 Minuten verringern. Die andere Möglichkeit betrifft die Kosten, welche von 20 auf 15€ sinken würden. Die Eigenschaften der Option MIV bleiben unverändert, im Mittel beträgt die Reisezeit hier ca. 48 Min und die Kosten belaufen sich im Mittel auf 13,78€. Im Modell können nun die exogenen Variablen verändert werden, die Parameter bleiben gleich.

Für die Maßnahme Eins ergibt sich eine Steigerung der Wahlwahrscheinlichkeit für den ÖV auf 76%. Bei der Maßnahme Kostenreduktion fällt die Änderung der Auswahlwahrscheinlichkeit geringer aus. Hier ergibt sich lediglich eine Steigerung von 12% auf 45%. Ebenso könnten beide Maßnahmen zugleich durchgeführt werden, womit sich beide Variable in der Nutzenfunktion ändern. Die somit errechnete Auswahlwahrscheinlichkeiten beträgt 84%.



**Abbildung 6:** Änderung des Modal Split durch Infrastrukturmaßnahmen im Beispiel

## 6.9 Diskussion der Ergebnisse

Die Ergebnisse des Beispiels sollen im Folgenden noch einmal kritisch betrachtet werden, um in einer realen Anwendung einige Fehler zu vermeiden. Es muss stets darauf hingewiesen werden, dass oben gerechnete Ergebnisse aus einem Datensatz stammen, der rein fiktiv ist und keine empirische Grundlage hat. Das Beispiel soll der besseren Verständlichkeit des in der Theorie dargestellten Verfahrens aus den Kapiteln 4 und 5 dienen. Da der Fokus dieser Arbeit nicht in der Erstellung eines geeigneten Datensatzes oder eines Beispiels lag, ist hier vieles vereinfacht worden.

Zunächst wird ein Modell mit nur zwei Wahlalternativen für die meisten Zwecke in der Realität nicht ausreichend sein. Einem Entscheidungsträger steht wenigstens neben den Verkehrsmittelalternativen, welche auf betrachteter Relation existieren noch die Möglichkeit offen, die Reise gar nicht anzutreten, die sog. *no choice* Variante, in der der Wähler sich für keines der Alternativen entscheidet.

Des Weiteren sollte versucht werden, stets so viele Einflussfaktoren wie möglich in die Nutzenfunktion aufzunehmen. Allerdings auch nicht zu viele, da jedes Mal geprüft werden sollte, ob ein Faktor tatsächlich relevant ist. Im betrachteten Fall wäre beispielsweise eine Aufteilung der komplexen Reisezeit in die einzelnen Elemente sinnvoll, da somit auch Maßnahmen wie die Verkürzung der Wartezeit oder Ähnliches geprüft werden kann.

Im Beispiel wurde davon ausgegangen, dass die generischen Variablen für beide Alternativen genau bestimmt werden können. In der Realität muss hinterfragt werden, ob beispielsweise Kosten, welche entstünden, wenn Person x einen PKW



nutzt, für die Personen, welchen kein PKW zur Verfügung steht, überhaupt ermittelt werden können. Insbesondere stellt sich die Frage, ob für diese Personen die Alternative überhaupt PKW zur Wahl steht.

Weiterhin kann die Befragungsart abgeändert werden, da auch die Nutzung der *stated preference* viele Vorteile aufweist. Im Beispiel wurde eine Änderung der Variablen vorgenommen und daraufhin die Auswahlwahrscheinlichkeiten berechnet. Möglich ist es jedoch auch, solch neue Varianten direkt abzufragen, auch wenn sie nicht existieren. Somit können eventuell auch bisher unbedachte Faktoren mit in die Betrachtung aufgenommen werden.

Der Datensatz besteht der Einfachheit halber lediglich aus 32 Einzeldaten. Dieser geringe Stichprobenumfang ist für die meisten Fälle nicht repräsentativ. Auch die hier berechneten Zeitwerte und Änderungen im Modal Split sollten deshalb mit Vorsicht betrachtet werden.



## 7 Zusammenfassung und Ausblick

Zusammenfassend ist zu sagen, dass in der Verkehrsplanung bereits einige brauchbare Modelle zur Abbildung des Entscheidungsverhaltens in der Verkehrsmittelwahl existieren. Im Rahmen dieser Arbeit wurden einige der Modelle diskreter Entscheidungen beschrieben und das Logit-Modell im Besonderen ausführlich betrachtet.

