

Simulation und Analyse dynamischer Systeme in den Wirtschafts- und Sozialwissenschaften

Teil 2 von 4

Seite 164-398

Prof. Dr. Eckart Zwicker
Technische Universität Berlin
Fachgebiet Unternehmensrechnung und Controlling

2. Formen und Erschließungsmethoden dynamischer MZÄ-Modelle

Während bisher die allgemeinen Fragen des Aufbaus, der Analyse sowie der empirischen Akzeptierbarkeit dynamischer Modelle behandelt wurden, wendet sich dieses Kapitel den speziellen Typen dynamischer Modelle zu. Ausgehend von bestimmten polaren Begriffspaaren dynamischer Modelle, werden die formale Struktur dieser Modelltypen, ihre empirische Interpretation, die speziellen Verfahren ihrer Implikationenaufdeckung und die Existenz typenspezifischer Implikationen erläutert.

2.1. Lineare und nichtlineare Modellformen

Die Unterscheidung zwischen linearen und nichtlinearen Modellen resultiert aus dem Umstand, daß beide Modellformen sowohl verschiedene Methoden der Hypothesengewinnung durch statistische Schätzungen, als auch der Implikationenerschließung verlangen. Die Klassifizierung ist auch insofern fruchtbar, als sich für beide Modellformen typenspezifische empirisch interpretierbare Implikationen aufweisen lassen.

Die Abgrenzung zwischen linearen und nichtlinearen Modellen führt zur Teilkasse der linearen Modelle, für die eine geschlossene Theorie der Implikationenaufdeckung zur Verfügung steht. Diese Theorie linearer dynamischer Modelle wird in dem folgenden Abschnitt eingehend erörtert. Ihre Darstellung liefert grundlegende Einsichten über die Verknüpfung zwischen Modellprämissen und Modellimplikationen. So wohl die formelmäßige Aufweisung dieser Verknüpfungswege, als auch die formale Präzisierung der Modellimplikationen, schafft eine präzise und eindeutige Basis für die konkrete Durchführung, aber auch generelle Bewertung dynamischer Modellanalysen.

Die wesentlich kürzeren Ausführungen über nichtlineare dynamische Modelle charakterisieren die Formen nichtlinearer Hypothesen und versuchen, eine Übersicht über den Stand der Verfahren zur deduktiven Implikationenaufdeckung zu geben.

2.1.1. Lineare Modellformen

In der Algebra spricht man immer dann von einer linearen Funktion, wenn sich die abhängige Variable aus der Summe der mit einem konstanten Wert multiplizierten unabhängigen Variablen und einer Konstanten ergibt. Als Beispiel sei der Ausdruck

$$Y = 0,5X + 0,3Z + 10$$

angeführt.

In analoger begrifflicher Verwendung könnte man unter einem linearen dynamischen Modell einen Zusammenhang verstehen, in dem alle endogenen Variablen durch eine Linearkombination der zeitveränderlichen endogenen und exogenen Variablen bestimmt werden. Als Beispiel sei auf den Ansatz

$$\begin{aligned} Y_1(t) &= 0,5Y_1(t-1) + 0,7E(t-1) + Y_2(t-2) + 100 \\ Y_2(t) &= 0,3Y_2(t-2) + 1,1Y_1(t-1) \end{aligned}$$

verwiesen. Diese begriffliche Fassung der Linearität ist jedoch für dynamische Modelle zu einschränkend. Im vorliegenden Fall handelt es sich vielmehr um die allerdings wichtigste Teilkasse linearer Modelle: den linearen Modellen mit konstanten Koeffizienten. Zu den linearen Modellen zählen jedoch ebenfalls die linearen Modelle mit zeitvariablen Koeffizienten.

Das angeführte Beispiel würde in diese Modellklasse übergehen, falls die Gleichung von Y_2 durch

$$Y_2(t) = 0,3^t Y_2(t-2) + 1,1Y_1(t-1)$$

ersetzt würde. Der Koeffizient vor $Y_2(t-2)$ verändert sich in diesem Fall mit der Zeit, d.h. es handelt sich um einen zeitvariablen Koeffizienten. Während wir uns den linearen Modellen mit zeitvariablen

Koeffizienten nur relativ kurz zuwenden, werden die linearen Modelle mit konstanten Koeffizienten intensiv erörtert, da sie in großem Umfang zur Modellierung sozialer Systeme verwendet werden.

A. Lineare Modellformen mit zeitvariablen Koeffizienten

Modelle dieses Typs werden relativ selten zur Beschreibung sozialer Zusammenhänge verwendet.

Als Beispiel sei das bereits beschriebene Modell von VIDALE und WOLFE angeführt. Es wurde durch folgende Beziehung beschrieben.¹

$$U(t) = rW(t-1) - \left[\frac{rW(t-1)}{S} + \zeta - 1 \right] U(t-1)$$

Man erkennt, daß es sich um eine Differenzengleichung ersten Grades mit einem zeitvariablen Koeffizienten handelt. Denn die Werbeausgaben $W(t-1)$ ändern sich in der Zeit.

Nach der Einordnung des Modelltyps stellt sich zum einen die Frage, welche wünschenswerten Implikationen es in diesem Ansatz aufzudecken gibt, und zum anderen, welche Methoden der Erschließung von Modellimplikationen für diesen Modelltyp zur Verfügung stehen.

Von praktischem Interesse dürfte der Einsatz unterschiedlicher Werbestrategien auf den Umsatz sein. Man kann beispielsweise die prozyklische Werbepolitik

$$W(t) = aU(t-1) \quad a > 0$$

mit der antizyklischen Politik

$$W(t) = b[\bar{W} - U(t-1)] \quad b > 0, \bar{W} > 0$$

oder der zyklusneutralen Politik

$$W(t) = A \quad A > 0$$

vergleichen.

Als Implikation ist daher vor allem der Zeitverlauf des Umsatzes U von Interesse. Damit stellt sich die Frage, ob es möglich ist, eine Funktionslösung linearer Modelle mit zeitvariablen Koeffizienten zu entwickeln.

¹ Siehe Seite 25

Generell lässt sich sagen: Für Differenzengleichungen ersten Grades mit variablen Koeffizienten gibt es einen allgemeinen Ansatz zur Bestimmung einer Funktionslösung. Für Differenzengleichungen höheren Grades existiert keine generelle Methode zur Ermittlung von Funktionslösungen.²

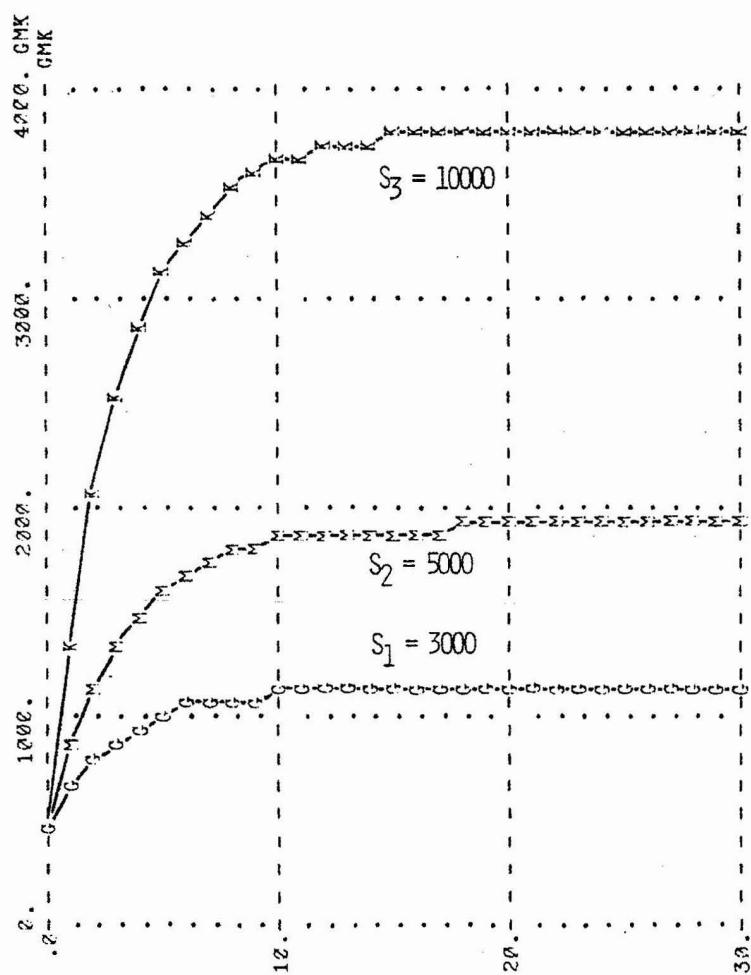


Abb. 21.1 Zeitverlauf des Umsatzes bei verschiedenen Sättigungsniveaus S_1 bis S_3 im Falle einer prozyklischen Werbepolitik $W(t)=0,1^3 U(t-1)$

² Vgl. [83, S. 59]

Wir wollen uns mit dieser generellen Methode nicht befassen, da sie wegen ihrer Beschränkung auf Gleichungen ersten Grades für praktische Anwendungen nahezu bedeutungslos ist.

Bei derartigen Modellformen sollte man daher von vornherein versuchen, den Zeitpfad mit Hilfe von Regressionslösungen zu ermitteln. Gehen wir von den von VIDALE und WOLFE ermittelten Werten für die Parameter von $r=0,5$, $S=5000$ und $\xi=0,1$ aus und unterstellen einen Anfangswert von $U(0)=500$, so ermittelt sich die Regressionslösung nach

$$U(t) = \left[\frac{0,5W(t-1)}{5000} + 0,1 - 1 \right] U(t-1) + 0,5W(t-1)$$

Die Sensitivität des Zeitverlaufes von U bezüglich des Sättigungsniveaus S zeigt Abbildung 21.1.³

Man erkennt, daß die Umsatzentwicklung eine starke Sensitivität bezüglich des Sättigungsniveaus aufweist. Nach diesem Beispiel wollen wir uns nunmehr den linearen Modellen mit zeitkonstanten Koeffizienten zuwenden.

B. Lineare Modellformen mit zeitkonstanten Koeffizienten

Dieser Modelltyp muß aus zwei Gründen sehr ausführlich behandelt werden. Zum einen gehören viele der heute verwendeten Modelle diesem Typ an. Auch unser Standardbeispiel eines MA-Modells fällt, wie man leicht erkennt, in diese Modellkategorie. Zum zweiten steht, wie erwähnt, für diesen Modelltyp eine geschlossene Theorie der Implikationenaufdeckung zur Verfügung. Dies ist von genereller Bedeutung, weil prinzipiell jedes nichtlineare Modell durch ein lineares Modell approximiert werden kann, so daß Einsichten über den Charakter linearer Modelle auch für die Beurteilung nichtlinearer Modelle von Interesse sind.

Es ist heute Mode geworden, unter Verwendung einfacher zu lernender Simulationssprachen komplexe nichtlineare dynamische Modelle zu simulieren, ohne dabei über Kenntnisse des strukturellen Aufbaus und der Analysemethoden derartiger Modelle zu verfügen. Nach Ansicht des Verfa-

³ Zur Simulationstechnik siehe Seite 522ff.

sers ist unter diesen Umständen ein ernsthaftes Arbeiten mit dynamischen Simulationsmodellen kaum möglich. Auch wenn im Rahmen der heute vorwiegend praktizierten Simulationen dynamischer Modelle die im weiteren zu erörternde Theorie nicht unmittelbar zur Anwendung kommt, so liefert ihre Kenntnis dennoch fruchtbare Leitlinien und Beurteilungsprinzipien für das Arbeiten mit dynamischen Modellen. So erhalten Konzepte und Begriffe, wie Gleichgewichtszustand eines Systems, Stabilitätstypen, Multiplikator oder Verhaltensweisen von Systemen, erst durch ihre Explizierung im Rahmen eines Kalküls die Präzision, die zu einer klaren Beurteilung ihres Stellenwertes von Bedeutung ist.

Da die Bestimmung des Zeitverlaufes einer endogenen Variablen die Ausgangsbasis für die Aufdeckung fast sämtlicher relevanter Modellimplikationen darstellt, wollen wir uns im folgenden ausführlich mit der Ermittlung der Funktionslösung einer endogenen Variablen beschäftigen. Anhand einer linearen Endgleichung zweiten Grades werden die verschiedenen Typen von Funktionslösungen hergeleitet und auf den Fall einer Endgleichung beliebigen Grades verallgemeinert. Auf dieser Basis erfolgt eine Erörterung mathematisch eindeutig beschreibbarer typenspezifischer Implikationen linearer Modelle. Abschließend wird die Theorie linearer Systeme unter Verwendung der Operatoren- und Matrizenrechnung in einem umfassenden Rahmen behandelt.

a) Zeitpfadermittlung durch Funktionslösungen

Die Bestimmung der Funktionslösung einer endogenen Variablen Y erfolgt wie erwähnt anhand ihrer Endgleichung. Fassen wir die Komponenten der exogenen Variablen in (12.10) mit $E(t)$ zusammen, d.h.

$$E(t) = \sum_{\eta=0}^s g_\eta E(t-\eta) \quad (21.1)$$

dann erhalten wir die Erklärungsform

$$Y(t) = \omega_1 Y(t-1) + \omega_2 Y(t-2) + \dots + \omega_n Y(t-n) + E(t) \quad (21.2)$$

und die Standardform

$$Y(t) + a_1 Y(t-1) + a_2 Y(t-2) + \dots + a_n Y(t-n) = E(t) \quad (21.3)$$

welche uns im folgenden als Ausgangsbasis dienen sollen. Man bezeichnet (21.2) und (21.3) als Endgleichungen n-ten Grades, wenn a_n bzw. $a_n \neq 0$ sind. Entsprechend dieser Festlegung ist die Endgleichung der Variablen Y in (12.9) eine Endgleichung 2-ten Grades. Wie erwähnt, ist es oft notwendig, erst die Endgleichung einer Variablen zu ermitteln.⁴

Es bieten sich verschiedene Verfahren an. Das sogenannte Einsetzungsverfahren wurde bereits zur Berechnung der Endgleichung von Y, (12.9), angewendet.⁵ Seine Anwendbarkeit dürfte aber wohl zumeist auf Modelle mit nicht mehr als vier bis fünf Variablen beschränkt bleiben, da der Rechenaufwand mit wachsender Variablenzahl stark ansteigt. Betrachten wir beispielsweise das anfangs bereits kurz erwähnte, noch recht einfache Modell der Anspruchsniveauanpassung von SIMON und MARCH, welches zuvor kurz beschrieben werden soll [128, S.48].

Das Modell beschreibt die Beziehung zwischen den für das Verhalten einzelner Organisationsteilnehmer relevanten psychischen Variablen: Zufriedenheitsniveau (Z), Anspruchsniveau (A), Suchintensität (S) und erwartete Belohnung (B). Die das Modell konstituierenden Hypothesen werden durch folgende Sätze gekennzeichnet:

- (1) Je niedriger das Zufriedenheitsniveau (Z) einer Verhaltenseinheit, um so stärker wird ihre Suchintensität (S) nach neuen Alternativen sein.
- (2) Je höher die Suchintensität (S), um so höher ist die erwartete Belohnung (B).
- (3) Je höher die erwartete Belohnung (B), um so höher das Zufriedenheitsniveau (Z).
- (4) Je höher die erwartete Belohnung (B), um so höher wird das Anspruchsniveau (A).
- (5) Je höher das Anspruchsniveau (A), um so höher das Zufriedenheitsniveau (Z).

⁴ Zur exakten mathematischen Kennzeichnung linearer Systeme, von denen eine Endgleichungsform berechenbar ist, siehe Seite 244

⁵ Vgl. Seite 37 f.

Dieses verballogische Aussagensystem wird durch folgende Gleichungen formalisiert:⁶

$$\text{Satz 1: } S(t) = \beta[\bar{Z} - Z(t)] \quad \text{mit } \bar{Z} > 0, \beta > 0 \quad (21.4)$$

\bar{Z} stellt das Sättigungsniveau der Zufriedenheit dar, bei dessen Erreichen die Suchintensität zum Erliegen kommt.

$$\text{Satz 2: } B(t+1) - B(t) = \gamma[S(t) - b - cB(t)] \quad \text{mit } \gamma > 0, b \geq 0, c > 0 \quad (21.5)$$

Aus der Beziehung erkennt man, daß eine bestimmte Suchintensität in Höhe von $[b + cB(t)]$ erforderlich ist, um die erwartete Belohnung konstant zu halten.

$$\text{Satz 3 und 5: } Z(t) = B(t) - A(t) \quad (21.6)$$

$$\text{Satz 4: } A(t+1) - A(t) = \alpha[B(t) - A(t) + a] \quad \alpha > 0, a > 0 \quad (21.7)$$

In dieser Gleichung kommt die zusätzliche Hypothese zum Ausdruck, daß zur Aufrechterhaltung des Anspruchsniveaus eine erwartete Belohnung in Höhe von $[A(t) - a]$ erforderlich ist.

Wie man sich überzeugen kann, ist die Ermittlung der Endgleichung von $A(t)$ durchaus kein trivialer Einsetzungsprozeß. Die Gleichungen (21.5) und (21.7) werden zuerst auf die Erklärungsform

$$B(t) = [1 - c\gamma]B(t-1) + \gamma S(t-1) - \gamma b \quad (21.8)$$

$$A(t) = [1 - \alpha]A(t-1) + \alpha B(t-1) + \alpha a \quad (21.9)$$

überführt.

(21.6) in (21.4) ergibt:

$$S(t) = \beta\bar{Z} - \beta B(t) + \beta A(t) \quad (21.10)$$

Verschiebt man das Zeitargument in (21.10) um eine Periode und setzt den $S(t-1)$ erklärenden Ausdruck in (21.8) ein, so ist das System auf die Gleichungen (21.9) und (21.11) reduziert.

$$B(t) = [1 - c\gamma]B(t-1) + \gamma\beta\bar{Z} - \gamma\beta B(t-1) + \gamma\beta A(t-1) - \gamma b \quad (21.11)$$

Formt man (21.9) um, so ergibt sich

$$B(t-1) = \frac{1}{\alpha} A(t) + \frac{1}{\alpha} [\alpha - 1]A(t-1) - a \quad (21.12)$$

⁶ Das Modell wurde von SIMON und MARCH in stetiger Form entwickelt und ist im folgenden in Differenzengleichungsform umgewandelt.

Wenn der Zeitindex von $B(t)$ eine Periode nach vorne geschoben wird, so ergibt sich aus (21.12)

$$B(t) = \frac{1}{\alpha} A(t+1) + \frac{1}{\alpha} [\alpha-1]A(t) - a \quad (21.13)$$

Setzt man (21.12) und (21.13) in (21.11) ein, so folgt

$$\begin{aligned} \frac{1}{\alpha} A(t+1) + \frac{1}{\alpha} [\alpha-1]A(t) - a &= [1-\gamma][\frac{1}{\alpha} A(t) + \frac{1}{\alpha} [\alpha-1]A(t-1) - a] + \gamma \bar{Z} \\ &\quad - \gamma \beta [\frac{1}{\alpha} A(t) + \frac{1}{\alpha} [\alpha-1]A(t-1) - a] + \gamma \beta A(t-1) - \gamma b \end{aligned}$$

oder

$$\begin{aligned} A(t+1) + [\alpha-1]A(t) - \alpha a &= [1-\gamma][A(t) + [\alpha-1]A(t-1) - \alpha a] + \gamma \beta \bar{Z} - \\ &\quad - \gamma \beta [A(t) + [\alpha-1]A(t-1) - \alpha a] + \gamma \beta a A(t-1) - \alpha \gamma b \end{aligned}$$

oder

$$\begin{aligned} A(t+1) + [\alpha-1]A(t) &= A(t) + [\alpha-1]A(t-1) - \gamma \alpha A(t) - \gamma [\alpha-1]A(t-1) + \\ &\quad + \alpha \gamma \alpha + \alpha \beta \bar{Z} - \gamma \beta A(t) - \gamma \beta [\alpha-1]A(t-1) + \\ &\quad + \alpha \alpha \gamma \beta + \gamma \beta \alpha A(t-1) - \alpha \gamma b. \end{aligned}$$

faßt man die Glieder nach Verzögerungen von $A(t)$ zusammen, ergibt sich

$$\begin{aligned} A(t+1) &= [2-\alpha-\gamma-\beta]A(t) + [\alpha-1-\gamma\alpha+\gamma+\beta]A(t-1) + \alpha \gamma \alpha + \alpha \beta \bar{Z} + \\ &\quad + \alpha \alpha \gamma \beta - \alpha \gamma b \end{aligned}$$

Eine Verschiebung des Zeitargumentes um eine Periode liefert die Erklärungsform der Endgleichung

$$\begin{aligned} A(t) &= [2-\alpha-\gamma-\beta]A(t-1) + [\alpha-1-\gamma\alpha+\gamma+\beta]A(t-2) + \alpha \gamma \alpha + \alpha \beta \bar{Z} + \\ &\quad + \alpha \alpha \gamma \beta - \alpha \gamma b \end{aligned}$$

Das beschriebene Einsetzungsverfahren ist im allgemeinen nur für die Ermittlung von Endgleichungen ersten und zweiten Grades zu empfehlen. In anderen Fällen ist es zu umständlich und sollte durch Verfahren, welche mit Operatoren arbeiten, ersetzt werden. Diese Verfahren werden jedoch erst später besprochen, weil weitere mathematische Voraussetzungen zu ihrem Verständnis erforderlich sind. Da wir uns im fol-

genden ausschließlich mit der Funktionslösung von Endgleichungen ersten und zweiten Grades befassen, reicht das Einsetzungsverfahren fürs erste aus.

Zur Ermittlung der Funktionslösung einer endogenen Variablen ist es wichtig, zwischen homogenen und inhomogenen Endgleichungen zu unterscheiden. Eine homogene Endgleichung liegt vor, wenn $E(t)$ in (21.2) oder (21.3) den Wert $E(t)=0$ ($t=0,1,\dots$) annimmt. Wird $E(t)$ dagegen durch einen Zeitpfad beschrieben, der nicht ständig Null ist, dann spricht man von einer inhomogenen Endgleichung. Der Ausdruck

$$A(t) = 0,5A(t-1) + 0,3A(t-2) \quad (21.14)$$

ist damit eine homogene Endgleichung, während

$$A(t) = 0,5A(t-1) + 0,3A(t-2) + 10^t \quad (21.15)$$

eine inhomogene Endgleichung darstellt.

aa) Funktionslösung von Endgleichungen ersten Grades

α) Funktionslösung homogener Endgleichungen ersten Grades

Eine homogene Endgleichung ersten Grades nimmt mit $n=1$, $\omega_1=a$ und $E(t)=0$ ($t=0,1,\dots$) in (21.2.) die Form

$$Y(t) = aY(t-1) \quad (21.16)$$

an. Unterstellen wir, es sei beispielsweise die Gleichung

$$Y(t) = 0,3Y(t-1) \quad (21.17)$$

mit dem Anfangswert $Y(0)=100$ gegeben. Die Regressionslösung von (21.17) läßt sich recht einfach ermitteln. Wir wollen aber $Y(t)$ als eine Funktion eines geschlossenen Formelausdruckes $F(t)$, d.h.

$$Y(t) = F(t) \quad (21.18)$$

beschreiben, mit der Folge, daß man durch Einsetzung von beispielsweise $t=10$ in die rechte Seite von (21.18) einen numerischen Wert für $Y(10)$ erhält.

Zu diesem Zweck betrachten wir das folgende Schema

$$\begin{aligned}Y(1) &= 0,3Y(0) = 0,3^1 Y(0) \\Y(2) &= 0,3Y(1) = 0,3^2 Y(0) \\Y(3) &= 0,3Y(2) = 0,3^3 Y(0)\end{aligned}$$

$$\vdots$$

$$Y(t) = 0,3Y(t-1) = 0,3^t Y(0)$$

in welchem die Variablenwerte zu den verschiedenen Zeitpunkten durch sukzessive Einsetzung der jeweils über jeder Zeile stehenden Gleichungen auf $Y(0)$ zurückgeführt werden.

Das Schema führt zu der Funktionslösung

$$Y(t) = Y(0)0,3^t \quad t=0,1,2,\dots \quad (21.19)$$

Für $Y(0)$ können wir den unterstellten Wert 100 einsetzen und erhalten die Funktionslösung

$$Y(t) = 100*0,3^t \quad (21.20)$$

Man erkennt, daß für 100 jeder beliebige Anfangswert $Y(0)$ hätte gewählt werden können, ohne die Gültigkeit der Lösung zu verletzen. Ersetzt man in dem entwickelten Schema den Wert 0,3 durch die einen beliebigen Wert repräsentierende Zahl a , so zeigt sich, daß die Herleitung auch für die Verallgemeinerung gilt und zu der allgemeinen Lösung von (21.16)

$$Y(t) = Y(0)a^t \quad (21.21)$$

führt.

Die vorangegangene Betrachtung legt es nahe, zwischen zwei Arten von Funktionslösungen einer Endgleichung zu unterscheiden: den generellen und den speziellen Lösungen.

Die spezielle Lösung einer Endgleichung beschreibt den numerisch konkreten Verlauf einer endogenen Variablen $Y(t)$. Das ist stets nur dann möglich, wenn die Parameter und Anfangswerte der Endgleichung für Y numerisch konkretisiert sind. (21.20) ist eine spezielle Lösung. Wir erkennen, daß in diesem Fall sowohl die Anfangswerte als auch die Koeffizienten der Endgleichung einen numerischen Wert besitzen.

Generelle Lösungen kann man nach anfangswertgenerellen und parametergenerellen Lösungen unterscheiden.

In anfangswertgenerellen Lösungen werden die Anfangswerte nicht durch Zahlen, sondern durch Buchstaben definiert. Sie umfassen damit den gesamten Bereich alternativ möglicher Anfangswerte. Gleichung (21.19) ist hierfür ein Beispiel.

Parametergenerelle Lösungen zeichnen sich dadurch aus, daß auch die neben den Anfangswerten in der Lösung auftretenden weiteren Parameter durch Buchstaben repräsentiert werden. Eine parametergenerelle Lösung umfaßt daher die Gesamtheit aller möglichen Einzellösungen.

Mit diesen Unterscheidungen zeigt sich schon der Vorteil einer Funktionslösung gegenüber einer Regressionslösung.

Mit einer Regressionslösung kann man nur den Zeitverlauf ermitteln, der durch eine spezielle Funktionslösung beschrieben wird. Mit einer parametergenerellen Lösung wie (21.21) kann man gewissermaßen auf einen Blick die Eigenschaft des Systems beurteilen. So wird beispielsweise im Falle $|a| < 1$ das System gegen Null konvergieren. Bei einer Untersuchung des Systems anhand von Regressionslösungen wäre es dagegen nicht möglich, diesen Konvergenzbereich durch beliebig viele Regressionslösungen zwingend abzuleiten.

β) Funktionslösung inhomogener Endgleichungen ersten Grades

Wählen wir in (21.2) $n=1$, $E(t) \neq 0$ ($t=0,1,\dots$) und setzen $\omega_1 = a$, dann erhalten wir die inhomogene Endgleichung ersten Grades

$$Y(t) = aY(t-1) + E(t) \quad (21.22)$$

Unser Ziel ist es, die parametergenerelle Funktionslösung von (21.22) zu gewinnen.

