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

Skript zur Vorlesung

Verkehrsökonometrie

für Bachelor-Studierende

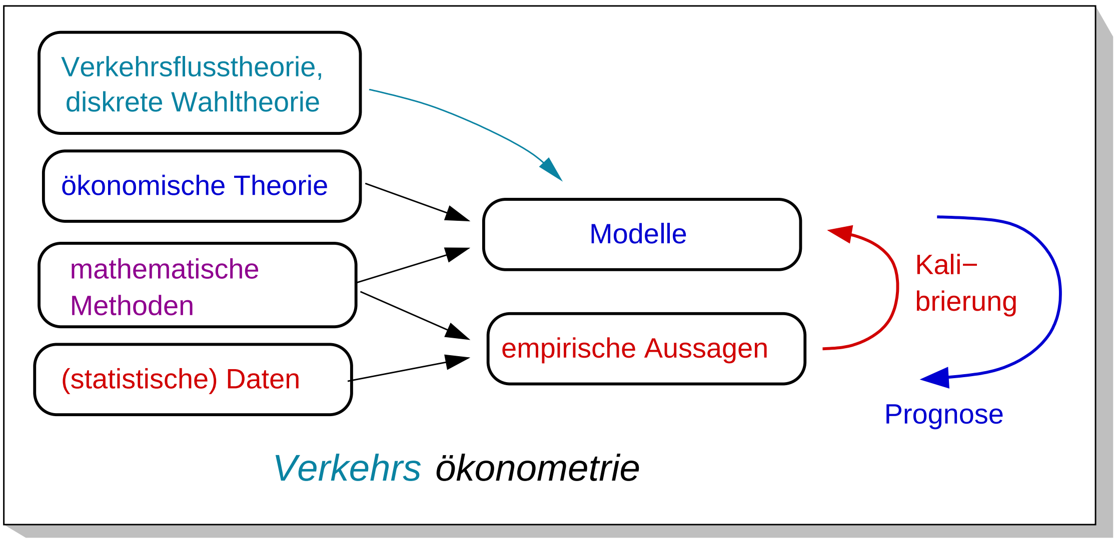
Dr. Martin Treiber

Sommersemester 2015

Inhaltsverzeichnis

1 Der Begriff der Verkehrsökonometrie

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



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

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

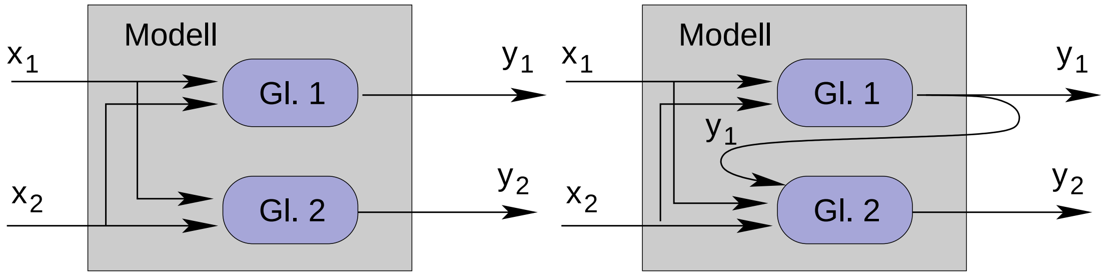


Abbildung 1.1: Flussdiagramm der exogenen und endogenen Variablen des allgemeinen ökonometrischen Modells (1.1). Gezeigt ist der Fall für je zwei exogene und endogene Variablen, $M = k = 2$ ohne (links) und mit (rechts) Kopplung der endogenen Variablen. Das Modell (eckige Box) besteht hier also aus zwei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

1.1 Modelle und Variablen der Verkehrsökonometrie

Da die Ökonometrie wirtschaftliche Zusammenhänge quantitativ beschreibt, besteht ihre Grundlage aus mathematischen Gleichungen. Meist werden Gleichungen folgender Struktur betrachtet (vgl. Abb. 1.1):

$$\boxed{Y_k = f_k(x_1, ..., x_m, .., x_M, \beta_0, ..., \beta_j, ..., \beta_J) + \epsilon_k.} \tag{1.1}$$

1.1.1 Endogene Variablen

Die Y_k sind die **erklärten Variablen**. Eine solche Variable wird bisweilen auch als **Explanandum** (lat. "zu erklärende" bzw. erklärte Variable), als **endogene Variable** (Griechisch: von innen heraus, also aus dem Modell kommend) oder gemäß allgemeinen mathematischen Sprachgebrauch als **abhängige Variable** bezeichnet. In der Systemtheorie entsprechen die Y_k dem *Output* des jeweiligen Modells. Da viele Modelle stochastischer Natur sind, also Zufallsgrößen ϵ_k enthalten (vgl. Abschnitt 1.1.4), sind die Y_k im Allgemeinen Zufallsvariablen und werden, der allgemeinen Konvention der Statistik entsprechend, "groß" geschrieben.

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

- **Eingleichungsmodelle**: Nur eine endogene Variable $Y_1 = Y$,
- **Mehrgleichungsmodelle**: Mehrere endogene Variable Y_k, die miteinander gekoppelt sind (vgl. Abb. 1.1 rechts),
- **Mehreren Eingleichungsmodelle**: Mehrere ungekoppelte endogene Variable wie im Flussdiagramm 1.1 links. Dies entspricht mehreren separat zu lösenden Eingleichungsmodellen.

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

1.1.2 Exogene Variablen

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

Je anch Anzahl der exogenen Variablen definiert man

- **univariate Modelle** bzw. **Einfachregression**: $M = 1$,
- **multivariate Modelle** bzw. **Mehrfachregression**[2]: $M > 1$.