Um Änderungen im Modal Split errechnen zu können, ist die Anwendung eines solchen Modells entscheidend. Die Modelle können auf verschiedene Weisen überprüft werden und bringen meist sehr gute Ergebnisse. Die Schwierigkeit in der Analyse besteht eher in einem anderen Aspekt. In der Analyse von Infrastrukturmaßnahmen bzw. der Prädiktion von deren Auswirkungen auf den Modal Split ist vor allem zu betrachten, welche Eigenschaftenn des Verkehrsangebots verändert werden, also welche der exogenen Variablen, die in das Modell eingehen von Bedeutung sind. Diese Variablen können dann mit Hilfe von statistischen Methoden genauer in der Einflussstärke bemessen werden.

Ebenfalls ist es wichtig, stets alle oder zumindest ausreichend viele Einflussfaktoren in einem solchen Modell zu betrachten, sodass keine falschen Rückschlüsse gezogen werden. Nicht nur dem im späteren betrachteten Faktor sollte Aufmerksamkeit geschenkt werden, sondern auch alle anderen Einflüsse sollten möglichst genau nachgebildet werden.

Größtes Problem beim Aufstellen solcher Art Modelle liegt in der Vielzahl an benötigten Daten. Diese müssen empirisch erfasst werden, um das Modell realistisch schätzen zu können. Um auf eine Gesamtheit schließen zu können, ist ein ausreichend großer Stichprobenumfang bei der Befragung von Bedeutung.

Für die Anwendung im Rahmen des Forschungsprojektes des DLR empfiehlt sich die Anwendung eines diskreten Entscheidungsmodells, wie das binomiale Logit-Modell, je nach Anwendungsfall auch eine Abwandlung dessen. Die in der Modellschätzung errechneten Parameter, welche die Gewichtung der Einflussfaktoren darstellen, können dann als Bewertung der Einzelaspekte in ein multikriterielles Verfahren übernommen werden.



## Literaturverzeichnis

**Axhausen, K. W.:** *Befragungsmethoden für hypothetische Märkte*, in G. Steierwald, H.-D. Künne und W.Vogt (Hrsg.) *Stadtverkehrsplanung*, Springer, Heidelberg, 2003

**Backhaus, K.,** Erichson, B., Plinke, W., Weiber, R.: *Multivariate Analysemethoden – Eine anwendungsorientierte Einführung*, 8.Auflage, Springer Verlag, Berlin, u.a., 1996

**Ben-Akiva, M.,** Lerman, S.: *Discrete Choice Analysis: Theory and Application to Travel Demand*, The MIT Press, Cambridge, 1985

**Brög, W.:** *Verkehrsmittelwahl im Stimmungswandel* in Giese, E. (Hrsg.): *Verkehr ohne (W)Ende?*, dgvt-Verlag, Tübingen, 1997

**Bühler, G.:** *Verkehrsmittelwahl im Güterverkehr – eine Analyse ordnungs- und preispolitischer Maßnahmen*, Schriftenreihe des Zentrums für europäische Wirtschaftsforschung, Heidelberg, 2006

**Eckey, H.-F.,** Stock, W.: *Verkehrsökonomie*, Gabler Verlag, Wiesbaden, 2000

**Eckey, H.-F.,** Kosfeld, R., Dreger, C.: *Ökonometrie Grundlagen – Methoden – Beispiele*, Gabler Verlag, Wiesbaden, 2001

**Gaudry, M.,** Dagenais, M.: *The Dogit Model*, in *Transportation Research* Vol. 13, Pergamon Press, Oxford, 1979

**Giese, E.** (Hrsg.): *Verkehr ohne W(Ende)?: psychologische und sozialwissenschaftliche Beiträge*, dgvt-Verlag, Tübingen, 1997

**Gorr, H.:** *Die Logik der individuellen Verkehrsmittelwahl, Theorie und Realität des Entscheidungsverhaltens im Personenverkehr*, Focus Verlag, Gießen, 1997