Folgendes Verfahren führt zu einer ersten Gliederung des Lösungssatzes:

Wir unterstellen die Existenz einer speziellen Lösung der inhomogenen Gleichung und wollen sie als $\bar{Y}(t)$ bezeichnen. Für $Y(t) = 0,5Y(t-1) + 50$ ist beispielsweise $\bar{Y}(t) = 100$ eine spezielle Lösung.

Die Subtraktion der Komponenten der speziellen Lösung \bar{Y} von dem Ansatz (21.22)

$$\begin{aligned}
 Y(t) &= aY(t-1) + E(t) \\
 -[\bar{Y}(t) = a\bar{Y}(t-1) + E(t)] \\
 \underline{Y(t) - \bar{Y}(t) = a[Y(t-1) - \bar{Y}(t-1)]} &\quad (21.23)
 \end{aligned}$$

ist deswegen statthaft, weil ex definitione auf jeder Seite der Gleichung die gleichen Werte subtrahiert werden.

Definiert man

$$\tilde{Y}(t) = Y(t) - \bar{Y}(t) \quad (21.24)$$

und setzt (21.24) in (21.23) ein, dann folgt:

$$\tilde{Y}(t) = a\tilde{Y}(t-1) \quad (21.25)$$

Gleichung (21.24) nach $Y(t)$ aufgelöst ergibt:

$$Y(t) = \tilde{Y}(t) + \bar{Y}(t) \quad (21.26)$$

Setzen wir für $Y(0)=C$, dann bestimmt sich die Funktionslösung von $\tilde{Y}(t)$ nach (21.21) mit

$$\tilde{Y}(t) = C a^t \quad (21.27)$$

Mit (21.27) in (21.26) folgt:

$$\boxed{Y(t) = C a^t + \bar{Y}(t)} \quad (21.28)$$

Gleichung (21.28) liefert eine generelle Information über die Form der Funktionslösung einer inhomogenen Endgleichung ersten Grades, nämlich:

Satz 21.1: Die parametergenerelle Funktionslösung einer Endgleichung ersten Grades bestimmt sich aus der Summe der parametergenerellen Funktionslösung ihrer homogenen Form $C a^t$ und einer speziellen Lösung $\bar{Y}(t)$.

Ein offenes Problem bleibt lediglich, eine spezielle Funktionslösung zu finden. Es sei schon vorweggenommen, daß 'die Kunst', eine beliebige inhomogene Differenzengleichung zu lösen, darin besteht, eine sie befriedigende spezielle Lösung zu bestimmen. Da derartige spezielle Lösungen bereits von anderen gefunden wurden, ist es überflüssig, hier die Auffindung dieser Lösungen im einzelnen zu erörtern. Wir wollen uns lediglich exemplarisch mit dem Spezialfall einer inhomogenen Endgleichung

$$Y(t) = aY(t-1) + E \quad E=\text{konstant}$$

befassen. Die spezielle Lösung gewinnt man, wenn man von der gedanklichen Vorstellung ausgeht, das System befände sich auf einem Niveaugleichgewicht, d.h. es sei $Y(t)=Y(t-1)=\bar{Y}(t)$ und damit

$$\bar{Y}(t) = a\bar{Y}(t) + E$$

womit sich $\bar{Y}(t)$ nach

$$\bar{Y}(t) = \frac{E}{1-a}$$

ermittelt. Setzen wir die spezielle Lösung in Gleichung (21.28) ein, so erhalten wir

$$Y(t) = Ca^t + \frac{E}{1-a} \quad (21.29)$$

Es handelt sich um eine parametergenerelle Lösung, in der der Einfluß eines bestimmten Anfangswertes $Y(0)$ in (21.29) jedoch nicht explizit zum Ausdruck kommt. Da der Anfangswert $Y(0)$ der Lösung (21.29) jedoch durch die Wahl von C bestimmt wird, geschieht die Explikation von $Y(0)$ auf folgende Weise: In der 0-ten Periode muß die Beziehung

$$Y(0) = Ca^0 + E/(1-a)$$

gelten. Die Auflösung dieser Gleichung liefert

$$C = Y(0) - [E/(1-a)] \quad (21.30)$$

Die Einsetzung von (21.30) in (21.29) liefert die parametergenerelle Lösung

$$Y(t) = [Y(0)-E/(1-a)]a^t + E/(1-a)$$

Lediglich im Falle $a=1$ versagt das geschilderte Verfahren. In diesem Fall bestimmt sich die spezielle Lösung aus:

$$\bar{Y}(t) = Et$$

Die Zurückführung des Parameters C in Gleichung (21.28) auf die Koeffizienten und Anfangswerte ergibt

$$Y(0) = Ca^0 + E*0$$

d.h.

$$C = Y(0)$$

und damit die Lösung

$$Y(t) = Y(0) + E*t$$

Zusammenfassend gilt: Satz 21.2: Die Funktionslösung der inhomogenen Endgleichung $Y(t)=aY(t-1)+E$, bestimmt sich nach

$$Y(t) = \begin{cases} [Y(0)-E/(1-a)]a^t + E/(1-a) & \text{für } a \neq 1 \\ Y(0) + Et & \text{für } a=1 \end{cases}$$

Als Beispiel sei die inhomogene Gleichung

$$Y(t) = 0,5Y(t-1) + 100 \quad \text{mit } Y(0)=50$$

angeführt. Entsprechend Satz 21.2 ergibt sich die Funktionslösung:

$$Y(t) = -150*0,5^t + 200$$

t	Regressionslösung				Funktionslösung			
	Y(t)	0,5Y(t-1)	100	Y(t-1)	Y(t)	-150*0,5 ^t	0,5 ^t	200
0	50	-	-	-	50	-150	1	200
1	125	25	100	50	125	-75	0,5	200
2	162,5	62,5	100	125	162,5	-32,5	0,25	200
3	181,25	81,25	100	162,5	181,25	-18,75	0,125	200
4	190,625	90,625	100	181,25	190,625	-9,375	0,0675	200

Tab. 21.1 Regressions- und Funktionslösung einer inhomogenen Differenzengleichung ersten Grades

Die Übereinstimmung der Funktionslösung mit der entsprechenden Regressionslösung bis zur vierten Periode zeigt die Tabelle 21.1.

ab) Funktionslösung von Endgleichungen zweiten Grades

a) Funktionslösung homogener Endgleichungen zweiten Gradesaa) Funktionslösung homogener Endgleichungen zweiten Grades mit ungleichen Wurzeln

Wählen wir in (21.3) $E(t)=0$ ($t=0,1,\dots$) und $n=2$, so erhalten wir die homogene Endgleichung zweiten Grades

$$Y(t) + a_1 Y(t-1) + a_2 Y(t-2) = 0 \quad (21.31)$$

Von dieser Gleichungsform soll die parametergenerelle Funktionslösung gefunden werden. Die Aufgabe besteht also darin, einen Ausdruck zu finden, der die Endgleichung (21.31) identisch Null macht.

Dieser Ausdruck kann, wie es sich gezeigt hat, nach folgendem Verfahren gefunden werden:

Definieren wir $a_1=a$ sowie $a_2=b$, dann folgt

$$Y(t) + aY(t-1) + bY(t-2) = 0 \quad (21.32)$$

In einer ersten Einschränkung unterstellen wir, daß der Formelausdruck

$$Y(t) = \lambda^t \quad (21.33)$$

eine Funktionslösung von (21.32) sei. Die Einsetzung von (21.33) in (21.32) ergibt

$$\lambda^t + a\lambda^{t-1} + b\lambda^{t-2} = 0 \quad (21.34)$$

Es zeigt sich, daß nicht jedes beliebige λ die Gleichung (21.31) befriedigt, sondern nur die Werte, die auch Gleichung (21.34) befriedigen. Um diese λ -Werte zu ermitteln, dividieren wir Gleichung (21.34) durch λ^{t-2} und erhalten die sogenannte charakteristische Gleichung

$$\lambda^2 + a\lambda + b = 0 \quad (21.35)$$

Es lassen sich für λ zwei Werte, λ_1 und λ_2 , finden, die die charakteristische Gleichung (21.35) befriedigen und sich nach

$$\lambda_{1,2} = -\frac{a}{2} \pm \sqrt{\frac{a^2}{4} - b} \quad (21.36)$$

ermitteln. λ_1 und λ_2 werden auch als die Wurzeln der charakteristi-

schen Gleichung bezeichnet. Man erhält damit (im Falle ungleicher Wurzeln) zwei Funktionslösungen:

$$Y(t) = \lambda_1^t \quad (21.37)$$

und

$$Y(t) = \lambda_2^t \quad (21.38)$$

Demonstrieren wir das Auffinden der zwei Funktionslösungen an einem Beispiel: Mit $a=1,5$ und $b=-1$ erhalten wir die homogene Endgleichung

$$Y(t) + 1,5Y(t-1) - Y(t-2) = 0 \quad (21.39)$$

Anhand von (21.36) bestimmen sich die Wurzeln $\lambda_1=0,5$ und $\lambda_2=-2$. Setzen wir die Lösung $Y(t)=0,5^t$ in Gleichung (21.39) ein, so zeigt sich mit

$$\begin{aligned} 0,5^t + 1,5*0,5^{t-1} - 0,5^{t-2} &= 0 \\ (0,5^2 + 1,5*0,5-1)0,5^{t-2} &= 0 \\ (0)0,5^{t-2} &= 0 \end{aligned}$$

daß $Y(t)=0,5^t$ eine Lösung von Gleichung (21.39) darstellt. Dasselbe gilt für $\lambda_2=-2$. Ermittelt man eine Regressionslösung von Gleichung (21.39), so erkennt man, daß der Zeitverlauf durch die Festlegung von $Y(0)$ und $Y(1)$ eindeutig bestimmt ist. Es liegt nahe, nach den Anfangswerten der Lösungen $Y(t)=0,5^t$ und $Y(t)=-2^t$ zu fragen. Für die erste ist $Y(0)=1$, $Y(1)=0,5$, für die zweite $Y(0)=1$, $Y(1)=-2$.

Nunmehr wird deutlich, daß die zwei Funktionslösungen spezielle Funktionslösungen repräsentieren, d.h. nur bezüglich bestimmter Anfangswerte gelten.

Da wir jedoch eine anfangswertgenerelle Funktionslösung von Gleichung (21.32) gewinnen wollen, ist es notwendig, eine Funktionslösung zu finden, in der die beiden Anfangswerte in allgemeiner Form, d.h. durch Buchstabensymbole ausgedrückt werden.

Es gilt: Satz 21.3: Sind λ_1^t und λ_2^t zwei spezielle Funktionslösungen der Endgleichung $Y(t)+aY(t-1)+bY(t-2)=0$, so ist die Linearkombination

$$c_1 \lambda_1^t + c_2 \lambda_2^t$$

ihre anfangswertgenerelle Funktionslösung.

Da die Gleichungen

$$\lambda_1^t + a\lambda_1^{t-1} + b\lambda_1^{t-2} = 0$$

und

$$\lambda_2^t + a\lambda_2^{t-1} + b\lambda_2^{t-2} = 0$$

die Endgleichung befriedigen, gilt dies auch für die Multiplikation aller Glieder der Gleichungen mit beliebigen Konstanten c_1 und c_2 , d.h.

$$c_1\lambda_1^t + a c_1\lambda_1^{t-1} + b c_1\lambda_1^{t-2} = 0$$

$$c_2\lambda_2^t + a c_2\lambda_2^{t-1} + b c_2\lambda_2^{t-2} = 0$$

Die Addition beider Gleichungen liefert den Ausdruck

$$(c_1\lambda_1^t + c_2\lambda_2^t) + a(c_1\lambda_1^{t-1} + c_2\lambda_2^{t-1}) + b(c_1\lambda_1^{t-2} + c_2\lambda_2^{t-2}) = 0 \quad (21.40)$$

Definiert man

$$Y(t) = c_1\lambda_1^t + c_2\lambda_2^t \quad (21.41)$$

so erkennt man, daß die Einsetzung dieses Ausdrucks in (21.40) zur Endgleichung (21.32) führt, d.h. Gleichung (21.41) stellt ebenfalls eine Funktionslösung von (21.32) dar. Es fragt sich jedoch, ob (21.41) auch die anfangswertgenerelle Funktionslösung repräsentiert. Unterstellen wir, daß dies der Fall sei, so erfolgt die numerische Konkretisierung von c_1 und c_2 anhand der Anfangswerte $Y(0)$ und $Y(1)$ aufgrund der Beziehung

$$\begin{aligned} Y(0) &= c_1\lambda_1^0 + c_2\lambda_2^0 \\ Y(1) &= c_1\lambda_1^1 + c_2\lambda_2^1 \end{aligned} \quad (21.42)$$

Man erkennt, daß die geeignete Wahl von c_1 und c_2 alle möglichen Anfangswertkombinationen $Y(0)$, $Y(1)$ zum Ausdruck bringt, Gleichung (21.41) also tatsächlich die anfangswertgenerelle Funktionslösung bildet. Sind die Anfangswerte $Y(0)$ und $Y(1)$ im Einzelfall vorgegeben, so gelangt man durch die Auflösung des Gleichungssystems (21.42) zur Konkretisierung von c_1 und c_2

$$c_1 = \frac{\lambda_2 Y(0) - Y(1)}{\lambda_2 - \lambda_1} \quad (21.43)$$

$$c_2 = \frac{\lambda_1 Y(0) - Y(1)}{\lambda_1 - \lambda_2} \quad (21.44)$$

Die parametergenerelle Funktionslösung einer homogenen Endgleichung zweiten Grades mit ungleichen Wurzeln bestimmt sich demnach durch

$$Y(t) = \frac{\lambda_2 Y(0) - Y(1)}{\lambda_2 - \lambda_1} \lambda_1^t + \frac{\lambda_1 Y(0) - Y(1)}{\lambda_1 - \lambda_2} \lambda_2^t \quad (21.45)$$

Bisher wurde (ohne Begründung) einschränkend angenommen, daß das beschriebene Verfahren von ungleichen Wurzeln ausgeht. Es liegt damit die Frage nahe, ob diese Einschränkung nicht aufhebbar ist. Dies ist jedoch nicht möglich. Vielmehr erweist sich diese Einschränkung nunmehr auch als zwingend notwendig, um das Gleichungssystem (21.42) lösbar zu machen. Denn stets nur im Falle $\lambda_1 = \lambda_2$ ist (21.42) nicht auflösbar.

aβ) Funktionslösung homogener Endgleichungen zweiten Grades mit gleichen Wurzeln

In diesem Fall ist als parametergenerelle Funktionslösung ein anderer Ansatz zu wählen. Überlegungen haben gezeigt, daß der Ansatz

$$Y(t) = [C_1 + t C_2] \lambda^t$$

die parametergenerelle Funktionslösung liefert. Die Präzisierung der Koeffizienten C_1 und C_2 anhand der Anfangswerte ergibt somit

$$Y(0) = [C_1 + 0 * C_2] \lambda^0$$

$$Y(1) = [C_1 + 1 * C_2] \lambda^1$$

Die Auflösung nach C_1 und C_2 liefert

$$C_1 = Y(0)$$

$$C_2 = Y(1)/\lambda - Y(0)$$

Die parametergenerelle Funktionslösung einer homogenen Endgleichung zweiten Grades mit gleichen Wurzeln ergibt sich damit aus

$$Y(t) = [Y(0) + t((Y(1)/\lambda) - Y(0))] \lambda^t \quad (21.46)$$

αγ) Numerische Beispiele von Funktionslösungen homogener Endgleichungen zweiten Grades

Im folgenden sollen drei Beispiele zur Ermittlung der Funktionslösung einer homogenen Differenzengleichung zweiten Grades beschrieben werden. Sie sind so gewählt, daß folgende Fälle zur Anwendung kommen:

Fall 1: Wurzeln sind reell und verschieden, d.h. $a^2/4 > b$

Fall 2: Wurzeln sind konjugiert komplex, d.h. $a^2/4 < b$

Fall 3: Wurzeln sind reell und gleich, d.h. $a^2/4 = b$

Der Begriff konjugiert komplexer Wurzeln wird später erörtert. Es genügt, vorläufig zu wissen, daß sich in diesem Fall für λ_1 und λ_2 zwei Ausdrücke finden lassen, die man als konjugiert komplexe Wurzeln bezeichnet.

(1) Beispiel zum Fall gleicher Wurzeln

Es ist die Funktionslösung der Gleichung

$$Y(t) + 4Y(t-1) + 4Y(t-2) = 0 \quad (21.47)$$

mit $Y(0)=50$ und $Y(1)=100$ zu bestimmen. Die Wurzeln der charakteristischen Gleichungen werden durch

$$\lambda = \lambda_1 = \lambda_2 = -2 \pm \sqrt{4-4} = -2$$

ermittelt. Mit der parametergenerellen Funktionslösung (21.46) ergibt sich:

$$Y(t) = [50+t[(100/-2)-50]][-2]^t$$

und damit

$$Y(t) = [50-100t][-2]^t$$

Tabelle 21.2. zeigt den Zeitverlauf der Variablen Y für 5 Perioden in Form der Funktions- und Regressionslösung

Funktionslösung				Regressionslösung			
t	$50-100t$	-2^t	$Y(t) = [50-100t](-2)^t$	$Y(t) = -4Y(t-1)-4Y(t-2)$	$-4Y(t-1)$	$-4Y(t-2)$	
0	50	1	50	50	-	-	
1	-50	-2	100	100	-	-	
2	-150	+4	-600	-600	-400	-200	
3	-250	-8	2000	2000	2400	-400	
4	-350	+16	-5600	-5600	-8000	2400	

Tab. 21.2 Funktions- und Regressionslösung einer homogenen, linearen Differenzengleichung zweiten Grades mit gleichen Wurzeln

(2) Beispiel zum Fall reeller und verschiedener Wurzeln

Wir betrachten die Gleichung

$$Y(t) + 1,5Y(t-1) - Y(t-2) = 0 \quad \text{mit } Y(0)=50 \text{ und } Y(1)=100$$

Ihre Wurzeln bestimmen sich durch

$$\lambda_1 = -0,75 + \sqrt{\frac{1,5^2}{4} + 1} = +0,5$$

$$\lambda_2 = -0,75 - \sqrt{\frac{1,5^2}{4} + 1} = -2,0$$

	Funktionslösung			Regressionslösung		
t	$80(0,5)^t$	$-30(-2)^t$	$Y(t)$	$Y(t)$	$-1,5Y(t-1)$	$Y(t-2)$
0	80	-30	50	50	-	-
1	40	60	100	100	-	-
2	20	-120	-100	-100	-150	+50
3	10	240	250	250	150	100
4	5	-480	-475	-475	-375	-100

Tab. 21.3 Funktions- und Regressionslösung einer homogenen, linearen Differenzengleichung zweiten Grades mit ungleichen Wurzeln

Da λ_1 ungleich λ_2 ist, und λ_1 und λ_2 reell sind, gilt die parametergenerelle Funktionslösung (21.45), d.h.

$$Y(t) = \frac{-2*50-100}{-2-0,5}(0,5)^t + \frac{0,5*50-100}{0,5+2}(-2)^t$$

und damit

$$Y(t) = 80(0,5)^t - 30(-2)^t$$

Aus Tabelle 21.3 ergibt sich der identische Zeitverlauf im Falle einer Funktions- und Regressionslösung.

(3) Beispiel zum Fall konjugiert komplexer Wurzeln

Als Ausgangspunkt wählen wir die Endgleichung

$$Y(t) - 4Y(t-1) + 13Y(t-2) = 0 \quad \text{mit } Y(0)=50 \text{ und } Y(1)=100 \quad (21.48)$$

Die Wurzeln dieser Endgleichung bestimmen sich mit

$$\lambda_1 = 2 + \sqrt{4-13} = 2 + 3\sqrt{-1}$$

$$\lambda_2 = 2 - \sqrt{4-13} = 2 - 3\sqrt{-1}$$

Bezeichnen wir zur Abkürzung $\sqrt{-1}=i$, so folgt

$$\lambda_1 = 2 + 3i \quad \text{und} \quad \lambda_2 = 2 - 3i$$

Man spricht in diesem Fall von konjugiert komplexen Wurzeln. Entsprechend der parametergenerellen Funktionslösung (21.45) ergibt sich durch Einsetzen die spezielle Funktionslösung der Endgleichung (21.48)

$$\text{oder } Y(t) = \frac{(2-3i)50-100}{(2-3i)-(2+3i)}(2+3i)^t + \frac{(2+3i)50-100}{(2+3i)-(2-3i)}(2-3i)^t$$

$$Y(t) = \frac{(-2+3i)50+100}{6i}(2+3i)^t + \frac{(2+3i)50-100}{6i}(2-3i)^t$$

und damit

$$Y(t) = 25(2+3i)^t + 25(2-3i)^t \quad (21.49)$$

Zur Berechnung von $Y(t)$ ist es notwendig, eine Potenzierung von i durchzuführen.

Mit $i^1 = \sqrt{-1}$ wird $i^2 = -1$, $i^3 = -i$, $i^4 = 1$, $i^5 = i$. Anhand dieser Beziehungen lässt sich die Funktionslösung in Tabelle 21.4 ermitteln.

t	$(2+3i)^t$	$25(2+3i)^t$	$(2-3i)^t$	$25(2-3i)^t$	$Y(t) = 25(2+3i)+25(2-3i)$
0	1	25	1	25	+50
1	$2+3i$	$50+75i$	$2-3i$	$50-75i$	100
2	$-5+12i$	$-125+300i$	$-5-12i$	$-125-300i$	-250
3	$-46+9i$	$-1150+225i$	$-46-9i$	$-1150-225i$	-2300
4	$-119-120i$	$-2975-3000i$	$-119-120i$	$-2975+3000i$	-5950
5	$122-597i$	$3050-14925i$	$122+597i$	$3050+14925i$	6100

Tab. 21.4 Funktionslösung einer homogenen, linearen Differenzengleichung zweiten Grades mit konjugiert komplexen Wurzeln

Die entsprechende Regressionslösung zeigt Tabelle 21,5

t	$4Y(t-1)$	$-13Y(t-2)$	$Y(t) = 4Y(t-1)-13Y(t-2)$
0	-	-	50
1	-	-	100
2	400	-650	-250
3	-1000	-1300	-2300
4	-9200	+3250	-5950
5	-23800	+29900	+6100

Tabelle 21.5 Regressionslösung einer homogenen, linearen Differenzengleichung zweiten Grades mit konjugiert komplexen Wurzeln

Man erkennt, daß die Funktions- und Regressionslösungen miteinander übereinstimmen. Angesichts der beiden Lösungen liegt die Frage nahe, welche Vorteile in diesem Fall eine Funktionslösung gegenüber ihrer entsprechenden Regressionslösung aufweist. Denn die Funktionslösung (21.49) ermöglicht wegen des Auftretens von i keine überschaubare Beurteilung des Zeitverlaufes von $Y(t)$.

Im Falle von Endgleichungen, in deren Wurzeln der Ausdruck i auftritt, wird diese Überschaubarkeit wiederhergestellt, wenn man die Funktionslösung in ihre sogenannte trigonometrische Form überführt. Es läßt

sich vorausgreifend beispielsweise zeigen, daß die Funktionslösung unseres Beispiels, d.h.

$$Y(t) = 25(2+3i)^t + 25(2-3i)^t$$

auch durch die Funktion

$$Y(t) = 3,60555^t [50\cos 56^\circ 18' t]$$

beschrieben werden kann. Diese Darstellungsform erlaubt eine bessere Beurteilung des Systemverhaltens. Daher wollen wir uns im nächsten Abschnitt mit dem Problem der Bestimmung der trigonometrischen Formen von Funktionslösungen befassen.

αδ) Trigonometrische Form der Funktionslösung homogener Endgleichungen zweiten Grades mit konjugiert komplexen Wurzeln

Stellt man bei der Analyse einer Endgleichung zweiten Grades fest, daß $b > a^2/4$, dann erhält man als Wurzeln der charakteristischen Gleichung Ausdrücke wie zum Beispiel

$$\lambda_1 = 2 + 3i$$

$$\lambda_2 = 2 - 3i$$

Derartige Ausdrücke bezeichnet man als komplexe Zahlen. Komplexe Zahlen setzen sich aus zwei Komponenten zusammen: einem Realteil, d.h. einer reellen Zahl und einem imaginären Teil, d.h. einer mit i multiplizierten reellen Zahl. Zwei komplexe Zahlen werden als konjugiert komplex bezeichnet, wenn sie sich nur im Vorzeichen ihres Imaginärteiles unterscheiden. Ist eine komplexe Zahl die Wurzel einer charakteristischen Gleichung, so ist ihr konjugiert komplexes Gegenstück stets auch eine Wurzel dieser charakteristischen Gleichung. Komplexe Zahlen lassen sich geometrisch auf der sogenannten Gaußschen Zahlenebene darstellen. Diese Gaußsche Zahlenebene wird durch ein rechtwinkliges Koordinatensystem beschrieben, dessen Abszissenwerte die Realteile einer komplexen Zahl beschreiben, während die Ordinatenwerte die Achse der imaginären Zahlen bilden. Jeder komplexen Zahl mit einem bestimmten Real- und Imaginärteil entspricht daher ein bestimmter Punkt in der Gaußschen Zahlenebene.

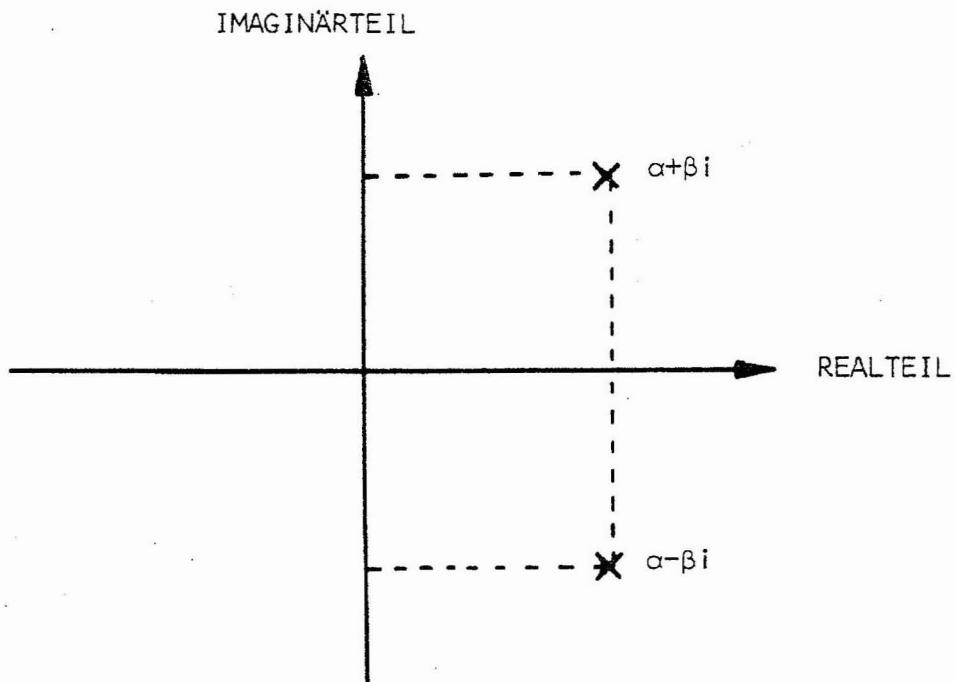


Abb. 21.2 Darstellung eines Paars konjugiert komplexer Wurzeln in der Gaußschen Zahlenebene

Bezeichnen wir eine konjugiert komplexe Wurzel mit $\lambda_{1,2} = \alpha \pm \beta i$, so ist der geometrische Ort beider Wurzeln aus Abbildung 21.2 zu erkennen. Komplexe Zahlen können in der Gaußschen Zahlenebene auch durch andere Maßsysteme als das bisher beschriebene kartesische Koordinatensystem gekennzeichnet werden. Als ein weiteres System zur Kennzeichnung der Lage komplexer Zahlen bietet sich das Polarkoordinatensystem an. Eine komplexe Zahl kann in einem Polarkoordinatensystem erschöpfend durch die Länge eines vom Nullpunkt ausgehenden Fahrstrahles und dem Winkel φ dieses Fahrstrahles mit der positiven Halbachse dargestellt werden.