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

1.1.3 Modellparameter

Die Größen β_j, $j = 0, \ldots, J$ sind die **Modellparameter**. Im Gegensatz zu den exogenen und endogenen Variablen, welche sich bei jedem System bzw. bei jeder Anwendung des Modells ändern, sollten die Modellparameter nach ihrer **Schätzung** bzw. **Modellkalibrierung** und einer solchermaßen bewirkten Anpassung des Modellverhaltens an die Wirklichkeit (vgl. Abschnitt 10.3) für alle Anwendungen einen konstanten Wert besitzen. *Diese Eigenschaft ist ein entscheidendes Kriterium für die Qualität eines Modells und seiner Aussage- und Prognosekraft!*

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

---

[1] In der Praxis werden sowohl die x_m als auch die Y_k aus Erhebungen gewonnen, beide Variablenkategorien sind damit prinzipiell stochastischer Natur. Man kann aber ohne Einschränkung der Allgemeinheit die Stochastizität der Input-Größen auf die Zufallsanteile ϵ_k verschieben, da die Modelle ja stochastische *Funktionen* darstellen, also die exogenen Variablen sowieso variabel sind. Dies ist sogar geboten, da andernfalls die Stochastizität überbestimmt ist. Näheres in Abschnitt 1.1.4.

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

1.1.4 Zufallsanteile

Die ϵ_k beschreiben unbestimmte oder **nicht erklärte Anteile** welche meist durch Zufallsvariablen mit Erwartungswert $E(\epsilon_k) = 0$ modelliert werden[3] Insbesondere gibt es folgende Gründe für die Notwendigkeit eines Zufallsanteils ϵ_k:

- Das Modell kann nicht alle Einflussfaktoren berücksichtigen. Wichtig ist aber, zumindest alle Faktoren mit *systematischen Einfluss* zu berücksichtigen, da das Modell sonst **fehlspezifiziert** ist und falsche Aussagen liefert (Näheres in der Master-Vorlesung).

- Die zur Modellkalibrierung verwendeten Messwerte sind fehlerbehaftet. Dieser Anteil von ϵ gibt direkt die kumulierten Messfehler wider.

- Der Mensch ist keine Maschine. Der entsprechende Anteil von ϵ spiegelt die Abweichung des in Wirklichkeit oft nichtrationalem Verhaltens vom Idealbild des *Homo Oeconomicus* wider.

Korrekt spezifizierte und eindeutig definierte Zufallsterme spielen eine wesentliche Rolle bei der Parameterschätzung (Abschnitt 10.3)

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

1.1.5 Modellfunktionen

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

Beispiel Verkehrsmittelwahl bei N Entscheidungen insgesamt:

$$Y_k = f_k(\boldsymbol{x}, \boldsymbol{\beta}) = N \frac{e^{V_k(\boldsymbol{x}, \boldsymbol{\beta})}}{\sum_{k'} e^{V_{k'}(\boldsymbol{x}, \boldsymbol{\beta})}} \tag{1.2}$$

mit den Nutzenfunktionen $V_k(\boldsymbol{x}, \boldsymbol{\beta})$, welche die Einflussfaktoren (exogenen Variablen) $\boldsymbol{x}$ (Reisezeit, Kosten,...) und die Modellparameter $\boldsymbol{\beta} = \{\beta_j\}$ (relative gewichtungen und bevorzugungen) enthalten. Näheres dazu im Abschnitt 1.3.2.

1.1.6 Das Prinzip der Sparsamkeit

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

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

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

1.2 Modellgleichungen

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

1.2.1 Inhaltliche Struktur

Hier unterscheidet man nach zwei Kategorien:

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

  $$Y(x) = \beta_0 + \beta_1 x + \epsilon \tag{1.3}$$

  und das geschätzte Modell

  $$\hat{Y}(x) = \hat{\beta}_0 + \hat{\beta}_1 x \tag{1.4}$$

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

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

  $$Y(x) = \beta_0 + \sum_{m=1}^{M} \beta_m x_m + \epsilon \tag{1.5}$$

- Die **Systemgleichungen**, manchmal auch als **Messgleichungen** bezeichnet, beschreiben die *Anwendung* des ökonometrischen Modells auf konkrete Systeme bzw. auf die von diesen Systemen gemessenen Werte der exogenen und endogenen Variablen. Im Gegensatz zum abstrakten Modell hängen die Systemgleichungen vom konkreten System ab. Die zum ökonometrische Modell (1.3) gehörigen Systemgleichungen lauten

  $$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad i = 1, \ldots, n \tag{1.6}$$

und werden auf n Datensätze $\{(x_i, y_i)\}$ angewandt. Die Systemgleichungen zum Mehrfachregressionsmodell (1.5) lauten

$$Y_i = \beta_0 + \sum_{m=1}^{M} \beta_m x_{mi} + \epsilon_i \tag{1.7}$$

und werden auf n Datensätze $\{(x_{1i}, x_{2i}, ..., x_{Mi}, y_i)\}$ angewandt. Mit Hilfe der Systemgleichungen kann man die Modellparameter derart schätzen, dass eine Funktion $F(\{\epsilon_i\})$ der Modellierungsfehler ϵ_i, z.B. die Fehlerquadratsumme, minimal wird (Kapitel 10.3)

Merke: Wichtig ist es, bei einfach indizierten exogenen Variablen zu unterscheiden, ob es sich um ein abstraktes Mehrfachregressionsmodell der Art (1.5) oder um Systemgleichungen eines Einfachregressionsmodells der Art (1.6) handelt.

Verständnisfrage:

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

1.2.2 Mathematische Struktur

Unterscheidung nach Linearität

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

- **Lineare Modelle**: Hier hängen die endogenen Variablen linear von den exogenen Variablen $\boldsymbol{x}$ und den Parametern $\boldsymbol{\beta}$ ab: Bei Eingleichungsmodellen haben sie also die Form

  $$Y(x) = \beta_0 + \sum_{m=1}^{M} \beta_m x_m + \epsilon. \tag{1.8}$$

  In vielen ökonometrischen Lehrbüchern werden ausschließlich lineare Modelle behandelt, in der Verkehrsökonometrie sind jedoch auch nichtlinieare Modelle wichtig.