**Hensher, D.,** Button, K.: *Handbook of Transport Modelling*, Elsevier, Amsterdam, 2000





**Infas GmbH:** *Mobilität in Deutschland 2008 – Methodenbericht*, [http://www.mobilitaet-in-deutschland.de/pdf/MiD2008\\_Methodenbericht\\_I.pdf](http://www.mobilitaet-in-deutschland.de/pdf/MiD2008_Methodenbericht_I.pdf), [aufgerufen am 17.07.2012], Bonn und Berlin, 2010

**Kern, G.:** *Verkehrsmodelle und ihre Anwendungen*, Skript zur Vorlesung Verkehrsplanung I, FH Nürnberg, Fachbereich Bauingenieurwesen, 2004

**Keuchel, S.:** *Wirkungsanalyse von Maßnahmen zur Beeinflussung des Verkehrsmittelwahlverhaltens, Eine empirische Untersuchung am Beispiel des Berufsverkehrs der Stadt Münster/Westfalen*, Vandenhoeck & Ruprecht, Göttingen, 1994

**Kirchhoff, P.:** *Städtische Verkehrsplanung*, Teubner Verlag, Stuttgart u.a., 2002

**Knapp, F.:** *Determinanten der Verkehrsmittelwahl*, Abhandlungen zur Nationalökonomie, Band 10, Duncker & Humblod, Berlin, 1998

**Krämer, T., Schwanhäuser, W.(Hrsg.), Wolf, P.(Hrsg.):** *Der Modal-Split im Personenverkehr*, Veröffentlichungen des verkehrswissenschaftlichen Institutes der RWTH Aachen, Aachen, 1992

**Lange, U.:** *Wirkung komplexer Maßnahmen auf die Verkehrsmittelwahl von Umlandpendlern*, Schriftenreihe des Lehrstuhls für Verkehrstechnik der Technischen Universität München, Heft 8, München, 2009

**Ließke, F.:** Projektseite von Mobilität in Städten – SrV, [http://tu-dresden.de/die\\_tu\\_dresden/fakultaeten/vkw/ivs/vip/srv](http://tu-dresden.de/die_tu_dresden/fakultaeten/vkw/ivs/vip/srv), 2011 [Zugriff am 06.06.2012]

**Lohse, D., Schneider, R.:** *Vergleichende Untersuchungen der aggregierten und disaggregierten Verkehrsplanungsmodelle*, Schriftenreihe des Instituts für Verkehrsplanung und Straßenverkehr Heft 3/1997, Dresden, 1997

**Louviere, J., Street, D.:** *Stated preference Methods*, in Handbook of Transport Modelling, Amsterdam, 2000

**Maier, G., Weiss, P.:** Modelle diskreter Entscheidungen, Springer-Verlag, Wien, 1990



**Ortúzar, J. de D.**, Willumsen, L. G.: *Modelling Transport*, John Wiley & Sons, Chichester, 1994

**PTV AG Deutschland:** *Stated Preferences*,

<http://logistik.ptv.de/consulting/verkehrsplanung/stated-preferences/>, [aufgerufen am 18.06.2012], Karlsruhe, 2012

**PTV AG Deutschland:** *PTV Vision*, <http://www.ptv-vision.com/de/produkte/vision-traffic-suite/ptv-visum/ueberblick/>, [aufgerufen am 14.07.2012], Karlsruhe, 2012a

**Sammer, G.:** *Untersuchung zur Verkehrsmittelwahl im Personennahverkehr*, Schriftenreihe des Bundesministeriums für Bauten und Technik, Straßenforschung, Heft 179, Wien, 1982

**Schnabel, W.**, Lohse, D.: *Grundlagen der Straßenverkehrstechnik und der Verkehrsplanung* Band 2: Verkehrsplanung, Beuth Verlag, Berlin, 2011

**Steierwald, G.**, Künne, H..D., Vogt, W. (Hrsg.): *Stadtverkehrsplanung*, Springer Verlag, Heidelberg, 2003

**Train, K.:** *Discrete Choice Methods with Simulation*, Cambridge University Press, Cambridge, 2003

**Treiber, M.:** *Vorlesungsmanuskript zur Verkehrsökonomie Methoden und Modelle*, Dresden, 2011

**Tutz, G.:** *Die Analyse kategorialer Daten – Anwendungsorientierte Einführung in Logit-Modellierung und kategoriale Regression*, Oldenbourg Verlag, München, Wien Oldenbourg, 2000