Abbildung 21.3 zeigt eine derartige Polarkoordinatendarstellung einer komplexen Zahl. Der Winkel des Fahrstrahles wird auch Abweichung genannt.

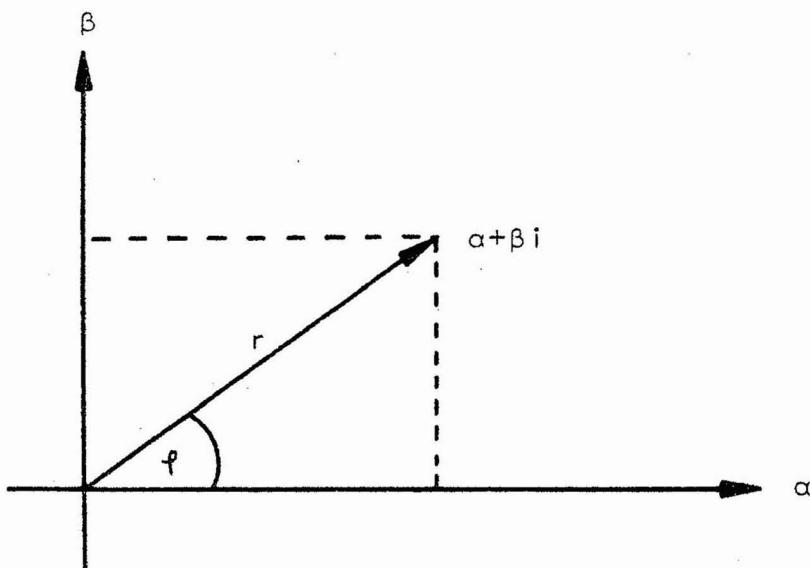


Abb. 21.3 Kennzeichnung einer komplexen Zahl in einem Polarkoordinatensystem

Die Größe r bezeichnet man als den Modul oder Absolutbetrag der komplexen Zahl. Er berechnet sich nach dem Satz des Pythagoras aus:

$$r = \sqrt{\alpha^2 + \beta^2} \quad (21.50)$$

Weiterhin bestehen die Beziehungen

$$\sin \varphi = \frac{\beta}{r} \quad (21.51)$$

und

$$\cos \varphi = \frac{\alpha}{r} \quad (21.52)$$

Die konjugiert komplexen Wurzeln

$$\lambda_1 = \alpha + \beta i$$

und

$$\lambda_2 = \alpha - \beta i$$

werden mit (21.51) und (21.52)

$$\lambda_1 = r(\cos \varphi + i \sin \varphi)$$

$$\lambda_2 = r(\cos \varphi - i \sin \varphi)$$

Nach dem Satz von Moivre gilt für die Potenzen von λ_1 und λ_2 :

$$\begin{aligned}\lambda_1^t &= (\alpha + \beta i)^t = r^t (\cos \varphi t + i \sin \varphi t) \\ \lambda_2^t &= (\alpha - \beta i)^t = r^t (\cos \varphi t - i \sin \varphi t)\end{aligned} \quad \text{für } t=1,2,3,\dots \quad (21.53)$$

Setzen wir die Ausdrücke für λ_1^t und λ_2^t in die allgemeine Funktionslösung (21.41) ein, dann folgt

$$Y(t) = C_1(\alpha + \beta i)^t + C_2(\alpha - \beta i)^t$$

oder

$$Y(t) = r^t [C_1(\cos \varphi t + i \sin \varphi t) + C_2(\cos \varphi t - i \sin \varphi t)]$$

oder

$$Y(t) = r^t [(C_1 + C_2) \cos \varphi t + (C_1 - C_2) i \sin \varphi t] \quad (21.54)$$

Unter Verwendung der Definition von C_1 und C_2 in (21.43) und (21.44) folgt

$$C_1 + C_2 = \frac{\lambda_2 Y(0) - Y(1)}{\lambda_2 - \lambda_1} + \frac{\lambda_1 Y(0) - Y(1)}{\lambda_1 - \lambda_2}$$

und

$$C_1 + C_2 = Y(0) \quad (21.55)$$

Für $(C_1 - C_2)i$ in (21.54) gilt

$$(C_1 - C_2)i = \left(\frac{\lambda_2 Y(0) - Y(1)}{\lambda_2 - \lambda_1} - \frac{\lambda_1 Y(0) - Y(1)}{\lambda_1 - \lambda_2} \right) i$$

und mit $\lambda_{1,2} = \alpha \pm \beta i$ folgt

$$(C_1 - C_2)i = [Y(1) - \alpha Y(0)]/\beta \quad (21.56)$$

Beide Koeffizienten vor den trigonometrischen Funktionen sind daher reell. Mit (21.55) und (21.56) in (21.54) erhalten wir die Funktionslösung

$$Y(t) = r^t \{Y(0) \cos \varphi t + [Y(1) - \alpha Y(0)]/\beta \sin \varphi t\} \quad (21.57)$$

Definiert man

$$A = Y(0) \quad \text{und} \quad B = \frac{\alpha Y(0) + 2Y(1)}{2 \sqrt{\left| \frac{a^2}{4} - b \right|}}$$

und berücksichtigt man, daß

$$\alpha = -\frac{a}{2} \quad \text{und} \quad \beta = \sqrt{\left| \frac{a^2}{4} - b \right|}$$

gilt, dann folgt

$$Y(t) = r^t (A \cos \varphi t + B \sin \varphi t)$$

Der Winkel φ berechnet sich wegen (21.52) mit

$$\varphi = \arccos\left(\frac{a}{r}\right) = \arccos\left(\frac{-a}{2r}\right)$$

Zur Erhöhung der Übersichtlichkeit wäre es wünschenswert, eine Funktionslösung von $Y(t)$ zu gewinnen, die im Gegensatz zu (21.57) nur von einer trigonometrischen Funktion abhängig ist. Das ist auf folgende Weise möglich. Setzen wir

$$\begin{aligned} C_1 + C_2 &= D \cos \omega \\ C_1 - C_2 &= D \sin \omega \end{aligned} \quad (21.58)$$

so erhalten wir mit (21.54)

$$Y(t) = r^t (D \cos \omega \cos \varphi t + D \sin \omega \sin \varphi t)$$

oder

$$Y(t) = r^t D (\cos \omega \cos \varphi t + \sin \omega \sin \varphi t) \quad (21.59)$$

Unter Verwendung des elementaren trigonometrischen Zusammenhangs

$$\cos \omega \cos \varphi t + \sin \omega \sin \varphi t = \cos(\varphi t - \omega)$$

folgt aus (21.59)

$$Y(t) = r^t D \cos(\varphi t - \omega) \quad (21.60)$$

Die Parameter D und ω in (21.60) bestimmen sich aus (21.58) mit:

$$D = \sqrt{2(C_1^2 + C_2^2)} \quad (21.61)$$

und

$$\tan \omega = \frac{\sin \omega}{\cos \omega} = \frac{C_1 - C_2}{C_1 + C_2} \quad (21.62)$$

Man erkennt, daß die Cosinusfunktion dieselben Werte für $t=0, 2\pi, 4\pi$ usw. annimmt. Da sich die Funktion alle $2\pi/\varphi$ Zeiteinheiten wiederholt, stellt sie eine periodische Funktion mit einer Zykluslänge von $2\pi/\varphi$ Zeiteinheiten dar. Als Beispiel ziehen wir die auf Seite 185 angeführte Endgleichung

$$Y(t) - 4Y(t-1) + 13Y(t-2) = 0$$

jedoch mit den Anfangswerten $Y(0)=20$ und $Y(1)=50$ heran. Sie besitzt, wie bereits festgestellt, die konjugiert komplexen Wurzeln

$$\lambda_{1,2} = 2 \pm 3i$$

Der Modul r berechnet sich nach (21.50) aus

$$r = \sqrt{2^2 + 3^2} = 3,60555$$

Nach (21.51) wird

$$\sin\varphi = \frac{3}{3,6} = 0,832$$

Anhand einer trigonometrischen Tabelle bestimmen wir den Winkel φ mit $\varphi \approx 56^\circ 18'$.

Mit (21.55) und (21.56) werden

$$c_1 + c_2 = 20$$

und

$$(c_1 - c_2)i = (50 - 2 \cdot 20)/3 = \frac{10}{3}$$

Unter Zugrundelegung von Gleichung (21.54) ergibt sich

$$Y(t) = 3,60555^t [20 \cos 56^\circ 18' t + \frac{10}{3} \sin 56^\circ 18' t]$$

Zur Ermittlung der nur von einer trigonometrischen Funktion abhängigen Lösung der Endgleichung sind ω und A zu ermitteln. Es ist gemäß (21.62)

$$\tan\omega = \frac{10}{20 \cdot 3} = \frac{1}{6}$$

Aus einer trigonometrischen Tabelle ermitteln wir: $\omega \approx 9^\circ 28'$.

Nach (21.61) wird

$$D = 2 \sqrt{2 \left[\left(\frac{35}{3}\right)^2 + \left(\frac{25}{3}\right)^2 \right]}$$

d.h.

$$D = 20,276$$

Die Funktionslösung lautet mit (21.60)

$$Y(t) = 3,60555^t * 20,276 \cos(56^\circ 18' t - 9^\circ 28')$$

Zum besseren Verständnis des dynamischen Verhaltens eines Zeitverlaufes, welcher durch die Gleichungen (21.57) oder (21.60) beschrieben wird, ist es sinnvoll, die Sinus- und Cosinuskomponenten im Hinblick auf ihren Beitrag zum Zeitverlauf etwas eingehender zu untersuchen.

Die Cosinus- und Sinusfunktion sind Funktionen, deren Ordinatenwerte in Abhängigkeit von dem Winkel φ bestimmt werden.

Man erkennt, daß es sich um periodische Funktionen mit einer Periode von 360° Grad handelt, die sich zwischen 1 und -1 bewegen.

In Gleichung (21.54) treten die Ausdrücke $\cos\varphi t$ und $\sin\varphi t$ auf. Der Ausdruck φt charakterisiert ein bestimmtes Winkelmaß. Ist φ beispielsweise 60° , so beschreibt $\cos 60t$ mit $t=0,1,2,\dots$ die Ordinatenwerte der Sinusfunktion bei den Abszissenwerten $0^\circ, 60^\circ, 120^\circ$ usw.. Grafisch gesehen wird aus der Cosinusfunktion in gleichbleibenden Abszissenabständen ein Ordinatenwert entnommen. In Abbildung 21.4 sind diese Ordinatenwerte durch Punkte gekennzeichnet.

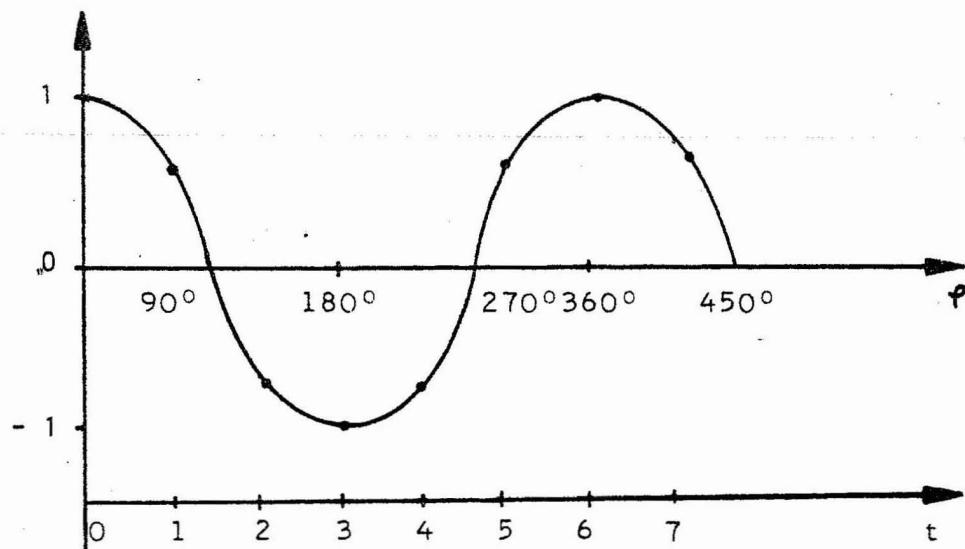


Abb. 21.4 Ordinatenwerte einer Cosinusfunktion in den Punkten $\varphi=60t$ für $t=0,1,2,\dots$

Analoges gilt für die Sinusfunktion. Die Ausdrücke $\cos\vartheta t$ und $\sin\vartheta t$ besitzen daher einen zwischen +1 und -1 fluktuierenden Verlauf. Ihre Multiplikation mit den Konstanten $Y(0)$ bzw. $[Y(1)-\alpha Y(0)]/\beta$ in Gleichung (21.57) vergrößert oder verkleinert die Amplitude ihres wellenförmigen Verlaufes. Der Ausdruck in der geschweiften Klammer der Gleichung (21.57) repräsentiert somit eine Überlagerung von zwei Schwingungen mit begrenztem Schwingungsausschlag. Entscheidend für die zeitliche Entwicklung erweist sich damit der Modul r . Ist $r>1$, so explodiert das System, während es im Falle $r<1$ gegen Null konvergiert.

β) Funktionslösung inhomogener Endgleichungen zweiten Grades

Eine inhomogene Endgleichung zweiten Grades, d.h.

$$Y(t) + aY(t-1) + bY(t-2) = E(t) \quad (21.63)$$

besitzt eine Funktionslösung, die sich wie im Falle inhomogener Endgleichungen ersten Grades aus der Summe der speziellen Lösung von Gleichung (21.63) und der Funktionslösung der entsprechenden homogenen Endgleichung von (21.63) zusammensetzt.

Der Beweis ergibt sich analog zum Fall einer Endgleichung ersten Grades und wird daher kurz gefaßt.

Ist $\bar{Y}(t)$ eine spezielle Lösung von (21.63), so folgt aus

$$Y(t) + aY(t-1) + bY(t-2) = E(t)$$

$$\bar{Y}(t) + a\bar{Y}(t-1) + b\bar{Y}(t-2) = E(t)$$

durch Subtraktion der unteren von der oberen Gleichung

$$Y(t) - \bar{Y}(t) + a[Y(t-1) - \bar{Y}(t-1)] + b[Y(t-2) - \bar{Y}(t-2)] = 0 \quad (21.64)$$

Definiert man

$$\tilde{Y}(t) = Y(t) - \bar{Y}(t) \quad (21.65)$$

so folgt mit (21.64)

$$\tilde{Y}(t) + a\tilde{Y}(t-1) + b\tilde{Y}(t-2) = 0 \quad (21.66)$$

Die allgemeine Lösung von $\tilde{Y}(t)$ ist im Falle ungleicher Wurzeln

$$\tilde{Y}(t) = C_1 \lambda_1^t + C_2 \lambda_2^t \quad (21.67)$$

Aus Gleichung (21.65) folgt daher

$$Y(t) = C_1 \lambda_1^t + C_2 \lambda_2^t + \bar{Y}(t) \quad (21.68)$$

Im Falle gleicher Wurzeln ergibt sich unter Berücksichtigung der in diesem Falle zu verwendenden Funktionslösung der homogenen Gleichung:

$$Y(t) = [C_1 + t C_2] \lambda^t + \bar{Y}(t) \quad (21.69)$$

Im Falle ungleicher reeller Wurzeln werden die Konstanten C_1 und C_2 durch:

$$Y(0) = C_1 + C_2 + \bar{Y}(t)$$

$$Y(1) = C_1 \lambda_1 + C_2 \lambda_2 + \bar{Y}(t)$$

auf die Anfangswerte und die spezielle Lösung zurückgeführt und bestimmen sich mit

$$C_1 = \frac{[Y(0) - \bar{Y}(t)] \lambda_2 - Y(1) + \bar{Y}(t)}{\lambda_2 - \lambda_1}$$

und

$$C_2 = \frac{-[Y(0) - \bar{Y}(t)] \lambda_1 + Y(1) - \bar{Y}(t)}{\lambda_2 - \lambda_1} \quad (21.70)$$

Im Falle gleicher Wurzeln bestimmen sich C_1 und C_2 aus dem Gleichungssystem:

$$Y(0) = C_1 + \bar{Y}(t)$$

$$Y(1) = (C_1 + C_2) \lambda + \bar{Y}(t)$$

d.h.

$$C_1 = Y(0) - \bar{Y}(t)$$

und

$$C_2 = \frac{Y(1) - \bar{Y}(t)}{\lambda} - Y(0) + \bar{Y}(t) \quad (21.71)$$

Die spezielle Lösung $\bar{Y}(t)$ soll nur für den Fall $E(t) = E = \text{konst.}$ betrachtet werden. Sind die Wurzeln ungleich, so ist als spezielle Lösung der Ansatz

$$\bar{Y}(t) = \frac{E}{1+a+b} \quad (21.72)$$

zu wählen. Bei gleichen Wurzeln dagegen ist auf den Ansatz

$$\bar{Y}(t) = \frac{4E}{(a+2)^2} \quad (21.73)$$

zurückzugreifen.

Ein Beispiel soll das geschilderte Verfahren abschließen. Zu ermitteln ist die Funktionslösung der inhomogenen Endgleichung

$$Y(t) + 1,5Y(t-1) - Y(t-2) = 120$$

mit den Anfangswerten $Y(0)=30$ und $Y(1)=80$. Die Wurzeln ihrer charakteristischen Gleichung bestimmen sich nach (21.36) mit $\lambda_1=0,5$ und $\lambda_2=-2,0$. Da die Wurzeln ungleich sind, wird die spezielle Lösung entsprechend Gleichung (21.72) bestimmt, d.h.

$$\bar{Y}(t) = \frac{120}{1+1,5-1} = \frac{120}{1,5} = 80$$

Die Funktionslösung bestimmt sich nach Gleichung (21.68) durch

$$Y(t) = C_1(0,5)^t + C_2(-2)^t + 80$$

Unter weiterer Spezifizierung von C_1 und C_2 gemäß (21.70) folgt

$$Y(t) = -40(0,5)^t - 10(-2)^t + 80$$

t	Funktionslösung					Regressionslösung				
	$0,5^t$	$(-2)^t$	$-40(0,5)^t$	$-10(-2)^t$	80	$y(t)$	$y(t)$	$-1,5y(t-1)$	$y(t-2)$	120
0	1	1	-40	-10	80	30	30	-	-	-
1	0,5	-2	-20	20	80	80	80	-	-	-
2	0,25	4	-10	-40	80	30	30	-120	30	120
3	0,125	-8	-5	80	80	155	155	-45	80	120
4	0,0625	16	-2,5	-160	80	-82,5	-82,5	-232,5	30	120

Tab. 21.6 Funktions- und Regressionslösung einer inhomogenen, linearen Endgleichung zweiten Grades mit $E(t)=120$

Tabelle 21.7 liefert eine zusammenfassende Übersicht der verschiedenen parametergenerellen Funktionslösungen der homogenen Gleichung. Den Fall einer inhomogenen Endgleichung mit einer exogenen Variablen $E(t)=E$ ($E=\text{konst.}$) zeigt Tabelle 21.8

Bedingung	Wurzeln der charakt. Gleichung	Bestimmung der Konstanten	Parameter generelle Funktionslösung
$b < \frac{a^2}{4}$	$\lambda_1 = -\frac{a}{2} + \sqrt{\frac{a^2}{4} - b}$ $\lambda_2 = -\frac{a}{2} - \sqrt{\frac{a^2}{4} - b}$	$C_1 = \frac{\lambda_2 Y(0) - Y(1)}{\lambda_2 - \lambda_1}$ $C_2 = \frac{Y(1) - \lambda_1 Y(0)}{\lambda_2 - \lambda_1}$	$Y(t) = C_1 \lambda_1^t + C_2 \lambda_2^t$
$b > \frac{a^2}{4}$	$\lambda_1 = -\frac{a}{2} + i\sqrt{\frac{a^2}{4} - b}$ $\lambda_2 = -\frac{a}{2} - i\sqrt{\frac{a^2}{4} - b}$	$A = Y(0)$ $B = \frac{aY(0) + 2Y(1)}{2\sqrt{\frac{a^2}{4} - b}}$	$Y(t) = r^t [A \cos \varphi t + B \sin \varphi t]$ $\varphi = \arccos \left(-\frac{a}{2r} \right)$ $r = \sqrt{\frac{a^2}{4} + \frac{ a^2 - 4b }{4}}$
$b = \frac{a^2}{4}$	$\lambda_1 = \lambda_2 = \lambda = -\frac{a}{2}$	$C_1 = Y(0)$ $C_2 = \frac{Y(1) - Y(0)}{\lambda}$	$Y(t) = [C_1 + tC_2] \lambda^t$

Tab. 21.7 Parametergenerelle Funktionslösungen der homogenen Endgleichung zweiten Grades $Y(t) + aY(t-1) + bY(t-2) = 0$

Bedingung	Wurzeln der Endgleichung	Spezielle Lösung	Konstante C_1 und C_2	Parametergenerelle Funktionslösung
$b < \frac{a^2}{4}$	$\lambda_1 = -\frac{a}{2} + \sqrt{\frac{a^2}{4} - b}$ $\lambda_2 = -\frac{a}{2} - \sqrt{\frac{a^2}{4} - b}$	$\bar{Y}(t) = \frac{E}{1+a+b}$	$C_1 = \frac{(Y(0) - \bar{Y}(t)) \lambda_2 - Y(1) + \bar{Y}(t)}{\lambda_2 - \lambda_1}$ $C_2 = Y(0) - C_1 - \bar{Y}(t)$	$Y(t) = C_1 \lambda_1^t + C_2 \lambda_2^t + \bar{Y}(t)$
$b > \frac{a^2}{4}$	$\lambda_1 = -\frac{a}{2} + i\sqrt{\frac{a^2}{4} - b}$ $\lambda_2 = -\frac{a}{2} - i\sqrt{\frac{a^2}{4} - b}$	$\bar{Y}(t) = \frac{E}{1+a+b}$	$A = Y(0) - \bar{Y}(t)$ $B = \frac{a(\bar{Y}(t) - Y(0)) - 2Y(1) + 2\bar{Y}(t)}{-2\sqrt{\frac{a^2}{4} - b}}$	$Y(t) = r^t [A \cos \varphi t + B \sin \varphi t]$ $\varphi = \arccos \left(-\frac{a}{2r} \right)$ $r = \sqrt{\frac{a^2}{4} + \frac{ a^2 - 4b }{4}}$
$b = \frac{a^2}{4}$	$\lambda = -\frac{a}{2}$	$\bar{Y}(t) = \frac{4E}{(a+2)^2}$	$C_1 = Y(0) - \bar{Y}(t)$ $C_2 = \frac{Y(1) - \bar{Y}(t)}{\lambda} - Y(0) + \bar{Y}(t)$	$Y(t) = (C_1 + C_2) \lambda^t + \frac{4E}{(a+2)^2}$

Tab. 21.8 Parametergenerelle Funktionslösungen der inhomogenen Endgleichung zweiten Grades $Y(t) + aY(t-1) + bY(t-2) = E$

ac) Funktionslösung von Endgleichungen n-ten Grades

Das für Endgleichungen ersten und zweiten Grades besprochene Lösungsverfahren gilt im Prinzip auch für Endgleichungen beliebigen Grades:
Aus der homogenen Endgleichung n-ten Grades

$$Y(t) + a_1 Y(t-1) + \dots + a_n Y(t-n) = 0$$

gewinnt man durch Substitution von $Y(t-i) = \lambda^{n-i}$ die charakteristische Gleichung

$$\lambda^n + a_1 \lambda^{n-1} + \dots + a_{n-1} \lambda + a_n = 0$$

Nach dem Gaußschen Hauptsatz der Algebra besitzt ein derartiges Polynom n-ten Grades genau n Wurzeln $\lambda_1, \lambda_2, \dots, \lambda_n$. Der auszuwählende Lösungsansatz hängt von der Art der Wurzeln ab. Für die Teilmenge der reellen und verschiedenen Wurzeln ist der Lösungsansatz

$$c_1 \lambda_1^t + c_2 \lambda_2^t + \dots + c_{n-1} \lambda_{n-1}^t + c_n \lambda_n^t$$

aufzunehmen.

Im Falle einer s-fach auftretenden gleichen Wurzel ist der Ausdruck

$$(c_1 + c_2 t + c_3 t^2 + \dots + c_s t^{s-1}) \lambda^t$$

in die Lösung mit einzufügen.

Falls ein Paar konjugiert komplexer Wurzeln s-mal vorliegt, ist der Ausdruck

$$r^t [(A_1 + A_2 t + A_3 t^2 + \dots + A_s t^{s-1}) \cos \varphi t + (B_1 + B_2 t + B_3 t^2 + \dots + B_s t^{s-1}) \sin \varphi t]$$

in den Lösungsansatz einzufügen. In einer Endgleichung n-ten Grades treten demnach n Konstante auf, die unter Vorgabe der Anfangswerte $Y(0), Y(1), \dots, Y(n-1)$ bestimmt werden können.

Im Falle der inhomogenen Endgleichung

$$Y(t) + a_1 Y(t-1) + \dots + a_n Y(t-n) = E(t)$$

setzt sich die Funktionslösung stets aus einer speziellen Lösung $\bar{Y}(t)$ und der entsprechenden Funktionslösung der homogenen Endgleichung $\tilde{Y}(t)$ zusammen.