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

  $$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 \tag{1.9}$$

  mit den erklärenden Variablen

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

Als Aussagen lassen sich hier zum Beispiel Elastizitäten gewinnen wie

$$\epsilon_2 = \frac{x_2}{Y}\frac{\mathrm{d}Y}{\mathrm{d}x_2} = \frac{\beta_2 x_2}{Y}$$

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

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

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

  $$y(t) = y_0 e^{t/\tau+\epsilon} \tag{1.10}$$

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

  $$w(t) = \ln(y_0) + t/\tau + \epsilon = \beta_0 + \beta_1 t + \epsilon \tag{1.11}$$

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

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

  Beispiel: Modell für das *Sättigungsverhalten* eines ökonomischen Prozesses (z.B. die Verkaufszahlen von Autos oder Mobiltelefonen seit Erfindung der jeweiligen Produkte) beschreiben will. Das klassische **Modell beschränkten Wachstums** mit der Zeit x als exogenen Variablen hat die Form (vgl. Abb. 1.2)

  $$\hat{y}(x) = \frac{y_s}{1 + \left(\frac{y_s}{y_0} - 1\right) e^{-x/\tau}} \tag{1.12}$$

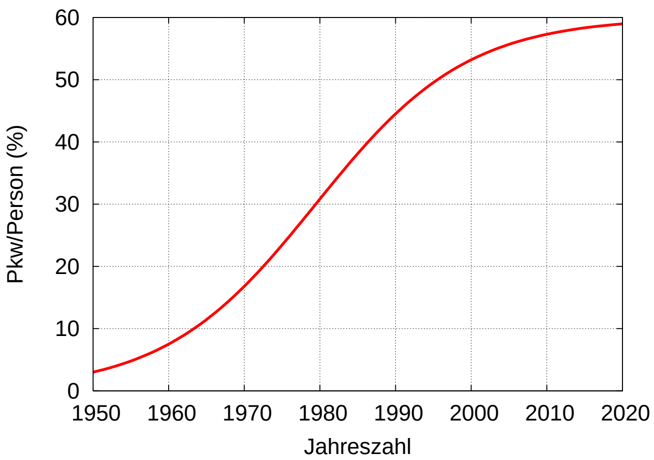


Abbildung 1.2: Graph des Modells (1.12).

Verständnisfrage:
Diskutieren Sie die Bedeutung der Parameter τ, y_0 und y_s im Modell (1.12). Kann man das Modell auch in der Form $\hat{y}(x) = y_s/[1+e^{-(x-x_0)/\tau}]$ schreiben? Was ist dann die Bedeutung des neuen Parameters x_0 und wie hängt er mit den Parametern der Formulierung (1.12) zusammen?

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

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

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

$$Y(\boldsymbol{x}, \boldsymbol{\beta}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \epsilon \tag{1.13}$$

mit den exogenen Variablen

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

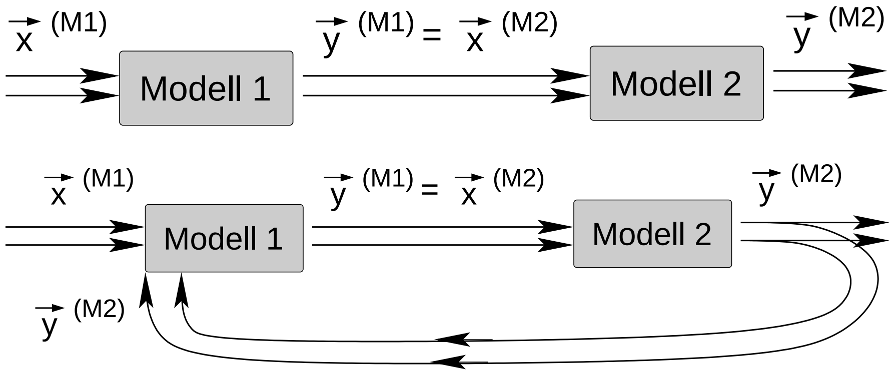


Abbildung 1.3: Flussdiagramm der exogenen und endogenen Variablen für verkette (oben) und rückgekoppelte (unten) Modell-Systeme. Das "Innenleben" der Modelle (also die je zwei Gleichungen, vgl. Abb. 1.1) ist nicht mehr gezeigt. Die Vektorpfeile symbolisieren den ganzen Satz jeweiliger endogener und exogener Variablen.

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

Mit den Transformationen

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

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

$$Y(\boldsymbol{z}, \boldsymbol{\beta}) = \beta_0 + \beta_1 z_1 + \beta_2 z_2 + \beta_3 z_3 + \epsilon \tag{1.14}$$

Die zugehörigen Anstiegsparameter bedeuten

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

Verständnisfrage:
Machen Sie sich die Vorzeichen der drei Parameter in diesem Beispiel klar. Warum ist bei Berücksichtigung auch extrem hoher Einkommen ein irreduzibel nichtlineares Modell notwendig? Zeigen Sie dies anhand der dann nicht plausiblen Aussagen des Modells (1.14).

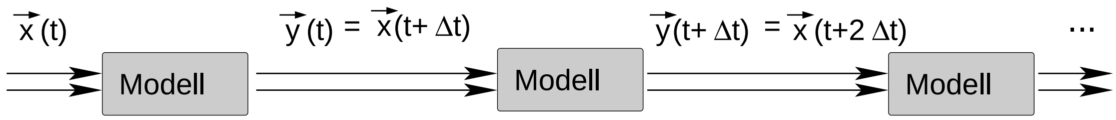


Abbildung 1.4: Flussdiagramm der Verkettung im Falle dynamischer Modelle. Das Modell selbst ist in der Regel in allen Schritten dasselbe, oft mit unveränderten Parametern (*autonomes* dynamisches Modell), manchmal mit von der Zeit abhängigen (*nichtautonomes* Modell).