**Urban, D.:** *Logit-Analyse – Statistische Verfahren zur Analyse von Modellen mit qualitativen Response-Variablen*, Gustav Fischer Verlag, Stuttgart, 1993

**Vrtic, M.**, Axhausen, K. W.: *Überprüfung von Prognoseansätzen im Personenverkehr – Ergebnisse einer Vorher-/Nachher-Untersuchung*, in Internationales Verkehrswesen Nr. 9, Sept. 2003



**Wikibooks:** Mathematik: Statistik: Tabelle der Chi-Quadrat-Verteilung, [http://de.wikibooks.org/wiki/Mathematik:\\_Statistik:\\_Tabelle\\_der\\_Chi-Quadrat-Verteilung](http://de.wikibooks.org/wiki/Mathematik:_Statistik:_Tabelle_der_Chi-Quadrat-Verteilung), Stand 22.11.2011, [aufgerufen am 30.07.2012]

**Winkler, C.:** *Ein integriertes Verkehrsnachfrage- und Bewertungsmodell – Ansatz einer Synthese von Mikroökonomie und Verkehrsplanung*, Dissertation, Dresden, 2011



## Anhang

A.	TABELLE DER CHI-QUADRAT-VERTEILUNG .....	XI
B.	EXEMPLARISCHE BEFRAGUNGSDATEN .....	XII
C.	MATLAB-SKRIPT ZUR MODELLSCHÄTZUNG .....	XIII
D.	MATLAB-SKRIPT ZUR MODELLÜBERPRÜFUNG .....	XV
E.	MATLAB-SKRIPT ZUR PARAMETERINTERPRETATION .....	XVI
F.	MATLAB-SKRIPT ZU BEISPIELMAßNAHMEN .....	XVII
G.	ÜBERSICHT DER ERGEBNISSE DER BEISPIELRECHNUNG .....	XVIII
H.	PLOT DER FUNKTION DER AUSWAHLWAHRSCHEINLICHKEITEN IM BEISPIEL .....	XIX



## A. Tabelle der Chi-Quadrat-Verteilung

Wahrscheinlichkeiten p										
FG	0,01	0,025	0,05	0,1	0,5	0,9	0,95	0,975	0,99	0,995
1	0,00	0,001	0,004	0,016	0,45	2,71	3,84	5,02	6,63	7,88
2	0,02	0,051	0,103	0,211	1,39	4,61	5,99	7,38	9,21	10,60
3	0,115	0,216	0,352	0,584	2,37	6,25	7,81	9,35	11,34	12,84
4	0,297	0,484	0,711	1,06	3,36	7,78	9,49	11,14	13,28	14,86
5	0,554	0,831	1,15	1,61	4,35	9,24	11,07	12,83	15,09	16,75
6	0,872	1,24	1,64	2,20	5,35	10,64	12,59	14,45	16,81	18,55
7	1,24	1,69	2,17	2,83	6,35	12,02	14,07	16,01	18,48	20,28
8	1,65	2,18	2,73	3,49	7,34	13,36	15,51	17,53	20,09	21,95
9	2,09	2,70	3,33	4,17	8,34	14,68	16,92	19,02	21,67	23,59
10	2,56	3,25	3,94	4,87	9,34	15,99	18,31	20,48	23,21	25,19

**Tabelle 10:** Quantile der Chi-Quadrat-Verteilung nach ausgewählten Wahrscheinlichkeiten p und Freiheitsgraden FG (vgl. wikibooks, 2011 und Eckey, 2001)