Besteht der die exogene Variable beschreibende Formelausdruck aus einer linearen Kombination der Terme a^t , $\sin ct$, $\cos ct$, t^p , dann ist

zur Auffindung der speziellen Lösung ein als 'Methode der unbestimmten Koeffizienten' bezeichnetes Verfahren anwendbar. [83, S.32f] Endgleichungen, deren Grad höher als drei ist, lassen sich allerdings nicht mehr in Form parametergenereller Funktionslösungen darstellen. Nach einem Theorem von GALOIS ist es nicht möglich, die Wurzeln von Polynomen vierten Grades und höher als Funktionen ihrer Koeffizienten auszudrücken. [162, S.92f]

Die Wurzeln der charakteristischen Gleichung sind in diesen Fällen nur durch Näherungen zu bestimmen. Es stehen dazu leistungsfähige EDV-Programme zur Verfügung. Die wohl aufwendigste Bestimmung der Wurzeln einer charakteristischen Gleichung wurde von HOWERY durchgeführt. Im Rahmen der Linearisierung eines dynamischen Modells der amerikanischen Wirtschaft untersuchte er eine charakteristische Gleichung 56sten Grades und bestimmte ihre 56 Wurzeln. [92, S.654]

b) Empirische Kennzeichen linearer Systeme

Mit den Verfahren zur Ermittlung von Funktionslösungen haben wir eine Grundlage geschaffen, um bestimmte typenspezifische Implikationen linearer Systeme zu erörtern.

Wie erwähnt, wird das Studium dynamischer Systeme oft unter Verwendung des Prinzips eines schwarzen Kastens vorgenommen. Als Eingangs- und Ausgangsgrößen dieses schwarzen Kastens können die unverzögerten exogenen und endogenen Variablen angesehen werden. Viele empirisch relevante Implikationen eines dynamischen Systems lassen sich dadurch ermitteln, daß man unter Festlegung bestimmter Eingänge (d.h. Verläufe der exogenen Variablen) die Ausgänge (d.h. die Verläufe der endogenen Variablen) betrachtet.

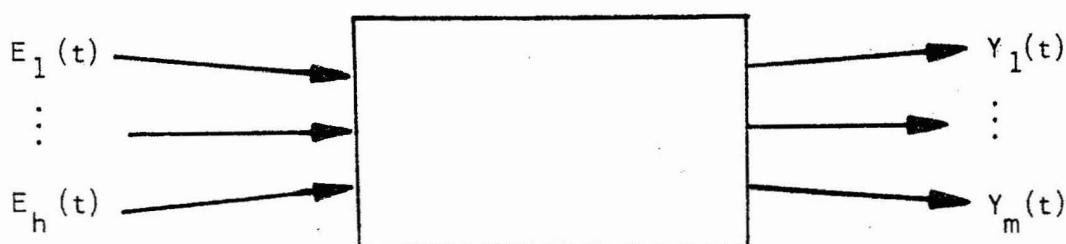


Abb. 21.5 Dynamisches System in der Deutungsweise eines schwarzen Kastens

In der Regelungstechnik werden derartige Ein- und Ausgangsuntersuchungen am System selbst durchgeführt. Wie erwähnt, ist dies in den wirtschafts- und sozialwissenschaftlichen Systemen aus praktischen Gründen nicht möglich. Man kann jedoch eine Schwarze-Kasten-Betrachtung am Modell durchführen, d.h. ein Modell eines Systems wird hinsichtlich der Beziehungen zwischen seinen Ein- und Ausgangsgrößen untersucht.

Die artspezifischen Kennzeichen des Transformationsmechanismus zwischen Ein- und Ausgängen linearer dynamischer Systeme lassen sich in Form zweier Postulate fassen, die als das Postulat der ungestörten Überlagerung und das Postulat der Adäquanz von Ursache und Wirkung bezeichnet werden können..

Das Postulat der ungestörten Überlagerung besagt, daß der durch eine bestimmte Eingangsgröße bewirkte Zeitverlauf einer Ausgangsgröße unabhängig von anderen Eingängen bestimmt werden kann. Der resultierende Ausgangsgrößenverlauf bestimmt sich dabei aus der Addition der isoliert ermittelten Ausgangsgrößenverläufe bezüglich einer Eingangsgröße. Der Zeitverlauf der endogenen Variablen $Y_v(t)$ kann daher in folgen Komponenten zerlegt werden

$$Y_v(t) = \Delta Y_v(t)[E_1] + \Delta Y_v(t)[E_2] + \dots + \Delta Y_v(t)[E_v]$$

Hierbei ist $\Delta Y_v(t)[E_j]$ die Komponente, die durch die exogene (Eingangs)-Variable E_j bewirkt wurde.⁷

Zur Verdeutlichung des Prinzips der ungestörten Überlagerung wird die Endgleichung eines MA-Systems

$$Y(t) = (\alpha + \beta)Y(t-1) - \beta Y(t-2) + I_a(t)$$

herangezogen.

Eine Endgleichung wird im Rahmen der Schwarze-Kasten-Betrachtung als Übergangsfunktion des schwarzen Kastens bezeichnet. Abbildung 21.6 zeigt die Deutung des Zusammenhangs zwischen den Variablen Y und I_a eines MA-Modells in Form eines schwarzen Kastens:

⁷ Es wird unterstellt, daß sich das System im Gleichgewicht befindet. Andernfalls müßte in die Gleichung ein weiteres Glied eingeführt werden, welches den Einfluß der Anfangswerte zum Ausdruck bringt.

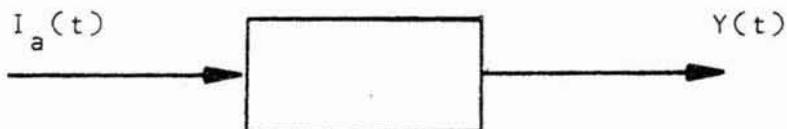


Abb. 21.6 Interpretation eines MA-Systems als ein schwarzer Kasten

Als Eingangsgrößenverlauf sei die Sprungfunktion $I_a(t)=100$ gewählt. Das System soll sich, wie erwähnt, im Gleichgewicht befinden. Wir wählen aus Einfachheitsgründen $Y(0)=0$ und $Y(1)=0$. Nach dem Postulat der ungestörten Überlagerung kann $Y(t)$ beispielsweise dadurch ermittelt werden, daß man die Zeitverläufe der Sprunganworten von einem Eingangsverlauf $E_1(t)=25$ und $E_2(t)=75$ ermittelt und aufsummiert. Es wird damit behauptet, daß die Summe der beiden endogenen Variablen $Y_1(t)$ und $Y_2(t)$

$$Y_1(t) = (\alpha + \alpha\beta)Y_1(t-1) - \alpha\beta Y_1(t-2) + 75$$

$$Y_2(t) = (\alpha + \alpha\beta)Y_2(t-1) - \alpha\beta Y_2(t-2) + 25$$

$$\underline{Y_1(t) + Y_2(t) = (\alpha + \alpha\beta)[Y_1(t-1) + Y_2(t-1)] - \alpha\beta[Y_1(t-2) + Y_2(t-2)] + 100}$$

dem Zeitpfad der Differenzengleichung

$$Y(t) = (\alpha + \alpha\beta)Y(t-1) - \alpha\beta Y(t-2) + 100$$

entspricht. Dieser Fall liegt genau vor, denn durch die Einsetzung von $Y_1(t) + Y_2(t) = Y(t)$ gelangen wir zu der gewünschten Gleichung. Das Prinzip der ungestörten Überlagerung ermöglicht eine wesentliche Vereinfachung der Analyse linearer Systeme, da die Einflüsse der exogenen Variablen isoliert voneinander betrachtet werden können.

Durch das Postulat der Adäquanz von Ursache und Wirkung wird ein weiteres Kennzeichen der Beziehungen zwischen Ein- und Ausgängen eines linearen Modells beschrieben.

Vergegenwärtigen wir uns den beliebigen Verlauf einer Eingangsgröße $E(t)$. Dieser Eingangsgrößenverlauf bewirkt wegen des Postulats der ungestörten Überlagerung einen bestimmten isoliert zu betrachtenden Ausgangsgrößenverlauf $Y_e(t)$. Das Postulat der Adäquanz von Ursache und Wirkung besagt, daß eine k -fache Erhöhung (Verminderung) des Ein-

gangsgrößenverlaufes $kE(t)$ stets eine k -fache Erhöhung (Verminderung) des Ausgangsgrößenverlaufes, d.h. $kY_e(t)$, zur Folge hat. Die Gültigkeit beider Postulate bringt zum Ausdruck, daß ein lineares dynamisches System eine ganz bestimmte von der Höhe der Eingänge unabhängige 'Maschinerie' besitzt, die Eingangsgrößen und Ausgangsgrößen umwandelt. Mit diesem durch die beiden Postulate bereits näher gekennzeichneten Transformationsmechanismus eines linearen dynamischen Systems wollen wir uns im folgenden ausführlicher beschäftigen.

ba) Übergangsverhalten linearer Systeme

a) Allgemeine Kennzeichnung des Übergangsverhaltens

Kennzeichnend für ein lineares dynamisches System ist, wie gesagt, sein Transformationsmechanismus, d.h. die Art und Weise, mit der eine bestimmte Eingangsgröße in eine Ausgangsgröße umgewandelt wird. Die bisher erörterten Prinzipien der ungestörten Überlagerung und der Adäquanz von Ursache und Wirkung liefern eine Information über typische Kennzeichen aller linearen dynamischen Modelle. Im Hinblick auf ein einzelnes System stellt sich jedoch die Frage nach der Art des Transformationsmechanismus in diesem speziellen Fall. Zur Einzelkennzeichnung linearer Systeme verwendet man sogenannte Testantworten. Einem System, welches sich in einem Niveaugleichgewicht befindet, wird eine standardisierte Eingangsgröße aufgeprägt, die auch als Testfunktion bezeichnet wird.

Als Testfunktionen werden am häufigsten der Einheitsimpuls und der Einheitssprung verwendet.

Ein Einheitsimpuls kennzeichnet eine Eingangsgröße, die im Zeitpunkt $t=0$ Eins beträgt und für alle sonstigen Zeitpunkte Null ist, d.h.

$$E^*(t) = \begin{cases} 1 & \text{für } t=0 \\ 0 & \text{für } t=\dots-2, -1, 1, 2, \dots \end{cases}$$

Ein Einheitssprung wird gekennzeichnet durch:

$$E^{**}(t) = \begin{cases} 1 & \text{für } t=0, 1, 2, \dots \\ 0 & \text{für } t=-1, -2, -3, \dots \end{cases}$$

Der von einem Einheitsimpuls (bei einem Niveaugleichgewicht von Null)

hervorgerufene Verlauf der Ausgangsgröße wird als Einheitsimpulsantwort $Y^*(t)$ bezeichnet. Entsprechend leitet sich die Einheitssprungantwort $Y^{**}(t)$ her, welche auf Seite 64 kurz gekennzeichnet wurde. Die graphische Darstellung der beiden Testfunktionsverläufe zeigen Abbildung 21.7 und 21.8.

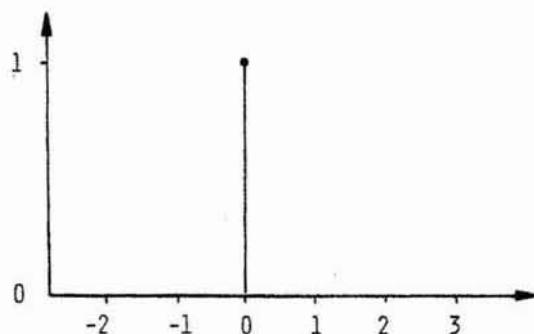


Abb. 21.7 Einheitsimpuls als Testgröße eines dynamischen Systems

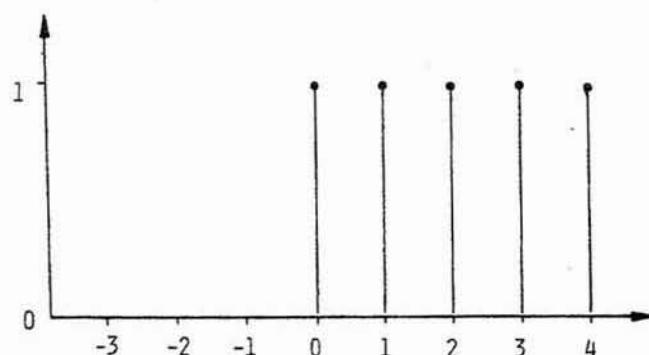


Abb. 21.8 Einheitssprung als Testgröße eines dynamischen Systems

Lineare Systeme zeigen sehr verschiedene Arten von Testantworten. Da im Prinzip unendlich viele Testantworten linearer Systeme möglich sind, ist es notwendig, in einem ersten Überblick bestimmte Teilmengen dynamisch linearer Systeme zu unterscheiden, die dasselbe qualitative Verhalten aufweisen. Um solche Teilmengen zu isolieren, betrachten wir die Funktionslösung der Impulsantwort einer Endgleichung

zweiten Grades. Die entwickelte Klassifizierung versuchen wir daran anschließend auf eine Endgleichung beliebigen Grades zu übertragen. Eine Einheitsimpulsantwort kann bei entsprechender Wahl der Anfangswerte durch eine homogene Endgleichung beschrieben werden. Im Falle der Endgleichung

$$Y(t) = -aY(t-1) - bY(t-2) + E^*(t) \quad Y(-1)=Y(-2)=0 \quad (21.74)$$

ergeben sich die Werte $Y(0)=1$ und $Y(1)=-a$. Setzt man $Y(0)=Z(0)$ und $Y(1)=Z(1)$, so beschreibt die homogene Endgleichung

$$Z(t) = -aZ(t-1) - bZ(t-2)$$

mit $Z(0)=1$ und $Z(1)=-a$ die Einheitsimpulsantwort, welche durch (21.74) bewirkt wird.

Bei Wahl dieser Anfangswerte erhält man entsprechend Tabelle 21.7 die Funktionslösung der Einheitsimpulsantwort

a) im Falle reeller und ungleicher Wurzeln

$$Z(t) = \frac{-\lambda_1}{\lambda_2 - \lambda_1} \lambda_1^t + \frac{\lambda_2}{\lambda_2 - \lambda_1} \lambda_2^t \quad (21.75)$$

b) im Falle reeller und gleicher Wurzeln

$$Z(t) = (1+t)\lambda^t \quad (21.76)$$

und

c) im Falle konjugiert komplexer Wurzeln

$$Z(t) = r^t [\cos \varphi t - \frac{a}{2\sqrt{|\frac{a^2}{4} - b|}} \sin \varphi t] \quad (21.77)$$

Als Beispiel soll die Einheitsimpulsantwort des Volkseinkommens Y im Rahmen des MA-Modells ermittelt werden. Die Endgleichung von Y lautet gemäß (12.9):

$$Y(t) = (\alpha + \alpha\beta)Y(t-1) - \alpha\beta Y(t-2) + I_a(t)$$

Die Wurzeln der charakteristischen Gleichung berechnen sich entsprechend (21.36) mit

$$\lambda_{1,2} = \frac{\alpha + \alpha\beta}{2} \pm \sqrt{\frac{(\alpha + \alpha\beta)^2}{4} - \alpha\beta}$$

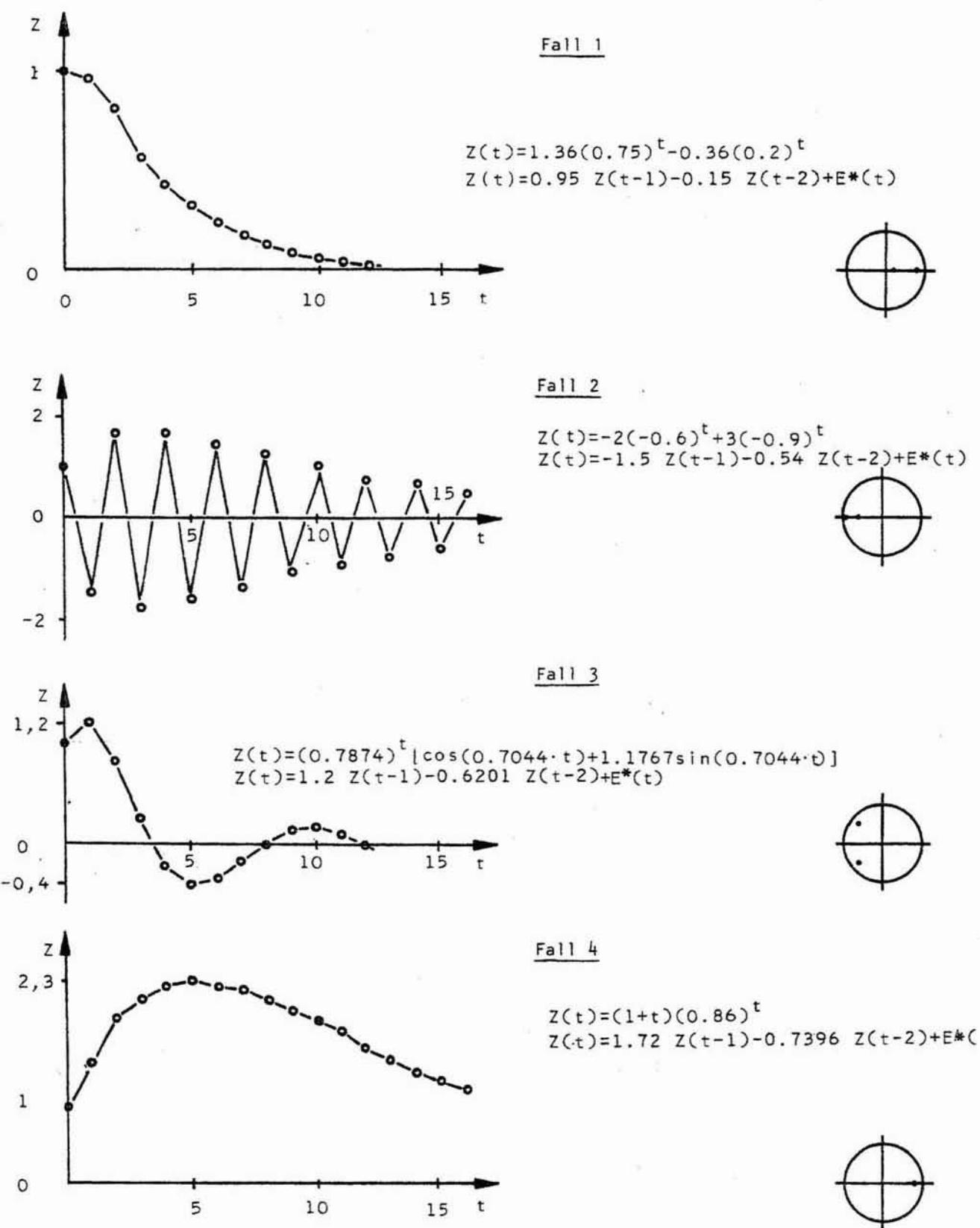


Abb. 21.9 Typische Einheitsimpulsantworten im Falle eines gedämpften Systemverhaltens

Mit $\alpha=0,72$ und $\beta=0,25$ werden $\lambda_1=0,3$ und $\lambda_2=0,6$, d.h. die Wurzeln sind reell und ungleich. Gemäß (21.75) ergibt sich die Funktionslösung der Einheitsimpulsantwort

$$z(t) = 2(0,6)^t - (0,3)^t$$

Die Einheitsimpulsantwort einer Systemvariablen lässt einige Rückschlüsse auf das dynamische Systemverhalten zu.

Sind in den Einheitsimpulsantworten (21.75) und (21.76) die Absolutwerte der Wurzeln $|\lambda_1|<1$ und $|\lambda_2|<1$, dann weist das System ein gedämpftes oder stabiles Verhalten auf. Dasselbe gilt für $r<1$ im Falle von Gleichung (21.77). $z(t)$ konvergiert hierbei gegen Null. Sind in den Fällen (21.75) und (21.76) die Wurzeln λ_1 und λ_2 positiv und kleiner Eins, d.h. $0<\lambda_{1,2}<1$, dann zeigt das System ein monoton gedämpftes Verhalten. Typisch für diese Verhaltensweise, die durch den Fall 1 in Abbildung 21.9 beschrieben wird, ist die fluktuationsfreie Annäherung des Zeitpfades an den Nullpunkt. Das Achsenkreuz mit dem Einheitskreis kennzeichnet die Gaussche Zahlenebene, aus welcher jeweils die Werte der Wurzeln zu erkennen sind.

Sind in den Einheitsimpulsantworten (21.75) und (21.76) die Wurzeln negativ und größer minus Eins, d.h. $0>\lambda_{1,2}>-1$, dann ist das System fluktuerend gedämpft, da die Werte der Wurzeln im Zeitverlauf ständig ihre Vorzeichen wechseln. Ein Beispiel eines derartigen fluktuerend gedämpften Verhaltens zeigt Fall 2 in Abbildung 21.9.

Sind die Wurzeln dagegen konjugiert komplex, so liegt ein oszillatorisch gedämpftes Verhalten mit einem Schwingungszyklus von $2\pi/\varphi$ vor, wie es durch Fall 3 beschrieben wird.

Ein System mit gleichen Wurzeln $\lambda_1=\lambda_2=0,86$ wird durch Fall 4 beschrieben. Einheitsimpulsantworten dieser Art zeichnen sich durch einen eingleipfigen Verlauf aus.

Im Falle eines instabilen oder ungedämpften Verhaltens sind zumindest die Absolutwerte einer Wurzel in (21.75) und (21.76) oder der Modul r in (21.77) größer als Eins. Auch hier kann man, wie aus Abbildung 21.10 ersichtlich ist, analog zu dem vorher Gesagten zwischen monoton ungedämpftem (Fall 5) und oszillatorisch ungedämpftem Verhalten (Fall 6 und 7) unterscheiden.

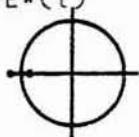
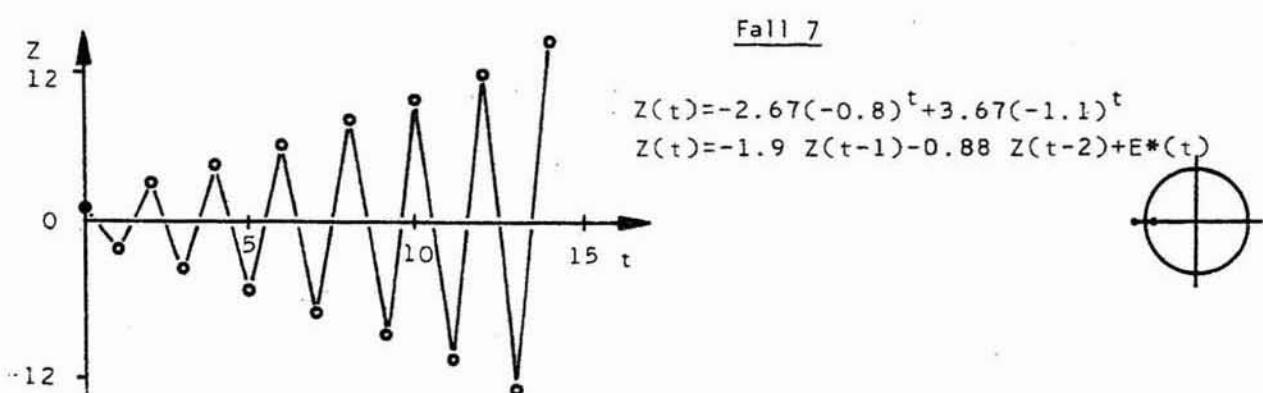
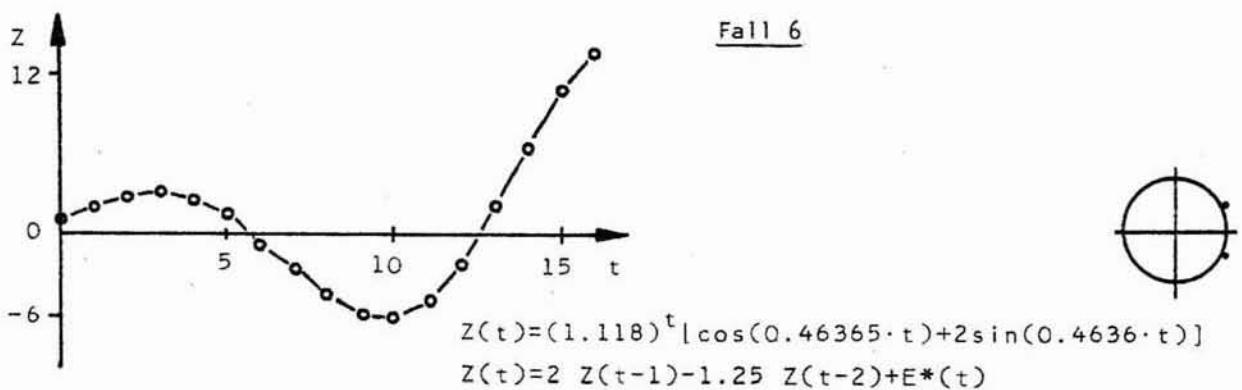
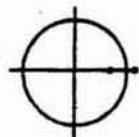
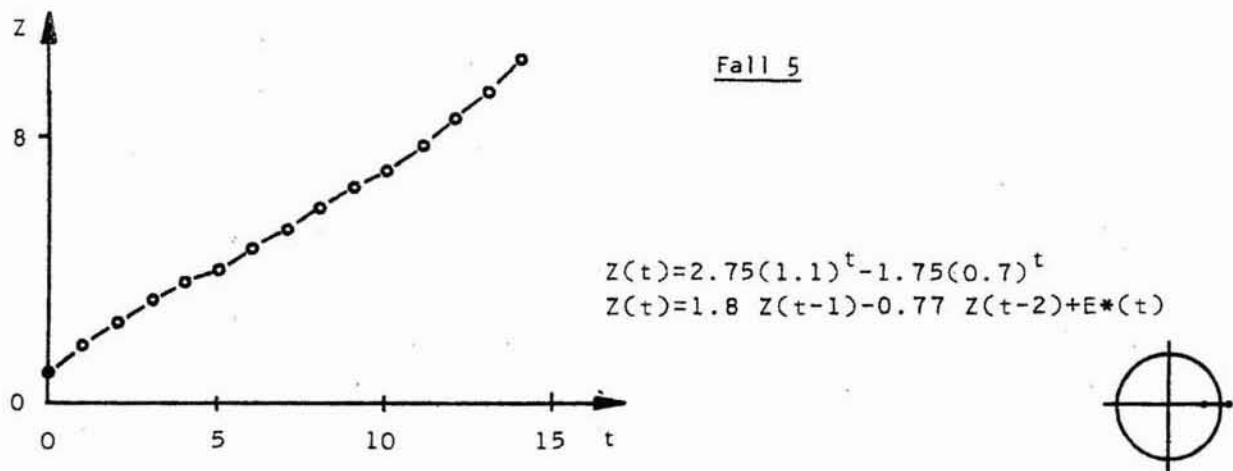


Abb. 21.10 Typische Einheitsimpulsantworten im Falle eines unge-dämpften Systemverhaltens

Für Systeme höherer Ordnung bestimmen sich die Funktionslösungen der Einheitsimpulsantworten aus den Komponenten der auf Seite 198 angegebenen allgemeinen Lösungsansätze. Man erkennt daher: Ist zumindest eine Wurzel oder ein Modul größer als Eins, so ist das System ungeämpft. Sind dagegen sämtliche Wurzeln und Module kleiner als Eins, so ist das System gedämpft.

b) Stabilität als Spezialfall des Übergangsverhaltens

Die Typisierung bestimmter Verhaltensweisen linearer Systeme steht in enger Verbindung zu dem bereits erörterten Stabilitätskonzept dynamischer Systeme.⁸ Im Sinne dieses Konzeptes kann man einen auf das System wirkenden Einheitsimpuls als eine Störung interpretieren und den Variablenverlauf $Y(t)$ für $t=0,1,\dots$, der ohne die Aufprägung des Einheitsimpulses realisiert worden wäre, als ungestörten Zustand oder Gleichgewichtspfad bezeichnen.