Unterscheidung nach weiteren mathematischen Kriterien

Weitere in der Ökonometrie relevante Unterscheidungsmerkmale sind

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

  Hingegen ist die Skalierung der exogenen Variablen nicht so bedeutsam, da man qualitative bzw. nominalskalierte exogene Variable durch kardinalskalierte **Pseudovariablen** der Art 1=Maurer, 2=Schreiner, 3=Physiker etc ausdrücken kann.
- Existenz von Verkettung oder Rückkopplung (siehe Abb. 1.3 und Abb. 1.5):
  - Bei **verketteten Modellen** sind die endogenen Variablen einer Modellstufe die exogenen der nächsten. Wichtigstes Beispiel dafür ist das Vier-Stufen-Schema der klassischen Verkehrsplanung (Abb. 1.5).
  - Bei **rückgekoppelten Modellen** koppeln die endogenen Variablen eines Modells einer späteren Verkettungsstufe auf die exogenen Variablen eines Modells einer früheren Verkettungsstufe zurück. Beim Vier-Stufen-Modell der Abb. 1.5 können als Folge der Routenwahl einzelne Netzelemente überlastet, also verstaut werden. Die damit verbundenen Resiezeitverlängerungen beeinflussen wiederum die Ziel- und ggf. die Verkehrsmittelwahl.
- Dynamische *vs.* statische Modelle: **dynamische Modelle**, also Modelle mit expliziten dynamischen Zeitbezug, stellen einen Sonderfall der Verkettung im Zeitbereich dar: Die Entwicklung des aktuellen Zeitschrittes hängt von den Ergebnissen

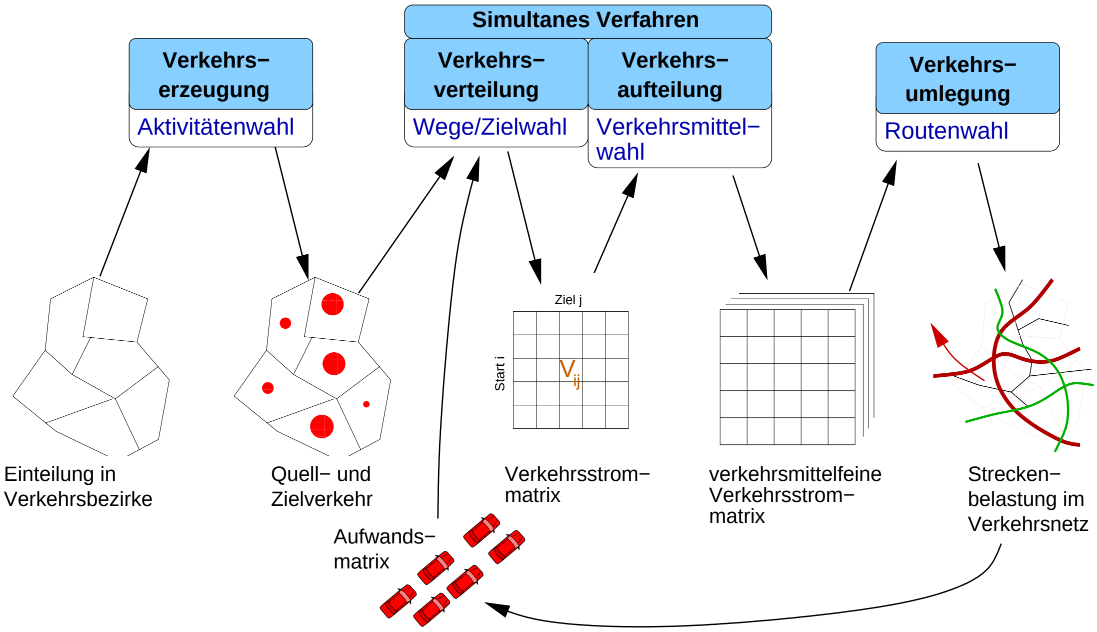


Abbildung 1.5: Das klassische Vierstufenmodell der Verkehrsplanung als Beispiel der Verkettung und Rückkopplung von Modellen. Die endogenen Variablen der ersten Modellstufe "Aktivitätenwahl", die im Verlauf eines Tages anfallenden Verkehrsströme von und nach jedem Bezirk (Quell- bzw. Zielsummen), spielen gleichzeitig die Rolle von exogenen Variablen bei der Modellierung der "Zielwahl". Die endogenen Variablen der Zielwahl (Quelle-Ziel-Relationen bzw. Wege zwischen jeweils zwei Bezirken) sind wiederum die exogenen Variablen der Verkehrsmittelwahl und in weiterer Folge der Routenwahl. Das Ergebnis der Routenwahl führt ggf. zu einer Überlastung einzelner Netzelemente und zu einer Erhöhung der Reisezeit auf einigen Relationen. Diese sogenante "Reisezeitmatrix" bzw. "Aufwandsmatrix" ist nun wiederum eine weitere exogene Variable der Zielwahl ("wenn der Weg von Bezirk i nach j immer verstopft ist, gehe ich vielleicht doch woanders einkaufen"), so dass die Routenwahl auf die Zielwahl rückkoppelt.

des vergangenen Zeitschrittes bzw. der vergangenen Zeitschritte ab (Abb. 1.4). In der vertiefenden Master-Vorlesung) werden wir einen einfachen Typ dieser Modelle kennen lernen, das **dynamische Input-Output-Modell**.

1.3 Zwei Anwendungsbesipiele

Hinweis: In diesem Abschnitt werden konkrete Anwendungsmöglichkeiten der Methoden und Modelle der Verkehrsökonometrie an zwei umfangreicheren Beispielen erläutert. Die Komplexität der mathematischen Modellierung dieser Beispiele geht über die im weiteren Verlauf des Skripts behandelten Modelle und Rechnungen hinaus und soll eher als "Kostprobe" dienen, wohin die Richtung in den vertiefenden Master-Vorlesungen und in der tatsächlichen Anwendung geht.