## B. exemplarische Befragungsdaten

Zeit ÖV [min]	Zeit MIV [min]	Kosten ÖV [€]	Kosten MIV [€]	Einkommen [Tsd. €/Monat]	Umwelt- bewusstsein [0=wenig, 1=mäßig, 2=stark]	Wahl ÖV [1=ja, 2=nein]	Wahl MIV [1=ja, 2=nein]
40	50	15	15	4	2	1	0
45	55	17	13	3	1	0	1
40	30	24	20	2	0	0	1
40	35	10	14	5	0	0	1
80	70	15	24	1	2	1	0
25	40	16	10	4	1	1	0
50	45	12	10	3	1	0	1
40	30	14	12	4	0	0	1
60	50	13	10	3	0	0	1
45	50	20	12	2	0	0	1
60	45	12	9	4	0	0	1
70	55	15	18	4	1	0	1
40	60	13	12	5	1	1	0
50	55	10	8	2	2	1	0
50	35	8	12	4	1	0	1
45	35	12	12	2	0	0	1
50	40	8	14	1	0	1	0
70	50	15	15	3	1	0	1
25	35	17	15	3	1	1	0
45	40	10	9	4	1	0	1
50	40	8	7	3	0	0	1
70	55	16	10	2	0	0	1
47	45	10	15	1	2	1	0
50	60	12	17	3	1	1	0
75	80	20	22	2	1	1	0
65	60	17	13	3	0	0	1
50	40	12	15	4	0	0	1
45	43	10	17	3	1	1	0
70	55	14	9	2	1	0	1
40	42	10	17	1	1	1	0
60	45	20	22	4	0	0	1
50	65	8	13	2	1	1	0

**Tabelle 11:** Tabelle der fiktiven Befragungsdaten



## C. Matlab-Skript zur Modellschätzung

```

%Daten
Zeit_PT=[40 45 40 40 80 25 50 40 60 45 60 70 40 50 50 45 50 70 25 45 50
70 47 50 75 65 50 45 70 40 60 50];
Zeit_MIV=[50 55 30 35 70 40 45 30 50 50 45 55 60 55 35 35 40 50 35 40 40
55 45 60 80 60 40 43 55 42 45 65];
Kosten_PT=[15 17 24 10 15 16 12 14 13 20 12 15 13 10 8 12 8 15 17 10 8 16
10 12 20 17 12 10 14 10 20 8];
Kosten_MIV=[15 13 20 14 24 10 10 12 10 12 9 18 12 8 12 12 14 15 15 9 7 10
15 17 22 13 15 17 9 17 22 13];
Einkommen=[4 3 2 5 1 4 3 4 3 2 4 4 5 2 4 2 1 3 3 4 3 2 1 3 2 3 4 3 2 1 4
2];
EZ_PT=Einkommen.*Zeit_PT;
EZ_MIV=Einkommen.*Zeit_MIV;
Umwelt=[2 1 0 0 2 1 1 0 0 0 0 1 1 2 1 0 0 1 1 1 0 0 2 1 1 0 0 1 1 1 0 1];
Wahl_PT=[1 0 0 0 1 1 0 0 0 0 0 0 1 1 0 0 1 0 1 0 0 0 1 1 1 0 0 1 0 1 0
1];
Wahl_MIV=[0 1 1 1 0 0 1 1 1 1 1 1 0 0 1 1 0 1 0 1 1 1 0 0 0 1 1 0 1 0 1
0];

%realisierte Merkmalssummen
ZEIT_PT_DATA = sum(Zeit_PT.*Wahl_PT);
ZEIT_MIV_DATA = sum(Zeit_MIV.*Wahl_MIV);
KOSTEN_PT_DATA = sum(Kosten_PT.*Wahl_PT);
KOSTEN_MIV_DATA = sum(Kosten_MIV.*Wahl_MIV);
EINK_PT_DATA = sum(EZ_PT.*Wahl_PT);
EINK_MIV_DATA = sum(EZ_MIV.*Wahl_MIV);
UMWELT_PT_DATA = sum(Umwelt.*Wahl_PT);
UMWELT_MIV_DATA = sum(Umwelt.*Wahl_MIV);
WAHL_PT_DATA = sum(Wahl_PT);
WAHL_MIV_DATA = sum(Wahl_MIV);
Personen = Wahl_PT+Wahl_MIV;

%Startbedingungen
b=zeros(5,1);
F=[1000; 1000; 1000; 1000; 1000];
i=0;

%Newton-Iterationen
while F(1)>1.0e-4 %Abbruchkriterium
    i=i+1;

%Nutzenfunktion für Alternative 1
V1=b(1)+b(2).*Zeit_PT+b(3).*Kosten_PT+b(4).*EZ_PT+b(5).*Umwelt;

%Wahlwahrscheinlichkeiten
P1=1./(1+exp(-V1));
P2=1-P1;

%Schätzbedingungen
F(1)=sum(Personen.*P1)-WAHL_PT_DATA;
F(2)=sum(Personen.*P1.*Zeit_PT)-ZEIT_PT_DATA;
F(3)=sum(Personen.*P1.*Kosten_PT)-KOSTEN_PT_DATA;
F(4)=sum(Personen.*P1.*EZ_PT)-EINK_PT_DATA;
F(5)=sum(Personen.*P1.*Umwelt)-UMWELT_PT_DATA;

%Elemente der Jacobi-Matrix
J=zeros(5,5);