Jede endliche Störung eines Systems läßt sich als eine Summe von Einheitsimpulsen interpretieren. Wegen des Postulats der ungestörten Überlagerung setzt sich die durch die Störung bedingte Systemantwort aus der Ordinatensumme der Einheitsimpulsantworten zusammen. Ist daher die Einheitsimpulsantwort begrenzt, so ist auch jede endliche Störung begrenzt, d.h. das betrachtete System ist stabil. Da ein System im Falle von Wurzeln und Modulen, deren Absolutwerte kleiner als Eins sind, eine gedämpfte Einheitsimpulsantwort besitzt, ist es stets unter diesen Bedingungen auch stabil. Diese Stabilitätseigenschaft ist unabhängig von der Höhe der Störung, was zur Folge hat, daß ein lineares System im Falle von Stabilität stets global stabil ist. Weiterhin ist diese Stabilität stets asymptotisch, denn die gedämpfte Einheitsimpulsantwort strebt, wie man zum Beispiel aus den Funktionslösungen (21.75) bis (21.77) erkennt, stets gegen Null.

Der Begriff eines stabilen Systems verlangt die Existenz eines Gleichgewichtspfades, und es stellt sich die Frage, wie in der Funktionslösung einer linearen Differenzengleichung ein derartiger Gleichgewichtspfad zum Ausdruck kommt. Die Funktion
ung einer inhomogenen Differenzengleichung

⁸ Vgl. Seite 75ff

$$Y(t) + a_1 Y(t-1) + \dots + a_n Y(t-n) = E(t) \quad (21.78)$$

setzt sich bekanntlich aus der generellen Funktionslösung ihrer entsprechenden homogenen Gleichung $\tilde{Y}(t)$ und einer speziellen Funktionslösung $\bar{Y}(t)$ zusammen, d.h.

$$Y(t) = \tilde{Y}(t) + \bar{Y}(t) \quad (21.79)$$

Die generelle Funktionslösung $\tilde{Y}(t)$ besitzt hierbei die Form

$$\tilde{Y}(t) = c_1 \lambda_1^t + \dots + c_n \lambda_n^t \quad (21.80)$$

Soll das durch (21.78) beschriebene System stabil sein, so darf eine Störung der Höhe S , der das System ausgesetzt wird, nur zu einer Systemantwort führen, deren aufsummierte Abweichungen vom Gleichgewicht einen endlichen Wert ergeben. Man kann in diesem Fall von einer endlichen Systemantwort sprechen. Analog zur funktionellen Beschreibung einer Einheitsimpulsantwort lässt sich die Störantwort eines Systems durch die Funktionslösung seiner homogenen Endgleichung beschreiben, wenn man als Anfangswerte

$$Y(0) = S/-a_n \quad \text{und} \quad Y(1) = Y(2) = \dots = Y(n-1) = 0$$

wählt.

Diese Darstellungsmöglichkeit aber macht deutlich, daß die Höhe der Störung S nur in den Koeffizienten c_1 bis c_n zum Ausdruck kommt, während die Frage, ob eine endliche Systemantwort gegeben ist, allein von den Werten der Wurzeln λ_1 bis λ_n abhängt. Da die generelle Funktionslösung $\tilde{Y}(t)$, die die Systemstörung zum Ausdruck bringt, stets gegen Null konvergiert, wird der Gleichgewichtspfad des Systems stets durch die spezielle Lösung $\bar{Y}(t)$ beschrieben, welche jedoch entscheidend von dem Verlauf der exogenen Variablen $E(t)$ bestimmt wird.

Besitzt die exogene Variable beispielsweise einen Verlauf der Form $t^p \alpha^t$, so kann die spezielle Lösung bzw. der Gleichgewichtspfad des Systems durch einen Ausdruck der Gestalt

$$\alpha^t (A_0 + A_1 t + A_2 t^2 + \dots + A_p t^p)$$

beschrieben werden. Die Koeffizienten A_0 bis A_p bestimmen sich im konkreten Fall in Abhängigkeit von den numerischen Werten der Parameter α

sowie a_1 bis a_n in (21.78).⁹

Für lineare Systeme zeigt sich damit, daß ihr Systemgleichgewichtspfad primär exogen bestimmt wird. Diese Erkenntnis läßt an der befriedigenden Anwendbarkeit linearer Modelle Zweifel auftreten, wenn gerade das zentrale Element eines linearen Modells, sein Gleichgewichtspfad, keine endogene Erklärung zuläßt. Hier deuten sich die Grenzen der Leistungsfähigkeit linearer Modelle im Hinblick auf eine adäquate Wirklichkeitsbeschreibung an.

γ) Multiplikatoren als Maßzahlen des Übergangsverhaltens

Der Begriff des Multiplikators kann durch folgende Überlegung verdeutlicht werden:

Wird einem System bei Null werdenden Anfangswerten in Periode 0 ein Einheitsimpuls E^* aufgeprägt, so hat dieser - wie erwähnt - eine bestimmte Impulsantwort $Y(0), Y(1), \dots$ zur Folge. Die aufsummierten Werte der Impulsantwort werden durch die unendliche Summe $\sum_{\tau=0}^{\infty} Y(\tau)$ beschrieben.

Falls das betrachtete System stabil ist, besitzt diese unendliche Summe einen endlichen Grenzwert. Dieser Grenzwert wird als Multiplikator oder präziser als Totalmultiplikator von Y bezüglich E bezeichnet. Er beschreibt, anschaulich formuliert, die für die gesamte Zukunft aufsummierten 'Wirkungen' einer 'Ursache' in Höhe von Eins. Bezeichnet man die Einheitsimpulsantwort eines stabilen Systems mit $Y^*(t)$, dann bestimmt sich sein Multiplikator M mit

$$M = \sum_{\tau=0}^{\infty} Y^*(\tau) \quad (21.81)$$

Als Interimsmultiplikator bezeichnet man die Komponenten $Y^*(\tau)$ der Einheitsimpulsantwort. Ihre Zahl ist unendlich. Von Interesse sind jedoch vorwiegend die Interimsmultiplikatoren der ersten Periode, weil sie darüber informieren, wie schnell ein System auf einen Eingang reagiert. Die Interimsmultiplikatoren werden durch die Funktionslösung einer Einheitsimpulsantwort beschrieben. Besteht ein System aus posi-

⁹ Die Anpassung dieser Koeffizienten erfolgt anhand der Methode der unbestimmten Koeffizienten, siehe Seite 198f.

tiven Interimsmultiplikatoren, so beschreibt die Halbwirkungszeit eines Systems die Zeitperiode, in der die aufsummierten Interimsmultiplikatoren die Hälfte des totalen Multiplikatorwertes erreichen. Nach der Definition des Multiplikators durch Gleichung (21.81) setzt seine Ermittlung die Lösung einer unendlichen Summe voraus. Es existiert jedoch ein einfacheres Verfahren zu seiner Bestimmung. Wir gehen von der Annahme aus, einem stabilen System mit Anfangswerten des Betrages Null werde ein Einheitssprung aufgeprägt.

Abbildung 21.11. zeigt einen derartigen Fall, in dem der Zeitpfad der Einheitssprungantwort einem Gleichgewichtspfad M zustrebt. Denkt man sich das System im Gleichgewicht, so befindet sich auch die Eingangsgröße vom Betrag Eins im Gleichgewicht mit der Ausgangsgröße vom Betrag M. M ist damit der Totalmultiplikator des Systems. Der Gleichgewichtspfad der Höhe M wird jedoch durch die spezielle Lösung der Endgleichung:

$$Y(t) + a_1 Y(t-1) + \dots + a_n Y(t-n) = E^{**}(t)$$

beschrieben und ergibt sich analog (21.72) durch

$$M = \frac{1}{1+a_1+\dots+a_n} \quad (21.82)$$

Da, wie wir später sehen werden, die Ungleichung

$$\sum_{\tau=1}^n 1+a_{\tau} > 0$$

eine notwendige Bedingung für die Stabilität eines linearen Systems bildet, muß M im Falle von Systemstabilität stets positiv sein, d.h. ein stabiles System besitzt nur positive Totalmultiplikatoren.¹⁰

Die entwickelten Begriffe sollen am Beispiel eines MA-Systems verdeutlicht werden:

Wir wählen $\alpha=0,9861$ und $\beta=0,7894$. Unter der Annahme, daß $I_a(t)$ durch einen Einheitsimpuls beschrieben wird, erhalten wir die Endgleichung der Einheitsimpulsantwort

$$Y(t) = 1,7647Y(t-1) - 0,7785Y(t-2)$$

mit $Y(0)=1$ und $Y(1)=\alpha+\alpha\beta$

¹⁰ Zur angeführten Bedingung siehe Seite 215

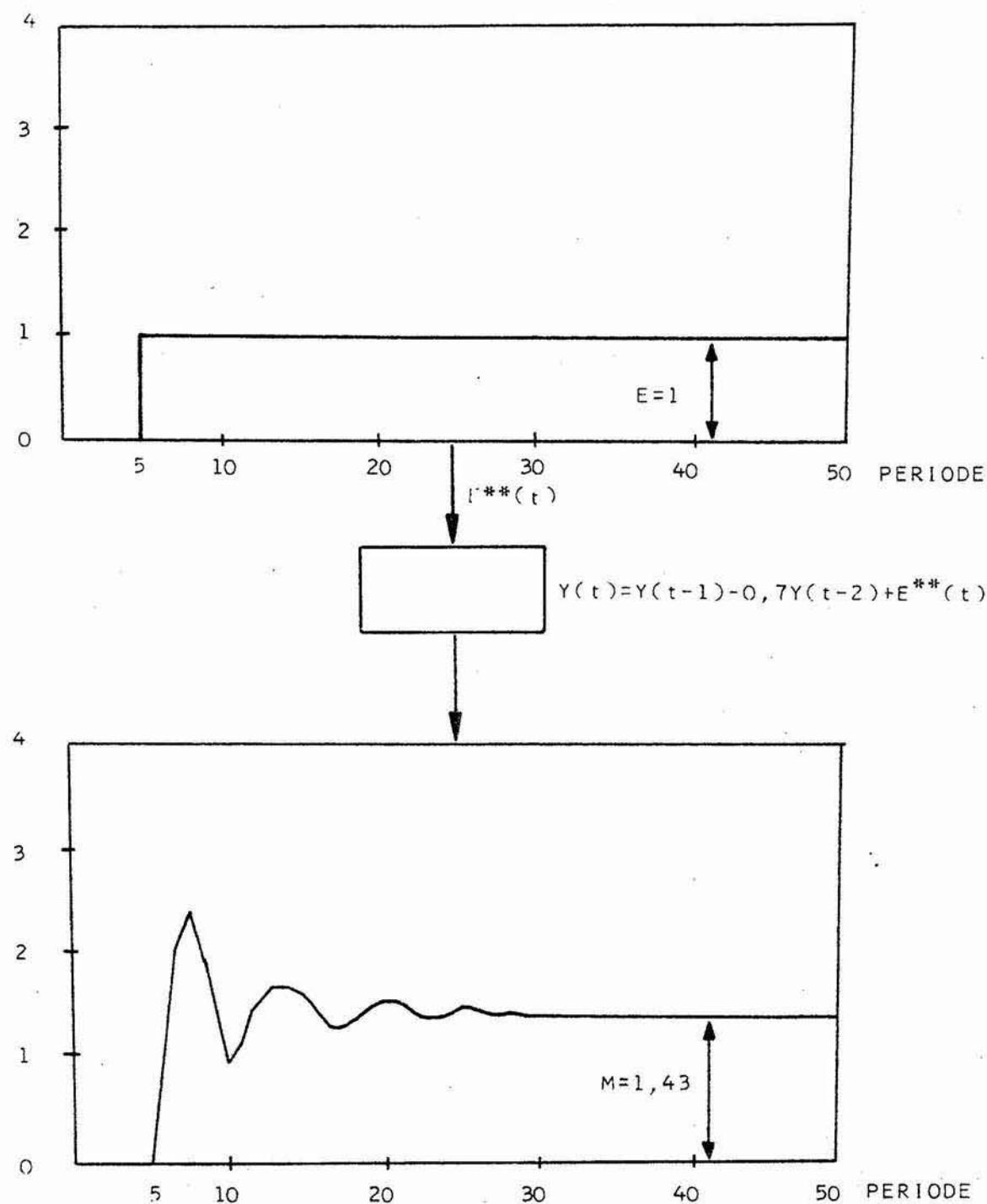


Abb. 21.11 Überlegung zur Bestimmung des Totalmultiplikators eines dynamischen Systems

Der Totalmultiplikator M bestimmt sich nach (21.82) mit

$$M = \frac{1}{1-\alpha(1+\beta)+\alpha\beta} = \frac{1}{1-\alpha}$$

womit M den Wert $M=72$ annimmt.

Die Interimsmultiplikatoren werden durch die Einheitsimpulsantwort des Systems beschrieben, welche gemäß (21.76) und mit $\lambda_1=\lambda_2=0,8826$ die Funktionslösung

$$Y^*(t) = (1+t)0,8826^t$$

besitzt. Der Verlauf der Interimsmultiplikatoren wird durch Abbildung 21.12 beschrieben, aus der auch die Halbwirkungszeit des Systems zu erkennen ist.

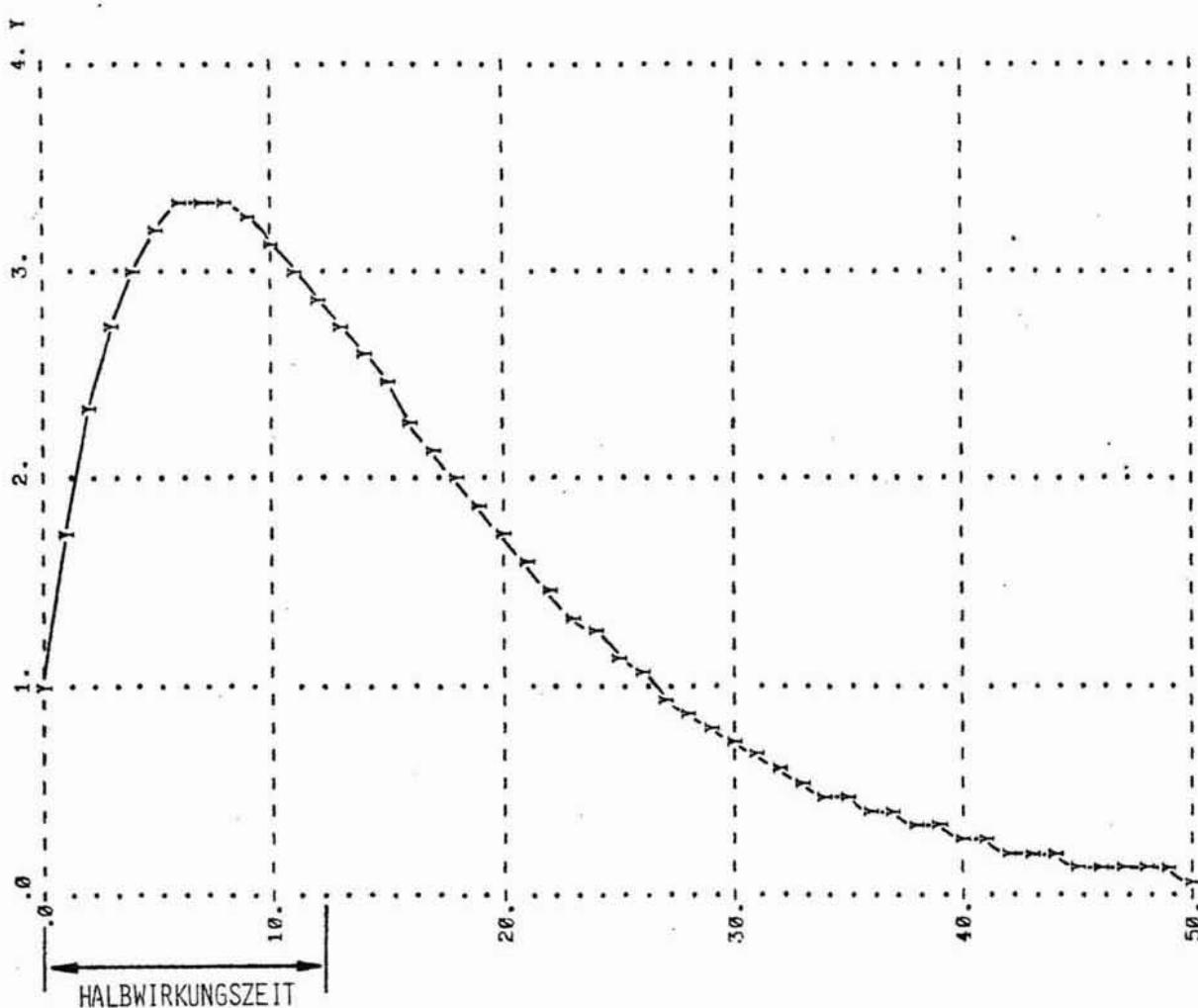


Abb. 21.12 Verlauf der Interimsmultiplikatoren und Halbwirkungszeit eines MA-Systems

δ) Koeffizientenkriterien des Übergangsverhaltens

Zur Gewinnung der Funktionslösung einer Endgleichung ist es stets notwendig, die Wurzeln der charakteristischen Gleichung zu ermitteln.

Bei Endgleichungen höher als dritten Grades ist, wie erwähnt, die Ermittlung der Wurzeln nur noch mit Hilfe iterativer Verfahren möglich. Hat man einen Rechner zur Hand, so ist dies kein Problem, denn es stehen eine Reihe leistungsfähiger Programme zur Errechnung der Polynomwurzeln zur Verfügung. So ist es beispielsweise mit den heute gebräuchlichen Programmen möglich, auf einer mittleren Anlage Polynome 10ten Grades in weniger als einer Sekunde Rechenzeit zu lösen, wenn alle Wurzeln reell sind, während man etwa zwei Sekunden benötigt, wenn die meisten Wurzeln komplex sind. [12]

Für eine schnelle Beurteilung der Stabilität eines linearen Systems bieten sich jedoch auch Verfahren an, mit denen man allein anhand der Koeffizienten einer Endgleichung zu einem Urteil gelangt.

Diese sogenannten Koeffizientenkriterien, die die Erfüllung bestimmter aus den Koeffizienten gebildeter Ungleichungen fordern, sollen im folgenden erörtert werden. Wir wollen zwischen zwei Typen unterscheiden: vollständigen und unvollständigen Koeffizientenkriterien.

Vollständige Koeffizientenkriterien geben die notwendigen und hinreichenden Bedingungen für die Stabilität eines Systems an. Unvollständige Bedingungen kennzeichnen dagegen nur entweder notwendige oder hinreichende Bedingungen.

Die knappsten und übersichtlichsten Bedingungen wurden von JURY entwickelt und stellen eine wesentliche Vereinfachung der Schuhr-Cohn-Kriterien dar, die heute noch in der wirtschaftswissenschaftlichen Literatur verwendet werden. (Siehe zum Beispiel [15])

Ein System mit der Endgleichung

$$Y(t) + a_1 Y(t-1) + \dots + a_n Y(t-n) = E(t)$$

ist dann und nur dann stabil, wenn die folgenden Ungleichungen erfüllt werden. [104]

- | | |
|---------|------------------------|
| $n = 2$ | 1) $1 + a_1 + a_2 > 0$ |
| | 2) $1 - a_1 + a_2 > 0$ |
| | 3) $a_2 - 1 < 0$ |

$$\begin{array}{ll}
 n = 3 & \begin{aligned} 1) \quad 1 + a_1 + a_2 + a_3 &> 0 \\ 2) \quad 1 - a_1 + a_2 - a_3 &> 0 \\ 3) \quad |a_3| &< 1 \\ 4) \quad a_3^2 - 1 &< a_3 a_1 - a_2 \end{aligned} \\
 n = 4 & \begin{aligned} 1) \quad 1 + a_1 + a_2 + a_3 + a_4 &> 0 \\ 2) \quad 1 - a_1 + a_2 - a_3 + a_4 &> 0 \\ 3) \quad 1 - a_4^2 + a_4 a_1 - a_3 &> 0 \\ 4) \quad 1 - a_4^2 - a_4 a_1 + a_3 &> 0 \\ 5) \quad a_4^3 + 2a_4 a_2 + a_3 a_1 - a_4 - a_2 - a_4 a_1^2 - a_4^2 - a_4^2 a_2 - \\ &- a_3^2 + a_4 a_3 a_1 + 1 > 0 \end{aligned}
 \end{array}$$

JURYs Kriterien sind für Endgleichungen beliebigen Grades formuliert. Wir beschränken uns jedoch auf den Fall von Endgleichungen zweiten bis vierten Grades, da im Falle von Endgleichungen höheren Grades die Zahl der Ungleichungen stark zunimmt und die Ausdrücke so unübersichtlich werden, daß sich eine praktische Anwendung nicht mehr empfiehlt.

Notwendige Stabilitätsbedingungen sind für das Vorhandensein von Stabilität erforderlich, garantieren jedoch keine Stabilität. Systeme, die die Notwendigkeitsbedingungen nicht erfüllen, sind aber auf jeden Fall instabil. Die Bedingungen 1) und 2) des Jury-Kriteriums führen verallgemeinert zu den folgenden notwendigen Stabilitätsbedingungen einer Endgleichung n-ten Grades:

$$\begin{aligned} a) \quad 1 + a_1 + a_2 + \dots + a_n &> 0 \\ b) \quad 1 + \sum_{i=1}^n (-1)^i a_i &> 0 \end{aligned}$$

Besonders a) gibt einen ersten Eindruck über das Modellverhalten. Weitere notwendige Bedingungen für die Stabilität einer Endgleichung lauten: [1,S.215-219]

$$\begin{aligned} n = 2 \quad |a_2| &< 1, |a_1| < 2 \\ n = 3 \quad |a_3| &< 1, -1 < a_2 < 3, |a_1| < 3 \\ n = 4 \quad |a_4| &< 1, |a_3| < 4, -2 < a_2 < 6, |a_1| < 4 \end{aligned}$$

Eine hinreichende Stabilitätsbedingung liefert:

$$1 > |a_1| + |a_2| + \dots + |a_n|$$

In den Wirtschaftswissenschaften werden manchmal Modelle in der Erklärungsform einer homogenen Endgleichung, d.h.

$$Y(t) = \omega_1 Y(t-1) + \omega_2 Y(t-2) + \dots + \omega_n Y(t-n)$$

verwendet, wobei $\omega_1, \omega_2, \dots$ positive Werte sein sollen. Eine hinreichende Stabilitätsbedingung für diesen Typ ist: [197, S.104ff.]

$$\omega_1 > \omega_2 > \dots > \omega_{n-1} > \omega_n > 0$$

d.h. das System ist stabil, wenn die Koeffizienten mit zunehmender Verzögerung der endogenen Variablen abnehmen.

bb) Verhaltensdiagramme linearer Systeme

Die Wurzeln einer Endgleichung bestimmen sich aus ihren Koeffizienten. Im Falle von Endgleichungen zweiten Grades mit den Koeffizienten a_1 und a_2 liegt es nahe, ein Verhaltensdiagramm dieses Endgleichungstyps aufzustellen. In einem derartigen Verhaltensdiagramm werden die Koeffizientenausprägungen als Ordinaten- bzw. Abszissenwerte definiert. Indem die Bereiche unterschiedlichen Systemverhaltens durch Ungleichungen mit a_1 und a_2 beschrieben werden können, erhält man im a_1/a_2 Koordinatensystem bestimmte Flächenbereiche unterschiedlichen Verhaltens.

Abbildung 21.13 zeigt das Verhaltensdiagramm der Endgleichung

$$Y(t) + a_1 Y(t-1) + a_2 Y(t-2) = E(t)$$

Die Gleichungen $a_1=a_2+1$, $a_2=-a_1-1$ und $a_2=1$ folgen aus JURYS Koeffizientenkriterien auf Seite 214. Die schraffierte Flächen beschreiben den Bereich stabilen Verhaltens. Es ist jedoch eine wesentlich differenziertere Klassifizierung des Systemverhaltens möglich. Man kann 14 Flächenbereiche unterscheiden, in denen unterschiedliche Verhaltensweisen auftreten. Als Beispiel seien die ersten fünf Bereiche angeführt:

- I Oszillatorisches stabiles Verhalten (Wurzeln konjugiert komplex, Modul kleiner als Eins)
- II Oszillatorisches instabiles Verhalten (Wurzeln konjugiert komplex, Modul größer als Eins)
- III Monoton stabiles Verhalten (beide Wurzeln positiv und kleiner als Eins)

- IV Monoton instabiles Verhalten (beide Wurzeln positiv, eine größer, eine kleiner als Eins)
 V Monoton instabiles Verhalten (beide Wurzeln positiv und größer als Eins)

usw.

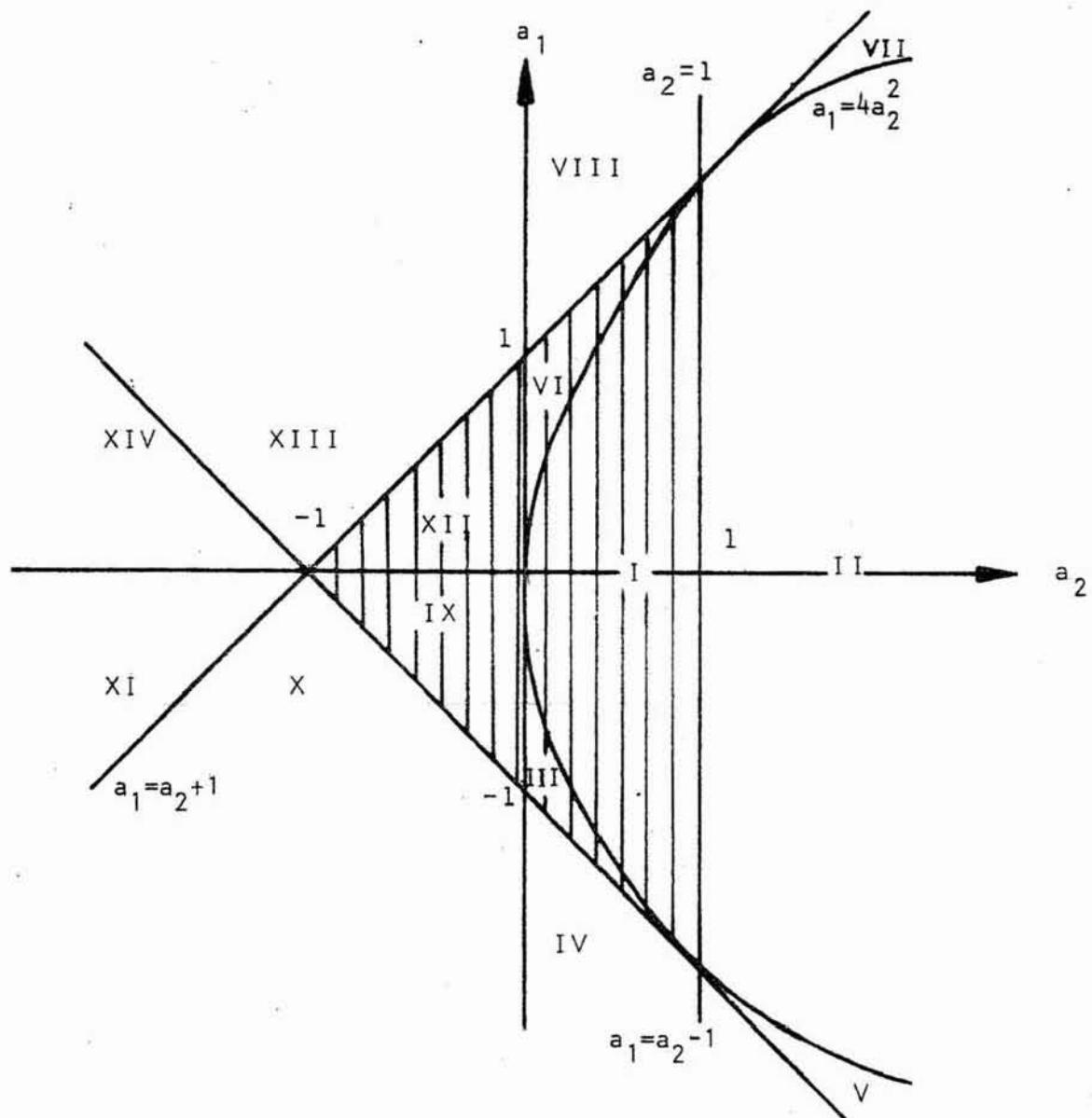


Abb. 21.13 Verhaltensdiagramm einer linearen Endgleichung zweiten Grades

Manche Flächenbereiche wie IV und V fallen in dieselbe Klassifizierungskategorie.