1.3.1 Staukosten

Die in Deutschland pro jahr anfallenden Kosten durch Staus werden kontrovers diskutiert. Schätzungen reichen von weniger als 1 Mrd €/Jahr über 10-20 Mrd €/Jahr (EU-Weißbuch, 2001) bis zu 100 Mrd €/Jahr (BDI 1999, BMW-Studie 1997). Selbst ein insgesamt negativer (!) Beitrag von Staus zu volkswirtschaftlichen Kosten wird diskutiert (Hermann Knoflacher, "Stehzeuge").

Da die endogenen Variablen, also die verschiedenen Kosten, (quasi-)stetige Größen darstellen,[5] ist hier eine Beschreibung mit *stetigen Modellen* sinnvoll. Die Abschätzung der gegenwärtigen und zukünftigen Staukosten kann in unterschiedlichen Detaillierungsgraden geschehen:

1. **Einfache globale Schätzung:** Hier wird das aktuelle Verkehrsaufkommen durch Mobilitätsbefragungen (siehe Kap. 8) empirisch erhoben und das zukünftige gemäß einem einfachen Modell abgeschätzt, z.B. mit dem Modell (1.12) für beschränktes Wachstum.

   Die durch Staus verursachte Gesamtkosten schätzt man dann durch einen einfachen globalen Ansatz ab, z.B.

   $$K_{\text{stau}} = \Delta K_{\text{Zeit}} + \Delta K_{\text{Treibstoff}} + \Delta K_{\text{Unfall}} + \Delta K_{\text{Umwelt}} \tag{1.15}$$

   Hierin sind

   - ΔK_{Zeit} die Zeit*mehr*kosten durch Staus. Diese kann man z.B. mit Hilfe der durch Mobilitätsbefragungen erhobernen Gesamt-Verkehrsleistung von $L =$ 800 Milliarden Personenkilometer pro Jahr abschätzen. Nimmt man eine mittlere Geschwindigkeit von $\bar{V} = 50\,\text{km/h}$ sowie einen durch Staus verursachten Zeitmehrverbrauch von $p = 5\%$ an und setzt für die Zeitkosten $\dot{K}$ =10 €/Pers/h an, ergibt sich

     $$\Delta K_{\text{Zeit}} = \frac{Lp\dot{K}}{\bar{V}} = 8\,\text{Mrd €/Jahr}. \tag{1.16}$$

     zu ähnlichen Größenordnungen kommt man bei Annahme von z.B. 5 min täglicher Stauzeit für die etwa 40 Millionen Autofahrer.

[5] Quasistetige Variablen sind, streng genommen, diskret, aber die Einheit der Diskretisierung (in unserem Falle Cent) ist um Größenordnungen kleiner als typische Werte der endogenen Variable (Milliarden von €)

- Die Treibstoff-Mehrkosten setzten ein Modell des Treibstoffmehrverbrauchs voraus. Nimmt man z.B. während der Stauzeit einen Verbrauch von $\dot{C}_{\text{Stau}} = 3\,\text{Liter/h}$, einen Treibstoffpreis $K = 1.50$ €/Liter sowie einen mittleren Besetzungsgrad von $B = 1.5\,\text{Pers/Kfz}$ an (Der Leerlaufverbrauch ist meist unter 1 Liter/h, aber die Stop&Go-Phasen haben natürlich einen deutlich höheren Verbrauch), erhält man

$$\Delta K_{\text{Treibstoff}} = \frac{Lp\dot{C}_{\text{Stau}}K}{B\bar{V}} = 2.4\,\text{Mrd €/Jahr}. \tag{1.17}$$

- Schwieriger wird die Abschätzung der durch Staus verursachten *zusätzlichen* Unfälle, z.B. Auffahrunfälle, und noch schwieriger die externen Umweltkosten.

2. **Schätzung mit CR-Funktionen** Dieser Ansatz ist vor allem für abgegrenzte Regionen (z.B. Städte) interessant, für deren Straßennetz eine Verkehrsumlegung vorliegt, d.h. eine modellierte (und durch Querschnittszählungen geprüfte) Verkehrsbelastung Q_l in Kfz/h auf allen gerichteten Kanten (Streckenelemente) l des Verkehrsnetzes. In diesem Falle kann die zeitliche Mehrbelastung direkt durch die **Capacity-Restraint-Funktionen** bzw. **CR-Funktionen** angenähert werden. Diese sind auch Basis der Umlegung (vgl. Abschnitt 7):

$$T_l(Q) = T_{l0}\left[1 + \left(\frac{Q}{K_l}\right)^{\gamma}\right]. \tag{1.18}$$

Hier bezeichnet $T_l(Q)$ die Fahrzeit auf einem Streckenelement (Kapazität K_l) bei der Verkehrsbelastung Q_l und T_0 die Fahrzeit ohne Belastung. Der Exponent γ wird meist zwischen 1 und 5 gewählt. Üblicherweise wird die Umlegung in Stundenabständen durchgeführt. In dieser Stunde t fallen auf einem Netzelement l die Stau-Zeitkosten $\Delta K_{lt} = \dot{K}Q_{lt}[T_l(Q_{lt}) - T_{l0}]$ an. Die Staukosten pro Tag ergeben sich einfach als die Summe über alle Streckenelemente und Stundenintervalle:

$$\Delta K_{\text{Zeit}} = \sum_{t=1}^{24}\sum_{l} \dot{K}Q_{lt}\left[T_l(Q_{lt}) - T_{l0}\right]. \tag{1.19}$$