```



```
J(1,1)=sum(Personen.*P1.*P2);
J(1,2)=sum(Personen.*Zeit_PT.*P1.*P2);
J(1,3)=sum(Personen.*Kosten_PT.*P1.*P2);
J(1,4)=sum(Personen.*EZ_PT.*P1.*P2);
J(1,5)=sum(Personen.*Umwelt.*P1.*P2);
J(2,1)=J(1,2);
J(2,2)=sum(Zeit_PT.*Zeit_PT.*P1.*P2);
J(2,3)=sum(Zeit_PT.*Kosten_PT.*P1.*P2);
J(2,4)=sum(Zeit_PT.*EZ_PT.*P1.*P2);
J(2,5)=sum(Zeit_PT.*Umwelt.*P1.*P2);
J(3,1)=J(1,3);
J(3,2)=J(2,3);
J(3,3)=sum(Kosten_PT.*Kosten_PT.*P1.*P2);
J(3,4)=sum(Kosten_PT.*EZ_PT.*P1.*P2);
J(3,5)=sum(Kosten_PT.*Umwelt.*P1.*P2);
J(4,1)=J(1,4);
J(4,2)=J(2,4);
J(4,3)=J(3,4);
J(4,4)=sum(EZ_PT.*EZ_PT.*P1.*P2);
J(4,5)=sum(EZ_PT.*Umwelt.*P1.*P2);
J(5,1)=J(1,5);
J(5,2)=J(2,5);
J(5,3)=J(3,5);
J(5,4)=J(4,5);
J(5,5)=sum(Umwelt.*Umwelt.*P1.*P2);

%Parameter für nächste Iteration
b=b-inv(J)*F;
end

%Ausgabe des Ergebnisses
disp('Die Parameter lauten:'),disp(b)
disp('Anzahl an Iterationen:'),disp(i)
disp('Schätzbedingungen F:'), disp(F)
```





## D. Matlab-Skript zur Modellüberprüfung

```

%Log-Likelihood mit geschätzten betas
for i=1:length(P1)
    a(i)=Wahl_PT(i)*log(P1(i))+Wahl_MIV(i)*log(P2(i));
    l_beta=sum(a);
end
disp('Log-Likelihood mit geschätzten betas'), disp(l_beta)

%Log-Likelihood bei betas=0
l_0=sum(Wahl_PT.*log(0.5)+Wahl_MIV.*log(0.5));
disp('Log-Likelihood bei betas=0'), disp(l_0)

%Goodness-of-Fit
likelihood_ratio_index=1-(l_beta/l_0)

%korrigierter Likelihood-Ratio-Index
likelihood_ratio_index_korr=1-((l_beta-5)/l_0)

%Likelihood-Ratio-Test
disp('H0= Die zusätzlichen Parameter sind irrelevant. ')
lambda=2*(abs(l_0)-abs(l_beta))
Freiheitsgrade=5-0
disp('Fehlerwahrscheinlichkeit = 5%')
Chiquadrat=11.7
if lambda>=Chiquadrat;
    disp('H0 abgelehnt, das unrestringierte Modell beschreibt die Daten
signifikant besser. '),
else
    disp('H0 angenommen, das restringierte Modell beschreibt die Daten
mindestens genauso gut. ')
end

%Bewertung des Prognoseerfolgs
for i=1:length(P1)
    if P1(i)>0.5
        Vorhersage(i)=1;
    else
        Vorhersage(i)=0;
    end
end
for i=1:length(Vorhersage)
    if Wahl_PT(i)==Vorhersage(i)
        Vorh_richtig(i)=1;
    else
        Vorh_richtig(i)=0;
    end
end
Prognoseerfolg=sum(Vorh_richtig)/sum(Personen);
disp('Anteil der richtig prognostizierten Entscheidungen'),
disp(Prognoseerfolg)