Die Ermittlung der verschiedenen Verhaltensbereiche hat sich an der

Wurzelgleichung

$$\lambda_{1,2} = \frac{-a_1}{2} \pm \sqrt{\frac{a_1^2}{4} - a_2} \quad (21.83)$$

zu orientieren.

Der Fall gleicher Wurzeln wird durch $a_1^2=4a_2$ beschrieben. Im Falle konjugiert komplexer Wurzeln muß die Ungleichung $a_1^2 < 4a_2$ erfüllt sein.

Diese Bedingung gilt damit für die Flächen I und II.

Es empfiehlt sich, zur weiteren Klassifizierung anhand von Gleichung (21.83) Ausdrücke zu finden, in denen a_1 und a_2 direkte Funktionen von λ_1 und λ_2 darstellen. Aus Gleichung (21.83) folgt

$$\lambda_1 + \lambda_2 = -a_1$$

oder

$$a_1 = -(\lambda_1 + \lambda_2) \quad (21.84)$$

Mit

$$\lambda_1 \lambda_2 = \left[\frac{-a_1}{2} + \sqrt{\frac{a_1^2}{4} - a_2} \right] \left[\frac{-a_1}{2} - \sqrt{\frac{a_1^2}{4} - a_2} \right]$$

folgt

$$\lambda_1 \lambda_2 = a_2 \quad (21.85)$$

Anhand von (21.84) und (21.85) kann man auf relativ einfache Art die Bereichsabgrenzungen finden, indem man von typischen Wertebereichen der Wurzeln ausgeht. Beispielsweise wollen wir den Bereich des Stabilitätsdiagrammes, für den das System ein monoton stabiles Verhalten aufweist, ermitteln.

Wir wissen bereits, daß nur die Flächen III, VI, XII und IX in Frage kommen. Für den Fall monotoner Stabilität gilt, daß beide Wurzeln positiv und kleiner als Eins sind. Daher muß im Falle positiver Wurzeln auch ihr Produkt $\lambda_1 \lambda_2 > 0$ sein. Gleichung (21.85) erfordert deshalb:

$$a_2 > 0$$

Damit kommen nur noch die Flächen VI und III in Frage. Da beide Wurzeln positiv sein sollen, muß gemäß (21.84) $a_1 < 0$ sein. Es zeigt sich also, daß allein die durch Fläche III gekennzeichneten Koordinatenwerte von a_1 und a_2 ein monoton stabiles Verhalten aufweisen.

Abbildung 21.14 zeigt einige ausgewählte Impulsantworten verschiedener Verhaltensbereiche.

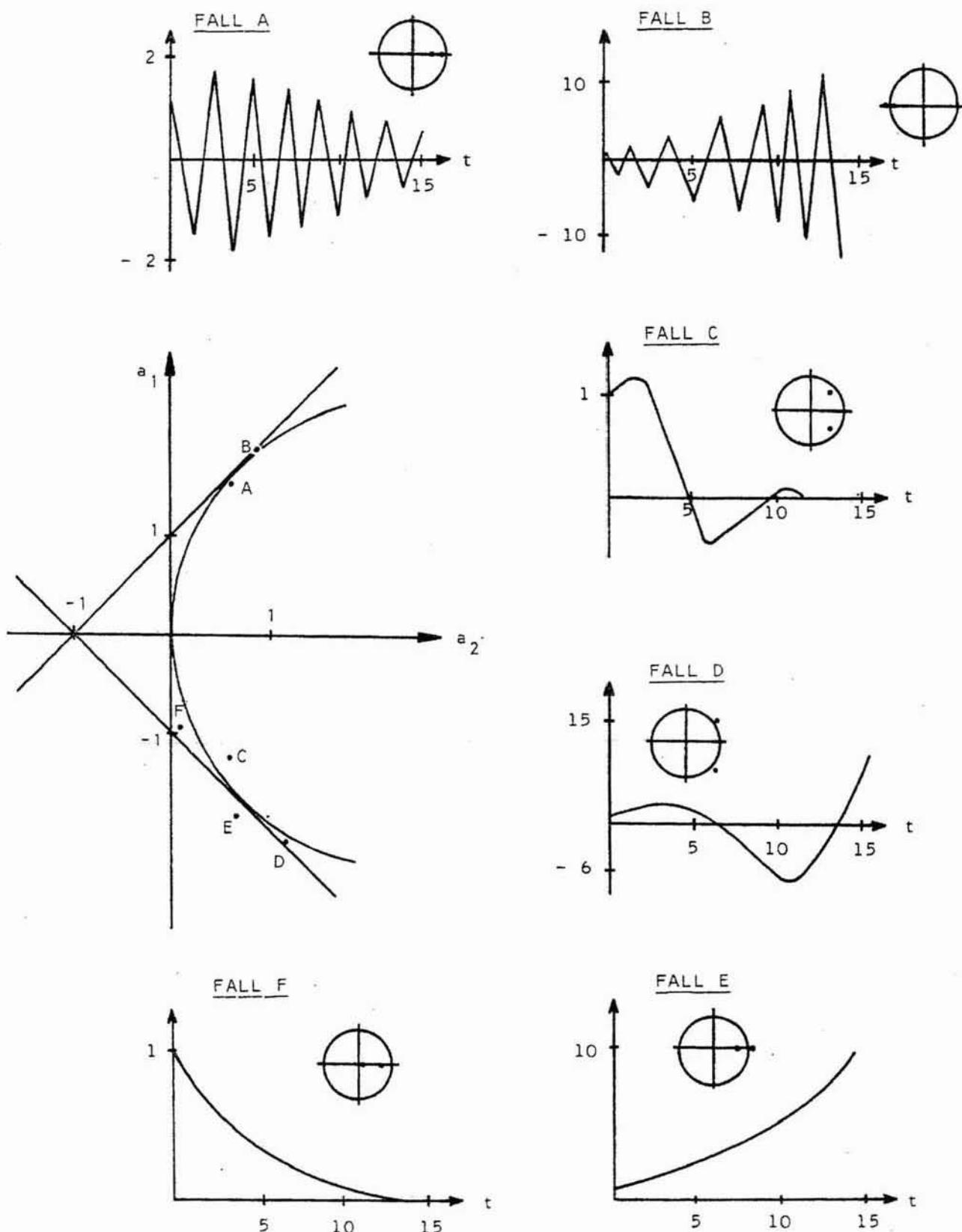


Abb. 21.14 Einheitsimpulsantwort und Lage der Wurzeln der charakteristischen Gleichung für ausgewählte Beispiele einer Endgleichung zweiten Grades

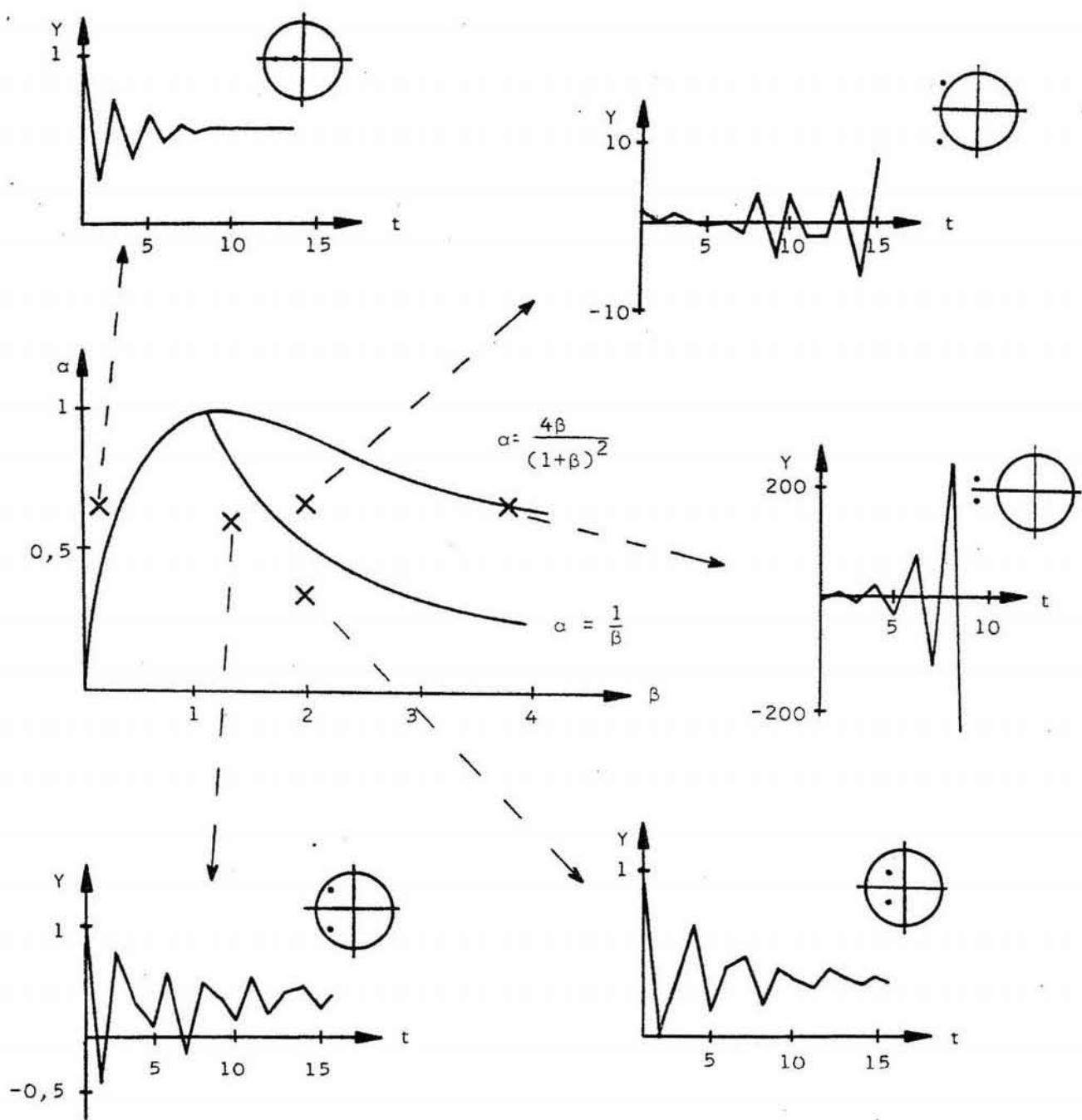


Abb. 21.15 Verhaltensdiagramm eines MA-Systems mit ausgewählten Einheitssprungantworten

Die Entwicklung von Verhaltensdiagrammen ist besonders dann ange-

bracht, wenn das betrachtete Modell genau zwei strukturelle Parameter enthält. In diesem Fall ist es sehr instruktiv, ein Verhaltensdiagramm zu entwickeln, dessen Koordinaten durch die (empirisch interpretierbaren) strukturellen Parameter gebildet werden. Denn man erkennt, wie die Veränderungen dieser Parameter zu unterschiedlichen Verhaltensweisen führen. Im Beispiel eines MA-Systems kann anhand der Endgleichung

$$Y(t) = (\alpha + \beta Y(t-1)) - \alpha \beta Y(t-2) + I_a(t)$$

das in Abbildung 15.8 auf Seite 92 bereits angeführte Verhaltensdiagramm für die Parameter α und β entwickelt werden.

Abbildung 21.15 zeigt das Verhaltensdiagramm eines MA-Systems einschließlich der Impulsantworten ausgewählter Parameterkombinationen. Jeder Impulsantwort ist das Koordinatensystem einer Gaußschen Zahlenebene mit einem Einheitskreis zugeordnet. Es beschreibt die Lage der Wurzeln der charakteristischen Gleichung im Falle der betreffenden Parameterkombination.

Die Entwicklung derartiger Verhaltensdiagramme zeigt die Sensitivität des Modells bezüglich der Parameter α und β und liefert damit wichtige Hinweise für seine Gültigkeit.

c) Höhere Analysemethoden linearer Systeme

ca) Verwendung von Operatoren in linearen Systemen

Zur Untersuchung von Differenzengleichungssystemen ist es oft sehr vorteilhaft, mit sogenannten Verschiebungsooperatoren zu arbeiten. Ein Verschiebungsooperator kann als ein Symbol angesehen werden, mit dessen Hilfe verzögerte Variable in einer anderen Schreibweise ausgedrückt werden. Mit dem Ausdruck K^{-n} kann die verzögerte Variable $X(t-n)$ durch

$$K^{-n} X(t) = X(t-n) \quad (21.86)$$

ersetzt werden.

Die Differenzengleichung

$$Y(t) + 0,5Y(t-1) + 0,3Y(t-2) = E(t) + 1,1E(t-1)$$

würde unter Verwendung dieser Operatorenenschreibweise die Gestalt

$$Y(t) + 0,5K^{-1}Y(t) + 0,3K^{-2}Y(t) = E(t) + 1,1K^{-1}E(t)$$

annehmen. Der Ausdruck K wird als Verschiebungsooperator bezeichnet. Die Einführung einer derartigen Schreibweise der verzögerten Variablen ist deswegen sinnvoll, weil sich zeigen läßt, daß der Verschiebungsooperator verschiedenen Regeln der Algebra gehorcht. Man kann mit K^{-n} daher Operationen wie mit einer Zahl durchführen.

Die Gültigkeit der wichtigsten algebraischen Regeln sei anhand von Beispielen aufgezeigt:

- (1) Es gilt das Kommutativgesetz, d.h. der Fall $a+b=b+a$ wird erfüllt.
So ist die Beziehung

$$K^{-3}Y(t) + K^{-2}Y(t) = K^{-2}Y(t) + K^{-3}Y(t)$$

zulässig, weil auch die entsprechende Differenzengleichung

$$Y(t-3) + Y(t-2) = Y(t-2) + Y(t-3)$$

gültig ist.

- (2) Ebenso gilt im Falle einer Multiplikation das Kommutativgesetz, d.h. $a \cdot b = b \cdot a$

Dies erkennt man an dem Beispiel

$$[K^{-3}(K^{-2})]Y(t) = [K^{-2}(K^{-3})]Y(t) = K^{-5}Y(t)$$

Der Ausdruck $[K^{-3}(K^{-2})]Y(t)$ entspricht $Y(t-5)$. Denn es ist $K^{-2}Y(t)=Y(t-2)$ und damit $K^{-3}Y(t-2)=Y(t-5)$. Für den Ausdruck $[K^{-2}(K^{-3})]Y(t)$ ergibt sich mit $K^{-3}Y(t)=Y(t-3)$ die Gestalt $[K^{-2}(Y(t-3))]$ und damit $[K^{-2}(Y(t-3))]=Y(t-5)$

- (3) Ferner ist sowohl für die Additionen als auch Multiplikationen das Assoziativgesetz anwendbar, d.h.

$$a + (b+c) = (a+b) + c$$

$$a(bc) = (ab)c$$

Als Beispiel sei die Beziehung

$$K^{-1}[K^{-1}+1]Y(t) = [K^{-2}+K^{-1}]Y(t)$$

angeführt. Wie man durch schrittweise Rückumwandlung der Glieder auf beiden Seiten der Gleichung erkennt, ist die Beziehung gültig.

Auf die Verwendung von Operatoren werden wir später noch mehrmals stoßen. Als erster Anwendungsbereich sei auf die vereinfachte Ermittlung der Endgleichung linearer Systeme verwiesen. Die Ermittlung der Endgleichung im Falle des dynamischen Modells der Anspruchsniveaumanpassung wurde auf Seite 171 beschrieben. Die Verwendung von Verschiebungsooperatoren vereinfacht das Verfahren wesentlich.

Mit der Einführung des Verschiebungsoperators $K^{-\eta}X(t)=X(t-\eta)$ erhält man aus (21.6) bis (21.9)

$$\begin{aligned} S(t) &= \beta[\bar{Z}-Z(t)] \\ B(t) &= \gamma[K^{-1}S(t)-b]/[1-K^{-1}+\gamma c K^{-1}] \\ Z(t) &= B(t) - A(t) \\ A(t) &= \alpha[K^{-1}B(t)+a]/[1-K^{-1}+\alpha K^{-1}] \end{aligned}$$

Man kann mit den sich ergebenden Gleichungen wie in einem linearen Gleichungssystem $A(t)$ berechnen. Durch sukzessives Einsetzen und Freistellen von $A(t)$ gelangt man zu der Gleichung:

$$A(t) = [2-\alpha-c\gamma-\gamma\beta]K^{-1}A(t) + [\alpha-1-c\gamma\alpha+c\gamma+\gamma\beta]K^{-2}A(t) + \alpha ac\gamma + \alpha\beta\bar{Z} + \alpha a\gamma\beta - a\gamma b$$

Die Rücktransformation $X(t-\eta)=K^{-\eta}X(t)$ führt schließlich zu der uns bereits bekannten Endgleichung

$$A(t) = [2-\alpha-c\gamma-\gamma\beta]A(t-1) + [\alpha-1-c\gamma\alpha+c\gamma+\gamma\beta]A(t-2) + \alpha ac\gamma + \alpha\beta\bar{Z} + \alpha a\gamma\beta - a\gamma b$$

Den durch eine Transformation der Form (21.86) in die Gleichungen eingebrochenen Operator nennt man einen Rückwärtsoperator, da er negative Hochzahlen besitzt. Oft wird jedoch auch mit Vorwärtsoperatoren gearbeitet, d.h. Operatoren mit positiven Hochzahlen. Beide Darstellungsformen sind ineinander überführbar. Die Differenzengleichung

$$Y(t) + 0,5Y(t-1) + 0,3Y(t-2) = E(t) + 1,1E(t-1)$$

führte bei Anwendung eines Rückwärtsoperators zu der Form

$$Y(t) + 0,5K^{-1}Y(t) + 0,3K^{-2}Y(t) = E(t) + 1,1K^{-1}E(t) \quad (21.87)$$

Bezeichnet man mit n den höchsten Verzögerungsgrad einer Endgleichung, dann ist in diesem Fall $n=2$. Die Multiplikation von Gleichung (21.87) mit K^2 führt zur Darstellung der ursprünglichen Differenzengleichung in Form von Vorwärtsoperatoren, d.h.

$$K^2Y(t) + 0,5KY(t) + 0,3Y(t) = K^2E(t) + 1,1KE(t) \quad (21.88)$$

Verallgemeinernd kann man feststellen, daß in diesem Fall von der Transformation

$$X(t-\eta) = K^{n-\eta}X(t)$$

ausgegangen wird. Diese Darstellungsweise hat den Vorteil, daß das sich ergebende Operatorenpolynom dieselbe Form wie die charakteristische Gleichung der zugrunde liegenden Differenzengleichung besitzt. Sie ist immer dann empfehlenswert, wenn schon die Endgleichung eines Systems vorliegt, weil n dann als numerischer Wert zur Verfügung steht.

Bevor wir uns dem generellen Fall der Ermittlung von Endgleichungen unter Verwendung von Operatoren im Rahmen der Matrizenrechnung zuwenden, wollen wir kurz auf bestimmte Verfahren eingehen, die es gestatten, Endgleichungen anhand der sukzessiven Umgestaltung von graphischen Systemdarstellungen zu ermitteln.

cb) Endgleichungsbestimmung anhand graphischer Systemdarstellungen

Der Leser wird sich erinnern, daß wir bei der Behandlung der graphischen Darstellung von Systemen die Beschreibung dynamischer Systeme anhand von Block- und Signalflußdiagrammen erwähnten. Es wurde an dieser Stelle darauf hingewiesen, daß Signalfluß- und Blockdiagramme, in denen Operatordarstellungen der Übergangsfunktion verwendet werden, allein dem Zweck dienen, die Endgleichung des betrachteten Systems auf einfache Weise zu ermitteln.¹¹ Die hier zur Anwendung kommenden Methoden sollen im folgenden behandelt werden.

¹¹ Vgl. Seite 65

a) Endgleichungsbestimmung anhand von Blockdiagrammen

Abbildung 21.16 zeigt noch einmal die Darstellung eines MA-Systems im Rahmen der beiden Diagrammtypen. Die Operatorenform der Übergangsfunktion ergibt sich durch die Einführung des Rückwärtsoperators $K^{-\eta} X(t) = X(t-\eta)$.

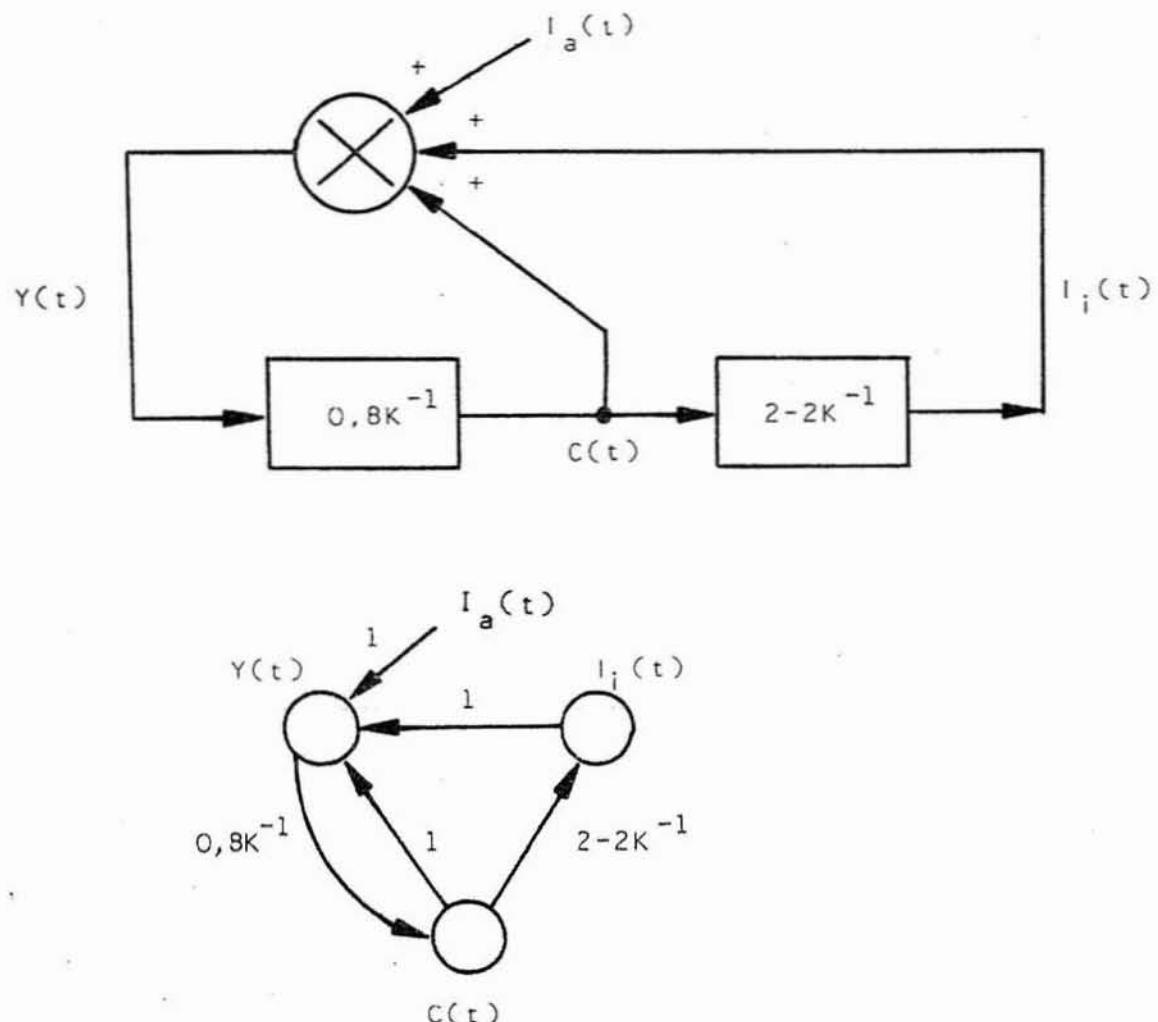


Abb. 21.16 Block- und Signalflußdiagramm eines MA-Systems unter Verwendung eines Rückwärtsoperators

Im Falle der Hypothese der induzierten Investition

$$i_i(t) = 2[c(t) - c(t-1)]$$

wird mit (21.86)

$$i_i(t) = 2[c(t) - K^{-1}c(t)]$$

oder

$$I_i(t) = [2 - 2K^{-1}]C(t)$$

Der Ausdruck $2 - 2K^{-1}$, welcher die beiden unverzögerten Variablen C und I miteinander verknüpft, wird als Operatorenübergangsfunktion von C nach I bezeichnet. Es ist üblich, sie in den entsprechenden Block des Diagrammes einzutragen oder über der entsprechenden Strecke des Signalflußdiagrammes anzuführen.

Bezeichnet man die ermittelte Operatorenübergangsfunktion mit $G(K)$, d.h.

$$G(K) = 2 - 2K^{-1}$$

dann läßt sich die Investitionshypothese durch

$$I_i(t) = G(K)C(t)$$

beschreiben.