Mit den CR-Funktionen kann man auch die interessante Frage beantworten, wie sehr man durch eine eigene Fahrt zur Rush-hour die Reisezeit aller anderen betroffenen Verkehrsteilnehmer verlängert. Vom egoistischen Standpunkt aus wären dies externe Kosten, da sie die anderen Fahrer betreffen.[6] Bezeichnet man die durch die eigene Verkehrsteilnahme bei anderen Verkehrsteilnehmern verursachten Zeitmehrkosten mit ΔK_{ext} und die eigenen Zeitmehrkosten durch die Verkehrsüberlastung mit ΔK_{indiv}, ergibt sich für jeden Streckenabschnitt und auch für die gesamte Fahrt das bemerkenswert einfache Ergebnis (vgl. Übungsaufgabe)

$$\Delta K_{\text{ext}} = \gamma \Delta K_{\text{indiv}}. \tag{1.20}$$

[6] üblicherweise werden externe Kosten aber definiert als nichtbezahlte Kosten, die anderen Gruppen (Fußgängern, ÖPNV-Teilnehmern etc) durch die Aktivität der betrachteten Gruppe (hier der Kfz-Nutzer) entstehen. In diesem Sinne sind die gesamten Stauzeitkosten interne Kosten.

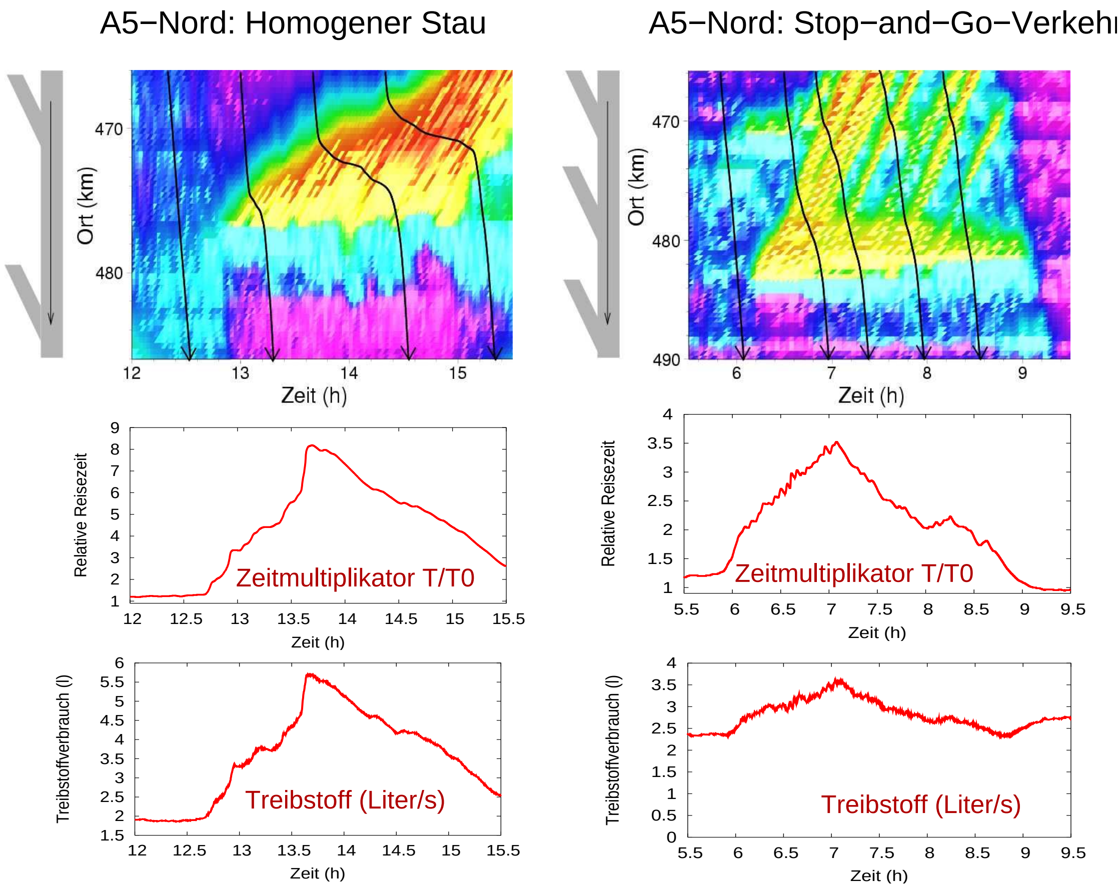


Abbildung 1.6: Empirische Ermittlung der durch zwei konkrete Staus auf der A5 (in der Nähe von Frankfurt) verursachten Zeit- und Treibstoffkosten. Oben: Aus Detektordaten rekonstruierte raumzeitliche Geschwindigkeitsprofile und einige beispielshaft daraus errechneten Trajektorien; Mitte: aus den Trajektorien bestinnte Reisezeit; unten: Aus Fahrzeug-Motor- und Getriebekenndaten eines VW Passat errechneter Treibstoffverbrauch aller im Testabschnitt befindlichen Fahrzeuge. Näheres über die hierbei verwendeten Modelle kann man in der Vorlesung Verkehrsdynamik und -simulation erfahren.

Der Zeitmehrverbrauch bei den anderen Verkehrsteilnehmern ist also gleich dem eigenen Zeitmehrverbrauch, multipliziert mit dem Exponenten der CR-Funktion (vgl. die folgende Aufgabe und die dazugehörige Lösung. Dies kann man plakativ ausdrücken:

"Durch meine Verkehrsteilnahme werden die anderen Verkehrsteilnehmer um den Faktor γ (Exponent der CR-Funktion) mehr verzögert als ich selbst"

Mehr Details zu dieser Problematik finden sich im Abschnitt 7 dieses Skripts.

Aufgabe:

Leiten Sie das Ergebnis (1.20) für einen Streckenabschnitt her, indem Sie für die Zeit T_l zum Durchfahren dieses Streckenabschnitts die Nachfrage Q_l um $\Delta Q_l = 1/T_l$ erhöhen (genau dies bewirkt das Durchfahren). Schätzen Sie die resultierende Reisezeiterhöhung durch $\Delta T_l = T_l'(Q_l)\Delta Q_l$ ab und multiplizieren Sie diese Erhöhung mit der Zahl $n = Q_l T_l$ der betroffenen Fahrzeuge.