```



## E. Matlab-Skript zur Parameterinterpretation

```

%Varianz-Kovarianz-Matrix der Schätzer
V=inv(J)/2;

%Standardabweichungen
Stdabw(1)=sqrt(V(1,1));
Stdabw(2)=sqrt(V(2,2));
Stdabw(3)=sqrt(V(3,3));
Stdabw(4)=sqrt(V(4,4));
Stdabw(5)=sqrt(V(5,5));
disp('Standardabweichungen der Schätzer:'), disp(Stdabw)

%t-Statistik
for i=1:5
tStatistik(i)=b(i)/Stdabw(i);
end
disp('t-Statistik'), disp(tStatistik)

%Quotienten
Quotient(1,2)=b(1)/b(2);
Quotient(1,3)=b(1)/b(3);
Quotient(1,4)=b(1)/b(4);
Quotient(1,5)=b(1)/b(5);
Quotient(2,3)=b(2)/b(3);
Quotient(2,4)=b(2)/b(4);
Quotient(2,5)=b(2)/b(5);
Quotient(3,4)=b(3)/b(4);
Quotient(3,5)=b(3)/b(5);
Quotient(4,5)=b(4)/b(5);
disp('Quotienten'), disp(Quotient)

%Elastizitäten
for i=1:length(P1)
E(1)=mean(b(1)*Wahl_PT(i)*P1(i)*(P2(i)));
E(2)=mean(b(2)*Zeit_PT(i)*P1(i)*(P2(i)));
E(3)=mean(b(3)*Kosten_PT(i)*P1(i)*(P2(i)));
E(4)=mean(b(4)*EZ_PT(i)*P1(i)*(P2(i)));
E(5)=mean(b(5)*Umwelt(i)*P1(i)*(P2(i)));
end
disp('Elastizitäten:'), disp(E)

%Zeitwerte
Zeitwert_1=b(2)/b(3)+(b(2)/b(4))*1
Zeitwert_2=b(2)/b(3)+(b(2)/b(4))*2
Zeitwert_3=b(2)/b(3)+(b(2)/b(4))*3
Zeitwert_4=b(2)/b(3)+(b(2)/b(4))*4
Zeitwert_5=b(2)/b(3)+(b(2)/b(4))*5

```



## F. Matlab-Skript zu Beispielmaßnahmen

*%Beispiel Ausgangssituation*

```
V_Bsp=b(1)+b(2)*40+b(3)*20+b(4)*4*40+b(5)*1  
P_Bsp=1/(1+exp(-V_Bsp))
```

*%Maßnahme 1*

```
V_M1=b(1)+b(2)*30+b(3)*20+b(4)*4*30+b(5)*1  
P_M1=1/(1+exp(-V_M1))
```

*%Maßnahme 2*

```
V_M2=b(1)+b(2)*40+b(3)*15+b(4)*4*40+b(5)*1  
P_M2=1/(1+exp(-V_M2))
```

*%beide Maßnahmen*

```
V_M3=b(1)+b(2)*30+b(3)*15+b(4)*4*30+b(5)*1  
P_M3=1/(1+exp(-V_M3))
```

## G. Übersicht der Ergebnisse der Beispielrechnung

$\beta_0$	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$
4,1273	-0,0175	-0,0987	-0,0418	4,5443
$\beta_1$ (Zeit in h)		$\beta_3$ (Zeit in h)		
-1,0516		-2,5078		
F(1)	F(2)	F(3)	F(4)	F(5)
0,0000	0,0009	0,0001	0,0029	0,0000
LL( $\beta$ )		LL(0)		
-8,4117		-22,1807		
$\rho^2$		$\rho^2$ korrigiert		
0,6208		0,3953		
$\sigma(\beta_0)$	$\sigma(\beta_1)$	$\sigma(\beta_2)$	$\sigma(\beta_3)$	$\sigma(\beta_4)$
2,5021	0,0386	0,1076	0,0144	1,4576
t( $\beta_0$ )	t( $\beta_1$ )	t( $\beta_2$ )	t( $\beta_3$ )	t( $\beta_4$ )
1,6495	-0,4539	-0,9176	-2,9016	3,1176
E( $\beta_0$ )	E( $\beta_1$ )	E( $\beta_2$ )	E( $\beta_3$ )	E( $\beta_4$ )
0,2180	-0,0463	-0,0417	-0,2208	0,2400
Zeitwert (e=1) [€/Min]	Zeitwert (e=2) [€/Min]	Zeitwert (e=3) [€/Min]	Zeitwert (e=4) [€/Min]	Zeitwert (e=5) [€/Min]
0,5969	1,0163	1,4356	1,8549	2,2743