Eine Beziehung zwischen einer Ein- und Ausgangsgröße, welche durch die Standardform der Endgleichung (12.10)

$$Y(t) + a_1 Y(t-1) + \dots + a_n Y(t-n) = g_0 E(t) + g_1 E(t-1) + \dots + g_s E(t-s)$$

beschrieben werden kann, besitzt daher unter entsprechender Verwendung eines Rückwärtsoperators eine Operatorenübergangsfunktion der Form

$$G(K) = \frac{g_0 + g_1 K^{-1} + g_2 K^{-2} + \dots + g_s K^{-s}}{1 + a_1 K^{-1} + a_2 K^{-2} + \dots + a_n K^{-n}}$$

Im folgenden sollen Verfahren beschrieben werden, mit denen unter Zu-
grundelegung der Operatorenübergangsfunktion von Verhaltengleichun-
gen die totale Übergangsfunktion eines Modells ermittelt werden kann.
Diese Ermittlung vollzieht sich anhand der sukzessiven Umgestaltung
bestimmter Block- und Signalflußdiagramme.

Die zu beschreibenden Verfahren sind primär im Rahmen der Regelungs-
theorie für stetige Systeme entwickelt worden, lassen sich jedoch ana-
log auch für zeitdiskrete Systeme verwenden. Ihre Anwendung besteht
darin, daß bestimmte Blöcke eines Blockdiagrammes unter Berechnung
der neuen Übergangsfunktion durch einfache Vorschriften zu einem ein-
zigen Block zusammengefaßt werden. Man kann drei Grundtypen der Zu-

sammenfassung von Blockdiagrammgliedern unterscheiden.

Erstens: die Zusammenfassung von Parallelgliedern. Hier gilt die Vorschrift

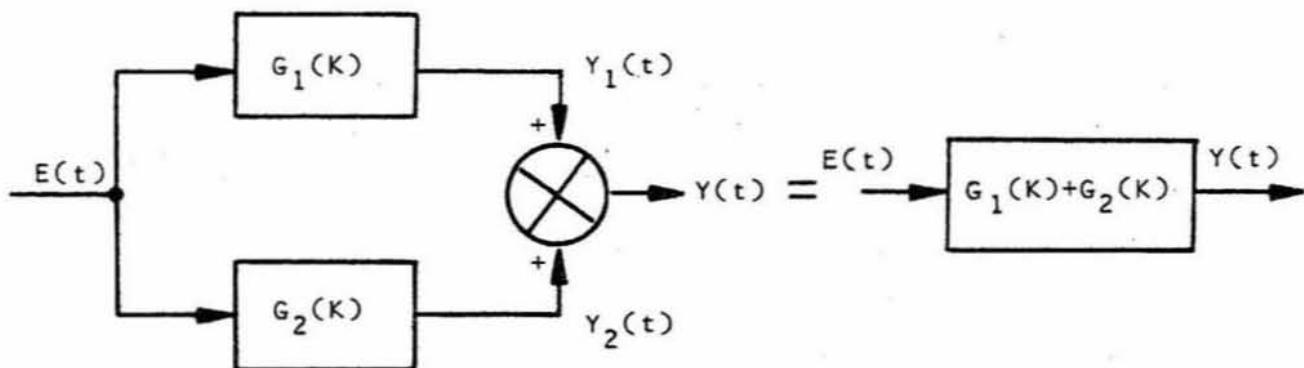


Abb. 21.17 Ermittlung der Übergangsfunktion eines linearen dynamischen Systems im Falle paralleler Glieder

Die Gültigkeit dieser Zusammenfassung wollen wir uns anhand des Gleichungssystems

$$\begin{aligned} Y_1(t) &= 0,5Y_1(t-1) + E(t) + 1,1E(t-1) \\ Y_2(t) &= 0,2Y_2(t-1) + 0,3E(t) \\ Y(t) &= Y_1(t) + Y_2(t) \end{aligned} \quad (21.89)$$

klar machen, welches bei Einführung eines Rückwärtsoperators durch

$$Y_1(t)[1-0,5K^{-1}] = [1+1,1K^{-1}]E(t) \quad (21.90)$$

$$Y_2(t)[1-0,2K^{-1}] = 0,3E(t) \quad (21.91)$$

beschrieben wird.

Die Auflösung der Gleichungen (21.90) und (21.91) nach Y_1 und Y_2 und ihre Einsetzung in Gleichung (21.89) liefert:

$$Y(t) = \frac{1+1,1K^{-1}}{1-0,5K^{-1}} E(t) + \frac{0,3}{1-0,2K^{-1}} E(t)$$

oder

$$Y(t) = \left[\frac{1+1,1K^{-1}}{1-0,5K^{-1}} + \frac{0,3}{1-0,2K^{-1}} \right] E(t)$$

Der erste und zweite Ausdruck in den Klammern sind jedoch mit der Ope-

ratorenübergangsfunktion $G_1(K)$ und $G_2(K)$ identisch. Denn es sind

$$G_1(K) = \frac{1+1,1K^{-1}}{1-0,5K^{-1}} \quad \text{und} \quad G_2(K) = \frac{0,3}{1-0,2K^{-1}}$$

Die Addition von $G_1(K)$ und $G_2(K)$ liefert die totale Übergangsfunktion

$$G(K) = \frac{1,3+0,75K^{-1}-0,22K^{-2}}{1-0,7K^{-1}+0,1K^{-2}}$$

und mit der Rücktransformation des Operators ihre entsprechende Endgleichung

$$Y(t) - 0,7Y(t-1) + 0,1Y(t-2) = 1,3E(t) + 0,75E(t-1) - 0,22E(t-2)$$

Die zweite Vorschrift bezieht sich auf die Zusammenfassung kaskadierender Glieder. Es gilt das Reduktionsschema:

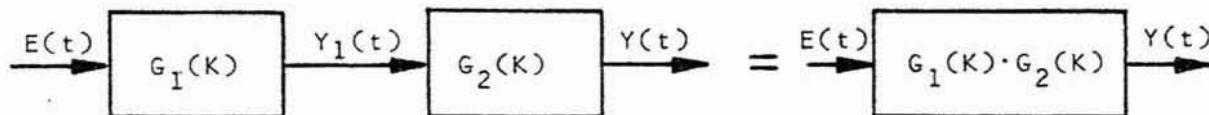


Abb. 21.18 Ermittlung der Übergangsfunktion eines linearen dynamischen Systems im Falle kaskadierender Glieder

Die Richtigkeit dieser Reduktion kann leicht aus der Verallgemeinerung des folgenden Beispiels erkannt werden. Zwei kaskadierende Systeme können durch

$$Y(t) = 0,3Y(t-1) + Y_1(t) \tag{21.92}$$

$$Y_1(t) = 1,1Y_1(t-1) + 1,3E(t-1) \tag{21.93}$$

beschrieben werden. $G_1(K)$ und $G_2(K)$ ergeben sich mit

$$G_1(K) = \frac{1,3K^{-1}}{1-1,1K^{-1}} \tag{21.94}$$

$$G_2(K) = \frac{1}{1-0,3K^{-1}}$$

Die Operatorentransformation von (21.92) und (21.93) liefert

$$Y(t)[1-0,3K^{-1}] = Y_1(t)$$

$$Y_1(t)[1-1,1K^{-1}] = 1,3K^{-1}E(t)$$

oder

$$Y(t) = \frac{1}{1-0,3K^{-1}} Y_1(t) \quad (21.95)$$

$$Y_1(t) = \frac{1,3K^{-1}}{1-1,1K^{-1}} E(t) \quad (21.96)$$

Mit (21.96) in (21.95) folgt:

$$Y(t) = \frac{1}{1-0,3K^{-1}} \cdot \frac{1,3K^{-1}}{1-1,1K^{-1}} E(t) \quad (21.97)$$

Anhand der Definition von $G_1(K)$ und $G_2(K)$ erkennt man, daß (21.97) mit

$$Y(t) = G_1(K)G_2(K)E(t)$$

identisch ist. Aus Gleichung (21.97) folgt

$$Y(t) = \frac{1,3K^{-1}}{1-1,4K^{-1}+0,33K^{-2}} E(t)$$

und damit die Endgleichung

$$Y(t) - 1,4Y(t-1) + 0,33Y(t-2) = 1,3E(t)$$

Die dritte Reduktionsvorschrift bezieht sich auf die Zusammenfassung von Kreisschaltungen. Sie behauptet die Äquivalenz der Schemata in Abbildung 21.19.

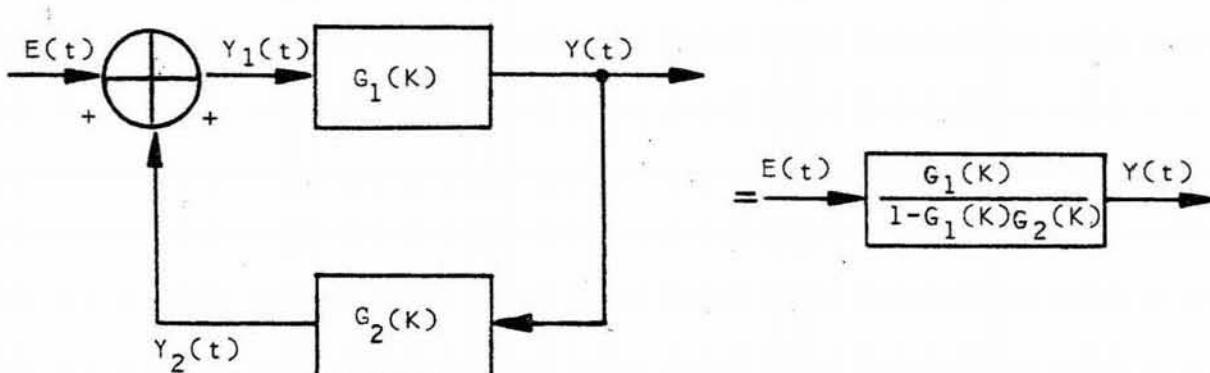


Abb. 21.19 Ermittlung der Übergangsfunktion eines linearen dynamischen Systems im Falle einer Kreisschaltung

Die Zulässigkeit dieses Reduktionsverfahrens soll im Gegensatz zu den vorangegangenen zwei Verfahren in allgemeiner Weise nachgewiesen werden. Die in Abbildung 21.19 dargestellte Kreisschaltung wird durch die Gleichungen

$$Y(t) = G_1(K)Y_1(t) \quad (21.98)$$

$$Y_2(t) = G_2(K)Y(t) \quad (21.99)$$

$$Y_1(t) = E(t) + Y_2(t) \quad (21.100)$$

beschrieben. Mit Gleichung (21.99) in (21.100) folgt

$$Y_1(t) = E(t) + G_2(K)Y(t) \quad (21.101)$$

Gleichung (21.101) in (21.98) liefert dann

$$Y(t) = G_1(K)[E(t) + G_2(K)Y(t)]$$

oder

$$Y(t)[1 - G_1(K)G_2(K)] = G_1(K)E(t)$$

und damit

$$Y(t) = \frac{G_1(K)}{1 - G_1(K)G_2(K)} E(t)$$

d.h. die in Abbildung 21.19 aufgezeigte Übergangsfunktion.

Durch die sukzessive Anwendung dieser drei Reduktionsvorschriften kann die totale Übergangsfunktion und damit zugleich auch die Endgleichung eines Systems auf übersichtliche Art ermittelt werden.

Als Beispiel sei das Blockdiagramm eines MA-Systems angeführt.

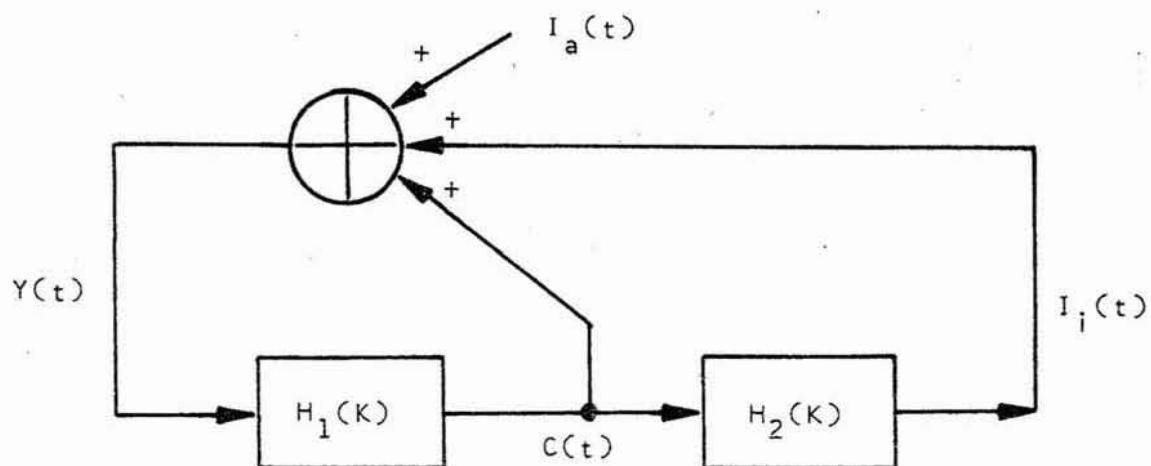


Abb. 21.20 Blockdiagramm eines MA-Systems

$C(t) = H_1(K)Y(t)$ beschreibt die Konsumfunktion und $I_a(t) = H_2(K)C(t)$ die Investitionsfunktion. Ermittelt werden soll die Übergangsfunktion zwischen I_a und Y .

Aufgrund der Vorschrift über die Zusammenfassung von Parallelgliedern folgt:

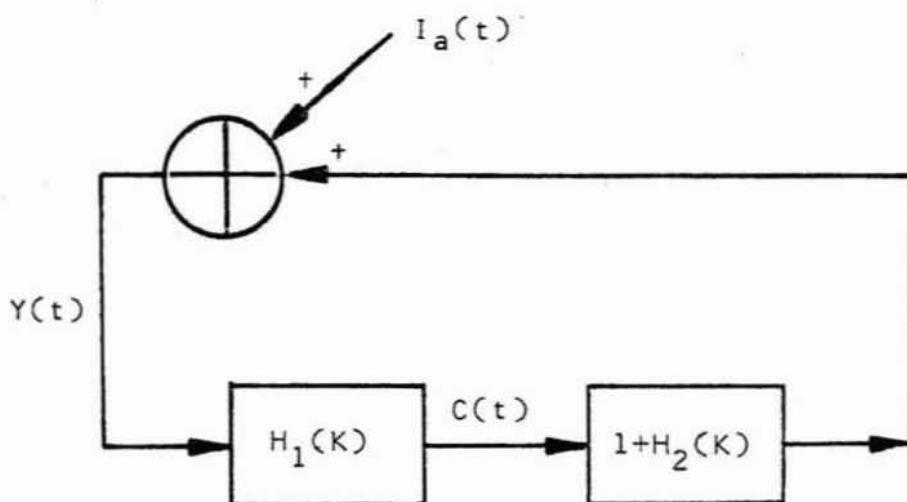


Abb. 21.21 Reduziertes Blockdiagramm eines MA-Systems durch Zusammenfassung von Parallelgliedern

Unter Beachtung der Reduktionsvorschrift für kaskadierende Glieder folgt die weitere Reduktion:

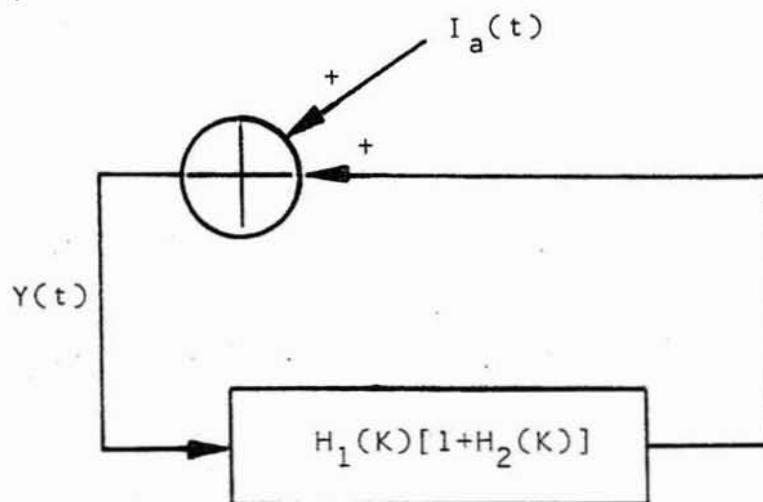


Abb. 21.22 Reduziertes Blockdiagramm eines MA-Systems durch Zusammenfassung kaskadierender Glieder

Schließlich wird die Kreisreduktionsvorschrift angewandt, die zur totalen Übergangsfunktion führt

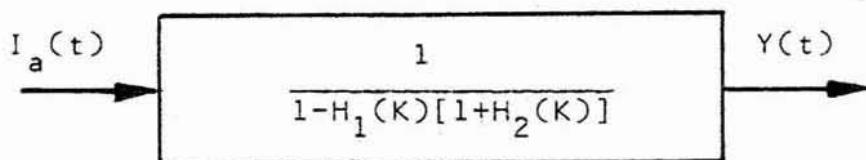


Abb. 21.23 Reduziertes Blockdiagramm eines MA-Systems durch Zusammenfassung von parallelen, kaskadierenden und Kreisgliedern

Da im Falle eines MA-Systems die Operatoren durch

$$H_1(K) = \alpha K^{-1}$$

und

$$H_2(K) = \beta - \beta K^{-1}$$

konkretisiert werden, ergibt sich mit der Einsetzung dieser Ausdrücke für $H_1(K)$ und $H_2(K)$ in die totale Übergangsfunktion die Fassung:

$$H(K) = \frac{1}{1 - (\alpha + \beta)K^{-1} + \alpha\beta K^{-2}}$$

und damit die uns schon bekannte Endgleichung:

$$Y(t) - (\alpha + \beta)Y(t-1) + \alpha\beta Y(t-2) = I_a(t)$$

b) Endgleichungsbestimmung anhand von Signalflußdiagrammen

Am Beispiel eines MA-Systems wurde deutlich, daß die Bestimmung der Übergangsfunktion anhand der Reduktion von Blockdiagrammen sehr mühsam sein kann. Eine andere Methode, die daher bei größeren Systemen zur Anwendung kommt, ist die Reduktion von Signalflußdiagrammen. Signalflußdiagramme liefern wie die Blockdiagramme eine bildhafte Darstellung der Systemzusammenhänge, auf deren Grundlage ein effektiveres Vorgehen bei der schrittweisen Lösung der Operatorengleichungen

bewirkt werden soll.

Der grundsätzliche Aufbau von Signalflußdiagrammen wurde bereits beschrieben. Ihr Unterschied zu einem Blockdiagramm besteht darin, daß die Blöcke in einem Signalflußdiagramm durch gerichtete Strecken ersetzt und die Eingangs- und Ausgangsgrößen in Knotenpunkte überführt werden. Diese Knotenpunkte werden zugleich als Summationspunkte definiert mit der Folge, daß sich der Wert einer Knotenpunktvariablen aus der Summe der auf einen Knotenpunkt führenden Variablen bestimmt.

Den Zusammenhang zwischen der Signalfluß- und Blockdiagrammdarstellung einer Übergangsfunktion zeigt Abbildung 21.24

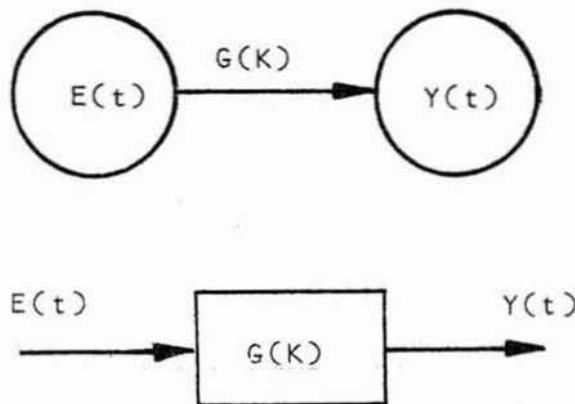


Abb. 21.24 Übergangsfunktion eines linearen dynamischen Systems in einer Signalfluß- und Blockdiagrammdarstellung

Analog zu den Blockdiagrammdarstellungen lassen sich bestimmte Reduktionsvorschriften angeben.

(1) Additionsvorschrift

Parallele Strecken können zu einer Strecke zusammengefaßt werden, deren Übergangsfunktion sich aus der Summe der Übergangsfunktionen der Parallelstrecken ergibt.

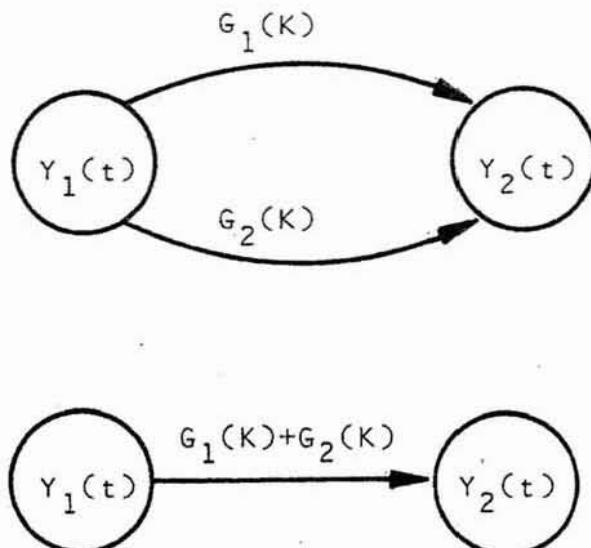
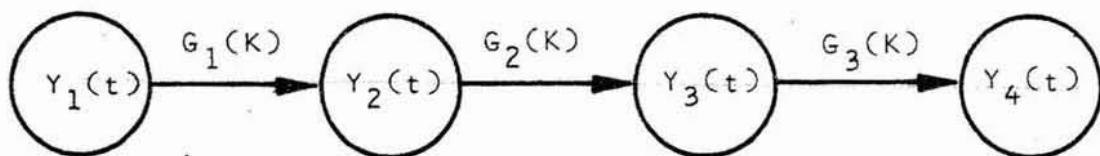


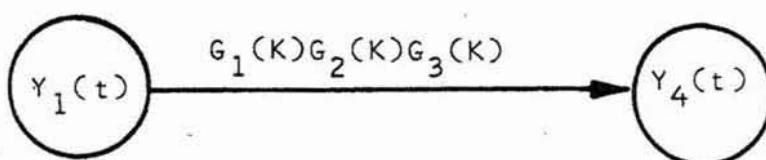
Abb. 21.25 Reduktion von parallelen Gliedern linear dynamischer Systeme in einer Signalflußdiagrammdarstellung

(2) Multiplikationsvorschrift

Eine Kette von Knotenpunkten kann durch eine Strecke ersetzt werden, deren Übergangsfunktion sich aus dem Produkt der Übergangsfunktionen zwischen den Knotenpunkten der Ketten bestimmt.



ist dem Signalflußdiagramm



äquivalent.

Abb. 21.26 Reduktion von kaskadierenden Gliedern linear dynamischer Systeme in einer Signalflußdiagrammdarstellung

(3) Schleifenreduktion

Für die Reduktion einer Schleife gilt die Vorschrift

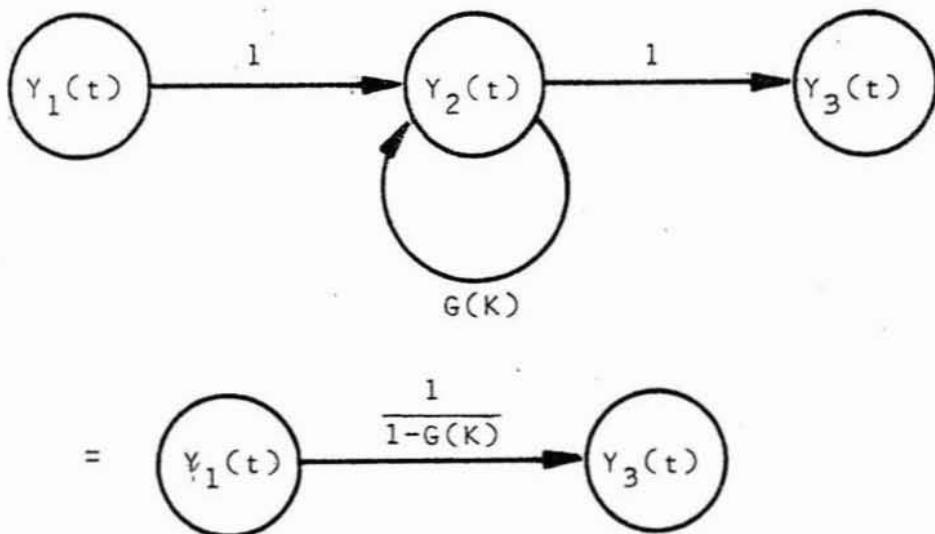


Abb. 21.27 Reduktion von Kreisgliedern linear dynamischer Systeme in einer Signalflußdiagrammdarstellung

Die Gültigkeit dieser Beziehung erkennt man anhand des Gleichungssystems

$$Y_2(t) = Y_1(t) + G(K)Y_2(t) \quad (21.102)$$

$$Y_3(t) = Y_2(t) \quad (21.103)$$

Die Auflösung von Gleichung (21.102) nach $Y_2(t)$ liefert

$$Y_2(t)(1-G(K)) = Y_1(t)$$

und

$$Y_2(t) = Y_1(t)/(1-G(K))$$

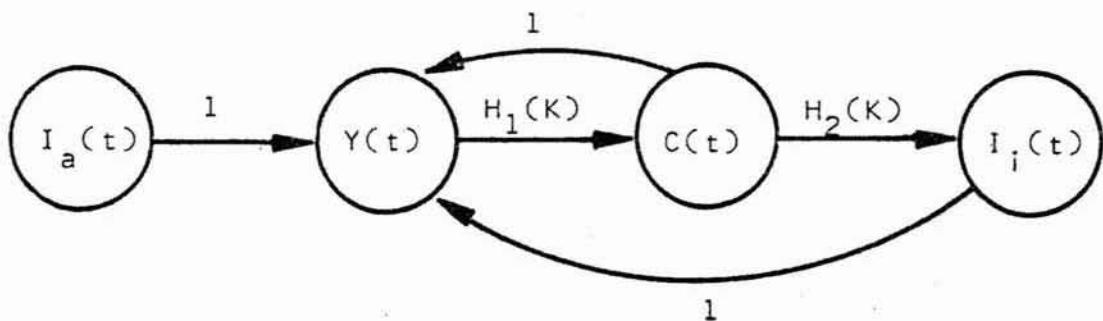
Mit Gleichung (21.103) folgt

$$Y_3(t) = \frac{1}{1-G(K)} Y_1(t)$$

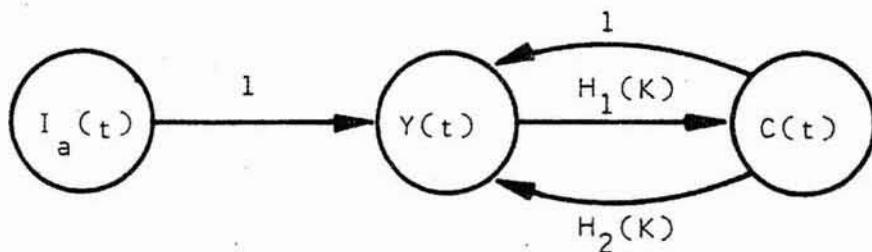
d.h. die Übergangsfunktion zwischen $Y_3(t)$ und $Y_1(t)$.

Die Reduktion von Systemgliedern und damit die Gewinnung der Übergangsfunktion eines Systems soll anhand eines durch ein Signalflußdiagramm beschriebenes MA-Systems demonstriert werden.