3. **Mikroskopische Schätzung mit Fahrzeugtrajektorien** Bei dieser detailliertesten Abschätzung beschränkt man sich wegen der Datenlage in der Regel auf eine einzelne Stausituation. Die Fragestellung lautet:

   Wieviel Zeitverlust und wieviel Treibstoff-Mehrverbrauch wird durch ein konkretes Stau-Ereignis verursacht?

   Diese Frage kann mit keiner anderen Methode beantwortet werden, da sowohl der Zeit-Mehrverbrauch als auch der Treibstoffverbrauch zum Beispiel davon abhängt, ob es sich um einen gleichmäßigen Stau ("zähfließender verkehr"), um Stop-and-Go-Verkehr oder um einen Totalstillstand handelt.

   Die Zeitverzögerung wird dabei direkt anhand der (simulierten oder gemessenen Fahrzeugtrajektorien abgelesen. Für den Treibstoffverbrauch kommen dezidierte mikroskopische Modelle zum Einsatz, welche den instantanen Verbrauch (Treibstoffmenge pro Zeiteinheit) in Abhängigkeit der Geschwindigkeit, der Beschleunigung sowie der Motor- und Fahrzeugparameter berechnet.[7]

   Abbildung 1.6 zeigt Beispiele für typische Stausituationen. In der Vorlesung Verkehrsdynamik und -simulation (innerhalb des Moduls "Vertiefung Verkehrsökonometrie") werden die zugrundeliegenden Modelle ausführlich behandelt.

1.3.2 Verkehrsmittelwahl

Die Verkehrsmittelwahl ist ein klassisches Beispiel für die in der Verkehrsökonometrie sehr wichtigen Modelle der **diskreten Wahltheorie**, welche in der vertiefenden Master-

[7] Es ist in Extremfällen – genau wie in der Wirklichkeit – nicht einmal ausgeschlossen, dass Staus zu *Minderverbräuchen* führen. Dies ist insbesondere bei gleichmäßig fließenden gebundenen Verkehr im Vergleich zu sehr schneller freier Fahrt möglich.

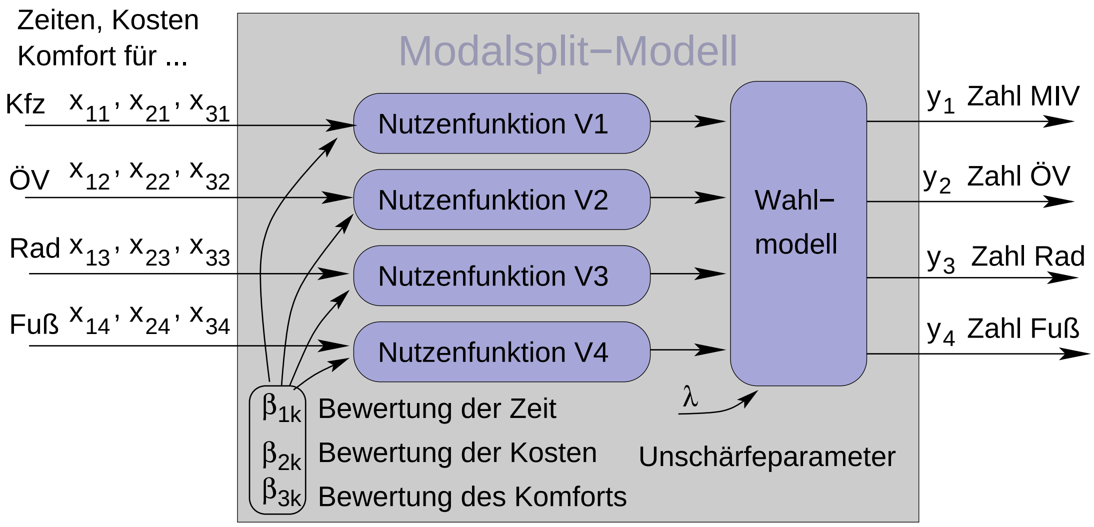


Abbildung 1.7: Beispiel eines ökonometrischen Modells mit – zumindest auf der Ebene des Individuums – diskreten endogenen Variablen: Modell der Verkehrsmittelwahl. Die exogenen Variablen x_m beeinflussen die Wahlentscheidung: Komplexe Reisezeit, Kosten und Bequemlichkeit für jedes infrage kommende Verkehrsmittel. Weitere Variable wie Geschlecht und Alter des Individuums bzw. Altersklasse, Quelle-Ziel-Gruppe (Kollektiv) kommen hinzu. Im eigentlichen Wahlmodell wird zunächst mit Hilfe der Wichtungsparameter β_1 und β_2 Nutzenfunktionen gebildet und diese dann, unter Berücksichtigung der Unschärfe λ des Zufallsnutzens als weiteren Parameter, in eine Wahlentscheidung (Anwendung auf der individuellen Ebene) bzw. einen Anteilswert (Anwendung auf ein Kollektiv) Zunächst wird unter Die exogenen Ausgängen mit diskreten endogenen

Vorlesung behandelt werden. Während die exogenen (erklärenden) Variablen diskret oder stetig sein können, ist der Ausgang (endogene bzw. erklärte Variablen) auf jedem Fall diskret, da man unter i.A wenigen Alternativen zu wählen hat. Die meisten Modelle der diskreten Wahltheorie sind mehrstufig (vgl. Abb. 1.7):

- Zunächst wird für jede Alternative k eine generalisierte Nutzenfunktion definiert, die den Zeitaufwand (komplexe Reisezeit, d.h. Haustür - zu - Haustür Reisezeit), die Kosten, die Bequemlichkeit, die Zuverlässigkeit und vieles mehr enthalten kann. Der deterministische Anteil V dieser Nutzenfunktion[8] ist oft quasilinear, d.h. linear in den Parametern, nicht jedoch notwendigerweise linear in den exogenen Variablen $\boldsymbol{x} = (x_1, ..., x_m, ..., x_M)$:

$$V_k(\boldsymbol{x}) = \beta_1 f_1(\boldsymbol{x}) + \beta_2 f_2(\boldsymbol{x}) + \dots$$

[8]Das Konzept des Zufallsnutzens und wie damit beispielsweise ein Logit-Modell der Art (1.21) hergeleitet werden kann, werden in der vertiefenden Master-Vorlesung betrachtet.