**Tabelle 12:** Ergebnisse des Beispiels

## H. Plot der Funktion der Auswahlwahrscheinlichkeiten im Beispiel

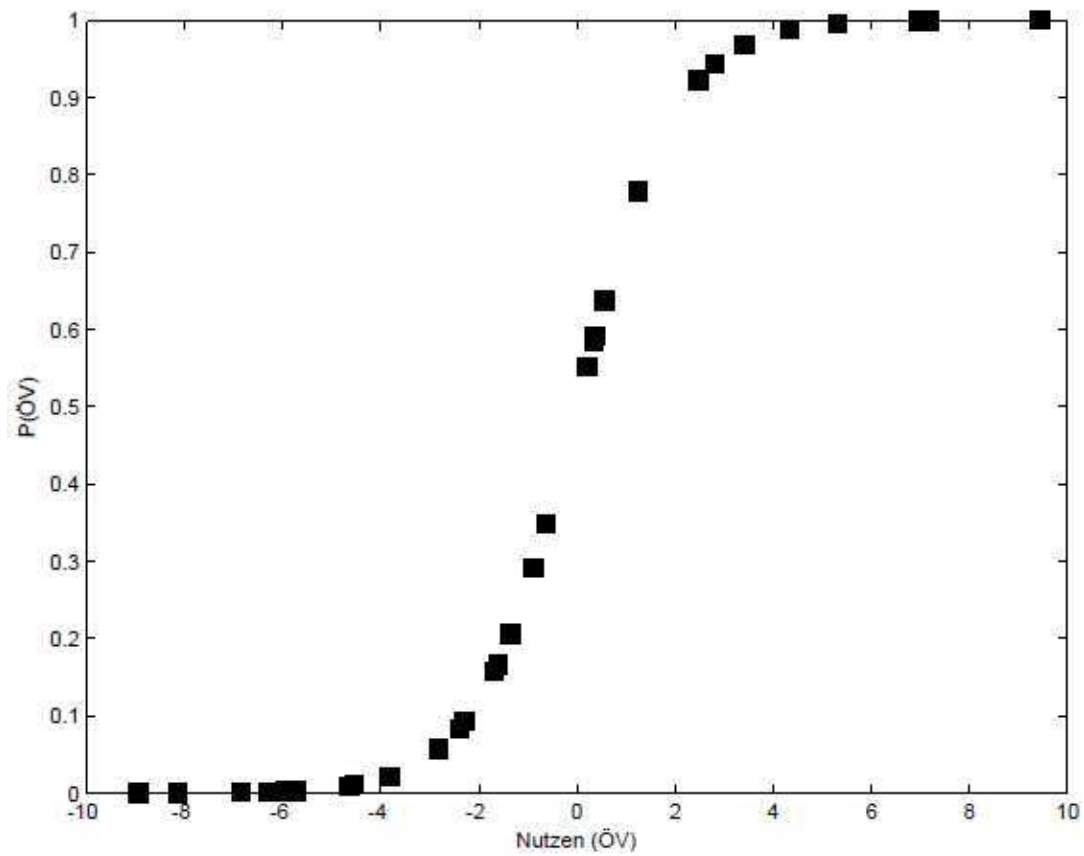


Abbildung 7: Auswahlwahrscheinlichkeiten des Beispiels



---

## **Ehrenwörtliche Erklärung**

Hiermit erkläre ich ehrenwörtlich, dass ich die vorliegende Arbeit selbst angefertigt habe. Die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sind als solche kenntlich gemacht. Die Arbeit wurde bisher noch keiner anderen Prüfbehörde vorgelegt und noch nicht veröffentlicht. Ich bin mir bewusst, dass eine unwahre Erklärung rechtliche Folgen haben wird.

Dresden, den 31.08.2012

---

Sabine Krause