- Jede Alternative wird nun anhand ihrer Nutzenfunktion bewertet. Um allgegenwärtige Unsicherheiten durch nicht berücksichtigte erklärende Variablen einerseits und irrationales Verhalten bei der Wahlentscheidung andererseits zu berücksichtigen, wird jedoch nicht unbedingt die Alternative mit dem höchsten (bewerteten) Nutzen gewählt, sondern mit geringeren Wahrscheinlichkeiten auch andere. Das Ausmaß, mit der auch suboptimale Alternativen gewählt werden können, wird dabei durch einen weiteren Parameter λ festgelegt. Die Auswahlwahrscheinlichkeiten können dann z.B. in der Form
$$P(A_k) = \frac{B(V_k, \beta)}{\sum_{k'} B(V_{k'}, \beta)} \tag{1.21}$$
definiert werden. Im bekanntesten dieser Modelle, dem Multinomial-Logit-Modell, ist die Bewertungsfunktion durch $B(\boldsymbol{x}; \boldsymbol{\beta}, \lambda) = e^{\lambda V(\boldsymbol{x};\boldsymbol{\beta})}$ gegeben. Die Nutzenfunktionen werden meist durch die negativen effektiven Zeitaufwände definiert, d.h. alle anderen Aufwandsarten (Kosten, Unbequemlichkeit ...) werden mit den Parametern β_j auf Zeitkosten umgerechnet (vgl. Abb. 1.7). Typische Werte von V_k liegen bei -30 min und typische Werte von λ sind durch $\lambda^{-1} = 10$ min charakterisiert.

Um dieses Modell zu kalibrieren, werden anhand von Befragungen Messgleichungen folgender Form aufgestellt:

	Alter	Ge-schl.	Zeit-bedarf Rad	Wetter	Kompl. Reisezeit ÖPNV	Ad-hoc Kosten ÖPNV	Wahl-entsch. Rad	Wahl-entsch. ÖPNV
Variable	x_1	x_2	x_3	x_4	x_5	x_6	y_{1i}	y_{2i}
Person 1	30	w	20 min	schlecht	30 min	1.00 €	0	1
Person 2	24	m	11 min	schön	20 min	2.00 €	1	0
Person 3	27	m	34 min	schön	15 min	2.00 €	0	1
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

In der Tabelle wurde eines der einfachsten nichttrivialen Sachverhalte gewählt:

- Es gibt nur zwei Alternativen: Rad und ÖPNV
- Die Nutzenfunktion enthält nur zwei "Kostenarten": Komplexe Reisezeit und Geldkosten (ÖPNV) bzw. Schlechtwetter (Rad). Die Kosten sind dabei als *Ad-hoc*-Kosten zu verstehen, also als zusätzliche Kosten, wenn man ÖPNV wählt: Dauerkarteninhaber haben deshalb Ad-Hoc-Kosten von Null.
- Neben den Abhängigkeit von weggebundenen erklärenden Variablen, also Reisezeit und Kosten für jede Alternative, wird noch eine Abhängigkeit von sozioökonomischen erklärenden Variablen (Alter und Geschlecht) berücksichtigt.

Anhand dieser Tabelle werden in der Modellkalibrierung die Werte der Modellparameter β_j bestimmt, für welche die durch die Werte y_{1i} und y_{2i} charakterisierten Beobachtungen im Modell am wahrscheinlichsten sind, das Produkt der entsprechenden Wahrscheinlichkeiten also maximal wird.

Aggregierung

Grundsätzlich sind Modelle der Verkehrsmittelwahl **mikroskopisch**, betrachten also jede Entscheidung einzeln. Damit kann auch nur eine der endogenen Variablen y_k für Alternative k gleich 1 sein, während alle anderen $y_{k'}$, $k' \neq k$, gleich 0 sein müssen. In drei Fällen ist eine *Aggregierung*, also eine **makroskopische Hernagehensweise** möglich, sinnvoll oder sogar notwendig:[9]

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

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

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

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

Aggregierung dieser Daten in Entfernungsklassen

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

1.4 Verwendete Symbole

Y	Abhängige bzw. erklärte bzw. endogene Variable. Bei Mehrgleichungsmodellen bekommen diese den Index k.
x_m, $m = 1, \ldots, M$	Unabhängige bzw. erklärende bzw. exogene Variablen
$\boldsymbol{x}$	Gesamtheit der exogenen Variablen, als Spaltenvektor geschrieben
y_i, x_{mi}, $i = 1, \ldots, n$	Wert der abhängigen bzw. der m-ten unabhängigen Variablen bei der i-ten Messung bzw. dem i-ten Element der Stichprobe.
β_0	Achsabschnitt (engl. *intercept*)
β_j, $j = 1, \ldots, J$	Lineare Anstiegsparameter
$\boldsymbol{\beta}$	Gesamtheit der Modellparameter, als Spaltenvektor geschrieben
$\hat{\beta}_j$, $\hat{Y}$	Das "Dach" ist ein Symbol für geschätzte Größen, also geschätzte Parameterwerte, exogene Variable des geschätzten Modells, etc.
ϵ	Ein additiver i.i.d. Zufallsterm mit Erwartungswert 0 und der Varianz σ^2
z_m, w	Transformierte exogene und endogene Variablen (mit dem Ziel, ein lineares Modell in den transformierten Variablen zu erhalten)
V_k	Deterministische Nutzenfunktion der Alternative k bei Wahlmodellen