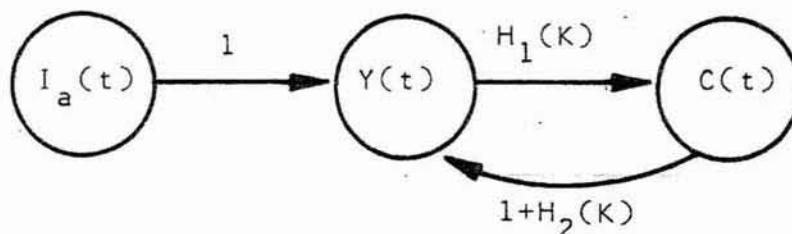
Ein solches Signalflußdiagramm wird unter Verwendung der bereits im Rahmen der Blockdiagrammdarstellung angeführten Operatoren H_1 und H_2 durch die folgende Abbildung dargestellt:



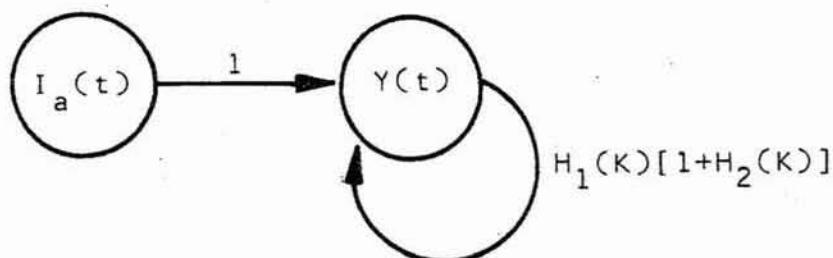
Aufgrund der Multiplikationsvorschrift folgt



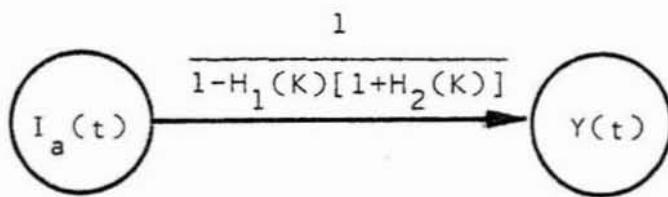
Die obere und untere Schleife von $Y(t)$ über $C(t)$ kann nach der Additionsvorschrift zusammengefaßt werden. Es ergibt sich damit



Nach der Multiplikationsvorschrift kann die Kette $Y(t)-C(t)-Y(t)$ reduziert werden durch



Die Schleifenreduktion führt zu dem nachfolgenden zweigliedrigen Signalflußdiagramm



und damit zu der totalen Übergangsfunktion eines MA-Systems, die bereits mit Hilfe der Blockdiagrammreduktion ermittelt wurde.

cc) Analyse linearer Systeme anhand von Matrizen

Eine umfassende Analyse linearer dynamischer Systeme ist nur mit Hilfe der Matrizenrechnung möglich. In diesem Abschnitt werden sowohl die Grundlagen der Matrizenrechnung erörtert als auch die daran anknüpfenden Verfahren der Analyse linearer Systeme.

a) Grundbegriffe der Matrizenrechnung

Eine Matrix ist ein rechteckiges Schema von Zahlen oder Elementen und wird mit runden oder eckigen Klammern

$$M = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \vdots & \ddots & & \vdots \\ \vdots & & & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nm} \end{bmatrix}$$

versehen. Die horizontalen Reihen werden Zeilen, die vertikalen Reihen Spalten genannt. Die Elemente werden hinsichtlich ihrer Stellung durch Indizes gekennzeichnet. Der erste Index benennt die Zeile, in der sich das betreffende Element befindet, während der zweite Index die Spalte charakterisiert. Das Element a_{ij} steht daher in der i-ten Zeile und j-ten Spalte. In abgekürzter Schreibweise kann man eine Matrix auch durch

$$M = [a_{ij}] \quad \begin{array}{l} i = 1, 2, \dots, n \\ j = 1, 2, \dots, m \end{array}$$

ausdrücken.

Wir betrachten im folgenden einige wichtige Typen von Matrizen:

Die Matrix

$$M = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 3 \end{bmatrix}$$

wird als Diagonalmatrix bezeichnet, da alle nicht auf der Hauptdiagonalen ($i=j$) liegenden Elemente Null werden. Sie ist zugleich eine quadratische Matrix, weil die Anzahl der Zeilen und Spalten übereinstimmt.

Quadratische Diagonalmatrizen werden oft durch folgende Schreibweise dargestellt:

$$M = \begin{bmatrix} a_{11} & & 0 \\ & \ddots & \\ 0 & & a_{mm} \end{bmatrix}$$

Einen Sonderfall unter den Diagonalmatrizen bildet die Einheitsmatrix, welche mit dem Buchstaben I bezeichnet wird. Als Beispiel sei folgende Matrix angeführt:

$$I = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Verallgemeinernd ist eine Einheitsmatrix eine Diagonalmatrix, deren Hauptdiagonale nur mit Einsen besetzt ist.

Eine einspaltige Matrix wird als Spaltenvektor bezeichnet, während eine einzeilige Matrix Zeilenvektor genannt wird.

So ist beispielsweise die 2×1 Matrix $\begin{bmatrix} 2 \\ 3 \end{bmatrix}$ ein Spaltenvektor, während die 1×2 Matrix $[3, 4]$ als Zeilenvektor bezeichnet wird.

Folgende Definitionen und Rechenoperationen sind für die nachfolgenden Betrachtungen von Belang:

a) Gleichheit zweier Matrizen

Zwei $n \times m$ Matrizen $A = [a_{ij}]$ und $B = [b_{ij}]$ sind gleich, wenn für alle i, j gilt: $a_{ij} = b_{ij} \quad i=1, 2, \dots, n \quad j=1, 2, \dots, m$

b) Summe zweier Matrizen

Die Summe zweier $n \times m$ Matrizen A und B bildet eine $n \times m$ Matrix C, die sich aus der Addition der entsprechenden Elemente ergibt, d.h.

$$C = A + B = [c_{ij}]$$

bestimmt sich mit

$$c_{ij} = a_{ij} + b_{ij} \quad i=1,2,\dots,n; \quad j=1,2,\dots,m$$

Beispiel:

$$\begin{bmatrix} 4 & 0 \\ 1 & 2 \end{bmatrix} + \begin{bmatrix} 1 & 6 \\ 2 & 0 \end{bmatrix} = \begin{bmatrix} 5 & 6 \\ 3 & 2 \end{bmatrix}$$

c) Multiplikation einer Matrix mit einer Konstanten

Werden in einer beliebigen Matrix A alle Elemente mit einer beliebigen Zahl c multipliziert, so wird die sich ergebende Matrix mit cA bezeichnet. Für das Rechnen mit Matrizen gelten analog zum Rechnen mit Zahlen die folgenden Sätze:

$$A + B = B + A$$

$$(c_1 + c_2)A = c_1A + c_2A$$

$$c_1(A+B) = c_1A + c_1B$$

d) Multiplikation von Matrizen

Bevor wir uns der Multiplikation von Matrizen zuwenden, sei die Multiplikation eines Zeilen- mit einem Spaltenvektor erklärt. U sei ein Zeilenvektor und W ein Spaltenvektor mit n Komponenten, d.h.

$$U = [u_1, u_2, \dots, u_n]$$

$$W = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix}$$

Das Produkt UW des Zeilenvektors U mit dem Spaltenvektor W ist eine 1×1 Matrix, die damit aus einer Zahl besteht, deren Wert sich nach

$$UW = \sum_{j=1}^n w_j u_j = w_1 u_1 + w_2 u_2 + \dots + w_n u_n \quad (21.104)$$

ermittelt.

Als Beispiel sei das Produkt von zwei Vektoren

$$\begin{bmatrix} 2 \\ 1 \end{bmatrix} [3, -2] = 2 \cdot 3 + 1 \cdot (-2) = 4$$

angeführt.

Nehmen wir an, es sei das Matrizenprodukt der $n \times m$ Matrix A mit der $m \times k$ -Matrix B zu bestimmen.

Die Matrix A kann durch

$$A = \begin{bmatrix} A_1 \\ A_2 \\ \vdots \\ A_n \end{bmatrix} = [a_{ij}] \quad i=1, \dots, n \quad j=1, \dots, m$$

beschrieben werden. A_i ist hierbei ein Zeilenvektor, der die i-te Zeile der Matrix beschreibt, d.h.

$$A_i = [a_{i1}, a_{i2}, \dots, a_{im}]$$

In ähnlicher Weise lässt sich die Matrix B durch

$$B = [B_1, B_2, \dots, B_k] = [b_{ij}] \quad \begin{matrix} (i=1, 2, \dots, m) \\ (j=1, 2, \dots, k) \end{matrix}$$

beschreiben, wobei

$$B_j = \begin{bmatrix} b_{1j} \\ b_{2j} \\ \vdots \\ b_{mj} \end{bmatrix}$$

den Spaltenvektor der j-ten Spalte der Matrix B repräsentiert. Das Vektorprodukt des Zeilenvektors A_i mit dem Spaltenvektor B_j ergibt sich analog (21.104) aus

$$A_i B_j = \sum_{l=1}^m a_{il} b_{lj} = a_{i1} b_{1j} + a_{i2} b_{2j} + \dots + a_{im} b_{mj} \quad (21.105)$$

Das Matrizenprodukt AB wird nunmehr folgendermaßen definiert:

$$AB = \begin{bmatrix} A_1B_1 & A_1B_2 & \dots & A_1B_k \\ A_2B_1 & A_2B_2 & \dots & A_2B_k \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ A_nB_1 & A_nB_2 & \dots & A_nB_k \end{bmatrix} \quad (21.106)$$

Die Matrizen A und B sind, wie man aus dieser Festlegung erkennt, nur dann miteinander multiplizierbar, wenn die Zahl der Spalten von A der Zahl der Zeilen von B entspricht.

Als Beispiel betrachten wir

$$A = \begin{bmatrix} 1 & -3 \\ 2 & 4 \end{bmatrix} \quad B = \begin{bmatrix} 2 & 1 \\ -1 & 3 \end{bmatrix}$$

Entsprechend (21.104) ergeben sich die Elemente A_iB_j der Produktmatrix aus

$$\begin{aligned} A_1B_1 &= 1 \cdot 2 + (-3)(-1) = 5 \\ A_2B_1 &= 2 \cdot 2 + 4(-1) = 0 \\ A_2B_2 &= 2 \cdot 1 + 4 \cdot 3 = 14 \\ A_1B_2 &= 1 \cdot 1 + (-3)3 = -8 \end{aligned}$$

und damit

$$AB = \begin{bmatrix} 5 & -8 \\ 0 & 14 \end{bmatrix}$$

e) Inverse Matrix

In der Zahlenalgebra hat jede Zahl (außer Null) einen Kehrwert. Eine Zahl b ist der Kehrwert von a, wenn $ab=1$ ist. In der Matrixalgebra kann eine dem Kehrwert analoge Matrix definiert werden:

Ist A eine beliebige $n \times n$ -Matrix und läßt sich eine Matrix B finden, daß

$$AB = I$$

wobei I die Einheitsmatrix ist, so wird B als inverse Matrix von A bezeichnet.

Die inverse Matrix A^{-1} bildet mit der adjungierten Matrix A^* und der

Determinante $|A|$ von A die folgende Beziehung

$$A^{-1} = \frac{A^*}{|A|} \quad (21.107)$$

Die Determinante $|A|$ ist eine nach bestimmten Vorschriften aus den Elementen a_{ij} zu ermittelnde Zahl. Die adjungierte Matrix kann auf folgende Weise ermittelt werden:

Streicht man in einer $n \times n$ -Matrix die i-te Zeile und j-te Spalte, so erhält man eine $(n-1) \times (n-1)$ -Matrix, deren Determinante man als Unter-determinante A_{ij} der Determinante $|A|$ bezeichnen kann. Setzt man vor diese Unter-determinante das Vorzeichen $(-1)^{i+j}$, so erhält man bei entsprechender Anordnung der Elemente die Matrix

$$\tilde{A} = [(-1)^{i+j} A_{ij}]$$

Durch Vertauschen der Zeilen und Spalten dieser Matrix erhält man ihre sogenannte transponierte Matrix, welche die adjungierte Matrix A^* von A darstellt, d.h.

$$A^* = \tilde{A}^T = [(-1)^{i+j} A_{ji}]$$

oder

$$A^* = [A_{ij}^*] \quad \text{mit} \quad A_{ij}^* = (-1)^{i+j} A_{ji} \quad (21.108)$$

Die Determinante $|A|$ kann nach dem Entwicklungssatz für Determinanten durch

$$|A| = a_{i1} A_{1j}^* + a_{i2} A_{2j}^* + \dots + a_{in} A_{nj}^* \quad (i=j) \quad (21.109)$$

entwickelt werden. $|A|$ ergibt sich somit aus der Summe des Produktes der jeweils miteinander korrespondierenden Werte der Elemente einer Zeile von A und einer Spalte von A^* . Das Gesagte sei an einem Beispiel verdeutlicht. Die Matrix

$$A = \begin{bmatrix} 1 & 3 & -1 \\ 3 & -2 & 0 \\ 2 & 3 & 2 \end{bmatrix}$$

führt zu den Unter-determinanten

$$\begin{aligned}
 A_{11} &= \begin{vmatrix} -2 & 0 \\ 3 & 2 \end{vmatrix} = -4 & A_{12} &= -\begin{vmatrix} 3 & 0 \\ 2 & 2 \end{vmatrix} = -6 & A_{13} &= \begin{vmatrix} 3 & -2 \\ 2 & 3 \end{vmatrix} = 13 \\
 A_{21} &= -\begin{vmatrix} 3 & -1 \\ 3 & 2 \end{vmatrix} = -9 & A_{22} &= \begin{vmatrix} 1 & -1 \\ 2 & 2 \end{vmatrix} = 4 & A_{23} &= -\begin{vmatrix} 1 & 3 \\ 2 & 3 \end{vmatrix} = 3 \\
 A_{31} &= \begin{vmatrix} 3 & -1 \\ -2 & 0 \end{vmatrix} = -2 & A_{32} &= -\begin{vmatrix} 1 & -1 \\ 3 & 0 \end{vmatrix} = -3 & A_{33} &= \begin{vmatrix} 1 & 3 \\ 3 & -2 \end{vmatrix} = -11
 \end{aligned}$$

Man erhält die Matrix

$$\begin{bmatrix} -4 & -6 & 13 \\ -9 & 4 & 3 \\ -2 & -3 & -11 \end{bmatrix}$$

deren Spiegelung zu der adjungierten Matrix

$$A^* = \begin{bmatrix} -4 & -9 & -2 \\ -6 & 4 & -3 \\ 13 & 3 & -11 \end{bmatrix}$$

führt. Die Determinante liefert nach der ersten Zeile entwickelt:

$$|A| = 1(-4) + 3(-6) - 1 \cdot 13 = -35$$

Im Falle der Entwicklung nach der zweiten Zeile ergibt sich ebenfalls:

$$|A| = 3(-9) - 2 \cdot 4 + 0 \cdot 3 = -35$$

Die inverse Matrix berechnet sich daher unter Beachtung von (21.107)

$$A^{-1} = \begin{bmatrix} \frac{4}{35} & \frac{9}{35} & \frac{2}{35} \\ \frac{6}{35} & -\frac{4}{35} & \frac{3}{35} \\ -\frac{13}{35} & -\frac{3}{35} & \frac{11}{35} \end{bmatrix}$$

Auf die Techniken der Berechnung von Determinanten und inversen Matrizen bei großen Matrizen soll hier nicht eingegangen werden, zumal sie von interessierten Anwendern ohne Schwierigkeiten mit Standard-EDV-Programmen ermittelt werden können. Für die sich anschließenden Betrachtungen ist es wichtig festzuhalten, daß die Berechnung der De-

terminanten adjungierter und inverser Matrizen ausschließlich auf Additionen, Subtraktionen, Multiplikationen und Divisionen beruht. Diese Feststellung ist von Bedeutung, wenn wir uns mit Polynommatrizen beschäftigen. In einer Polynommatrix werden die Elemente einer Matrix nicht durch Zahlen, sondern durch Polynome dargestellt. Da beim Rechnen mit Operatoren Additionen, Subtraktionen und Multiplikationen zugelassen sind, so ist es auch für den Sonderfall einer aus Operatorpolynomen bestehenden Polynommatrix nicht unsinnig, eine Determinante oder inverse Matrix zu ermitteln.

β) Endgleichungsbestimmung anhand von Polynommatrizen

Die Ermittlung der Endgleichungen linearer Systeme wurde bisher mehrmals erörtert und an Beispielen demonstriert. Es handelt sich jedoch um Verfahren, in denen die generelle Systematik einer Endgleichungsbestimmung nicht klar zum Ausdruck kommt.

Ein solcher genereller Ansatz zur Endgleichungsbestimmung soll im folgenden beschrieben werden.

Die Differenzengleichung

$$Y_1(t) + 2Y_1(t-1) - 3Y_1(t-3) = E_1(t)$$

kann bei Verwendung des Rückwärtsoperators $Y(t-\eta)=K^{-\eta}Y(t)$ in die Form

$$(1+2K^{-1}-3K^{-3})Y_1(t) = E_1(t) \quad (21.110)$$

überführt werden.

Definiert man $G(K)=1+2K^{-1}-3K^{-3}$, so wird (21.110) zu

$$G(K)Y_1(t) = E_1(t)$$

Dieselbe abkürzende Schreibweise ist auch bei Differenzengleichungssystemen möglich.

Das Modell

$$\begin{aligned} 2Y_1(t) + 0,4Y_1(t-2) + 3Y_2(t) + 0,25Y_2(t-1) &= E_1(t) \\ 0,5Y_1(t) + 0,1Y_1(t-1) + 2Y_2(t) - 0,1Y_2(t-1) &= E_2(t) \end{aligned}$$

kann in die Form

$$P_{11}(K)Y_1(t) + P_{12}(K)Y_2(t) = E_1(t)$$

mit

$$\begin{array}{ll} P_{11}(K) = 2 + 0,4K^{-2} & P_{12}(K) = 3 + 0,25K^{-1} \\ P_{21}(K) = 0,5 + 0,1K^{-1} & P_{22}(K) = 2 - 0,1K^{-1} \end{array}$$

überführt werden.

Ein beliebiges Differenzengleichungssystem kann daher in der folgenden allgemeinen Form beschrieben werden:

$$\begin{aligned} P_{11}(K)Y_1(t) + P_{12}(K)Y_2(t) + \dots + P_{1m}(K)Y_m(t) &= E_1(t) \\ P_{21}(K)Y_1(t) + P_{22}(K)Y_2(t) + \dots + P_{2m}(K)Y_m(t) &= E_2(t) \\ \vdots & \\ P_{m1}(K)Y_1(t) + P_{m2}(K)Y_2(t) + \dots + P_{mm}(K)Y_m(t) &= E_m(t) \end{aligned} \quad (21.111)$$

Diese Form eines Differenzengleichungssystems legt den Übergang zu einer Matrzenschreibweise nahe. Definiert man:

$$P(K) = \begin{bmatrix} P_{11}(K) & P_{12}(K) & \dots & P_{1m}(K) \\ P_{21}(K) & P_{22}(K) & \dots & P_{2m}(K) \\ \dots & \dots & \dots & \dots \\ P_{m1}(K) & P_{m2}(K) & \dots & P_{mm}(K) \end{bmatrix}$$

und

$$Y(t) = \begin{bmatrix} Y_1(t) \\ Y_2(t) \\ \vdots \\ Y_m(t) \end{bmatrix} \quad E(t) = \begin{bmatrix} E_1(t) \\ E_2(t) \\ \vdots \\ E_m(t) \end{bmatrix}$$

dann läßt sich das Gleichungssystem (21.111) durch

$$P(K)Y(t) = E(t) \quad (21.112)$$

ausdrücken. Die Matrix P ist eine Polynommatrix, da ihre Elemente keine Zahlen, sondern Polynome bilden. Die Operationsregeln der Matrizentheorie gelten jedoch, wie erwähnt, unverändert. Das Gleichungssystem (21.112) ist daher immer dann in seine Endgleichungsform überführbar, wenn die Polynommatrix P eine inverse Matrix P^{-1} besitzt. Entsprechend (21.107) besitzt eine Matrix immer dann eine Inverse,

wenn ihre Determinante ungleich Null ist. Die Bedingung zur Ermittlung eines Endgleichungssystems lautet daher

$$|P(K)| \neq 0$$

In diesem Fall wird aus (21.112) durch Multiplikation mit $P^{-1}(K)$

$$P^{-1}(K)P(K)Y(t) = P^{-1}(K)E(t)$$

und damit

$$Y(t) = P^{-1}(K)E(t) \quad (21.113)$$

Mit (21.107) wird (21.113)

$$Y(t) = \frac{P^*(K)E(t)}{|P(K)|}$$

und damit ergibt sich die Endgleichung

$$|P(K)|Y(t) = P^*(K)E(t) \quad (21.114)$$

Das Ermittlungsverfahren sei an einem Beispiel demonstriert. Gegeben sei das Differenzengleichungssystem:

$$\begin{aligned} 1,5Y_1(t) + 0,4Y_1(t-1) + Y_2(t) &= E_1(t) \\ 3Y_1(t) + 0,2Y_1(t-1) + 1,1Y_2(t) + 0,3Y_2(t-1) &= E_2(t) \end{aligned}$$

Die Operatorenpolynome sind:

$$\begin{aligned} P_{11}(K) &= 1,5 + 0,4K^{-1} & P_{12}(K) &= 1 \\ P_{21}(K) &= 3 + 0,2K^{-1} & P_{22}(K) &= 1,1 + 0,3K^{-1} \end{aligned}$$

Die Polynommatrix P bestimmt sich nach

$$P(K) = \begin{bmatrix} P_{11}(K) & P_{12}(K) \\ P_{21}(K) & P_{22}(K) \end{bmatrix}$$

Ihre Determinante wird durch

$$|P(K)| = (1,5+0,4K^{-1})(1,1+0,3K^{-1}) - (3+0,2K^{-1})$$

oder

$$|P(K)| = -1,35 + 0,69K^{-1} + 0,12K^{-2}$$

beschrieben.

Die Adjunkte $P^*(K)$ bestimmt sich aufgrund (21.108)

$$P^*(K) = \begin{bmatrix} 1,1 + 0,3K^{-1} & -3 - 0,2K^{-1} \\ -1 & 1,5 + 0,4K^{-1} \end{bmatrix}$$

Die Endgleichungsform erhalten wir mit (21.114)

$$(-1,35 + 0,69K^{-1} + 0,12K^{-2}) \begin{bmatrix} Y_1(t) \\ Y_2(t) \end{bmatrix} = \begin{bmatrix} 1,1 + 0,3K^{-1} & -3 - 0,2K^{-1} \\ -1 & 1,5 + 0,4K^{-1} \end{bmatrix} \begin{bmatrix} E_1(t) \\ E_2(t) \end{bmatrix}$$

Durch Rücktransformation des Operators erhält man in der üblichen Differenzengleichungsdarstellung die Endgleichung für $Y_1(t)$:

$$-1,35Y_1(t) + 0,69Y_1(t-1) + 0,12Y_1(t-2) = 1,1E_1(t) + 0,3E_1(t-1) - 3E_2(t) - 0,2E_2(t-1)$$

und für $Y_2(t)$

$$-1,35Y_2(t) + 0,69Y_2(t-1) + 0,12Y_2(t-2) = -E_1(t) + 1,5E_2(t) + 0,4E_2(t-1)$$

Die Division der beiden Gleichungen durch -1,35 führt zu der üblichen Standardform einer Endgleichung.

Die Tatsache, daß entsprechend Gleichung (21.114) der Spaltenvektor $Y(t)$ stets mit demselben Operatorpolynom $|P(K)|$ multipliziert wird, führt zu dem Schluß, daß die Endgleichungen der endogenen Variablen eines Differenzengleichungssystems stets dieselbe reduzierte Form

$$|P(K)|Y_v(t) = 0 \quad v=1,2,\dots,m$$

besitzen. Daraus könnte man den Schluß ziehen, daß die Funktionslösungen der endogenen Variablen eines Differenzengleichungssystems stets dieselben Wurzeln aufweisen und daher ein einheitliches dynamisches Verhalten zustande kommt. Dies ist jedoch nicht generell der Fall. Es soll nur kurz gezeigt werden, warum die oben geschilderte Schlußfolgerung falsch ist.

Eine endogene Variable besitzt nicht nur eine, sondern im Prinzip unendlich viele Endgleichungen. Betrachten wir die Endgleichung unseres MA-Modells

$$Y(t) - 0,3Y(t-1) + 0,2Y(t-2) = I_a(t)$$

deren charakteristische Gleichung die Wurzeln

$$\lambda_1 = 0,15 + 0,4213i \quad \text{und} \quad \lambda_2 = 0,15 - 0,4213i$$

besitzt, und schreiben wir diese Gleichung in Operatorenform, d.h.

$$Y(t)(1-0,3K^{-1}+0,2K^{-2}) = I_a(t)$$

dann ist es möglich, diese Gleichung mit einem beliebigen Operatorenpolynom zu multiplizieren, um damit eine weitere Endgleichungsform zu erhalten. Multiplizieren wir unsere Ausgangsgleichung beispielsweise mit $(K^{-1}-2)$, so erhalten wir:

$$Y(t)(-2+1,6K^{-1}-0,7K^{-2}+0,2K^{-3}) = (K^{-1}-2)I_a(t)$$

oder die Endgleichung

$$-2Y(t) + 1,6Y(t-1) - 0,7Y(t-2) + 0,2Y(t-3) = I_a(t-1) - 2I_a(t)$$

mit den Wurzeln

$$\lambda_1 = 0,15 + 0,4213i \quad \lambda_2 = 0,15 - 0,4213i \quad \text{und} \quad \lambda_3 = 2$$

Orientiert man sich an dieser Endgleichung, so würde das System durch drei Wurzeln bestimmt und einen gegenüber der ursprünglichen Gleichung völlig anderen Verlauf aufweisen: Das System würde wegen der Wurzel $\lambda_3=2$ explodieren, die in der Funktionslösung den Ausdruck $C_3 2^t$ bilden würde. Es zeigt sich, daß man durch Multiplikation des Ausgangspolynoms mit einem Ausdruck $(K^{-1}-a)$ beliebig viele Endgleichungen mit den unterschiedlichsten Wurzeln erhält. Es liegt die Frage nahe, welche der Wurzeln unter diesen Umständen das dynamische Verhalten des Systems bestimmen.

Um das zu klären, betrachten wir eine Endgleichung in Operatorschreibweise:

$$Y(t)(1+a_1K^{-1}t+\dots+a_nK^{-n}) = (g_0+g_1K^{-1}+\dots+g_sK^{-s})E(t)$$

Unterstellen wir $n > m$ und multiplizieren mit K^n , so erhalten wir:

$$Y(t)(K^n+a_1K^{n-1}+\dots+a_n) = (g_0K^n+g_1K^{n-1}+\dots+g_sK^{n-s})E(t) \quad (21.114)$$

Es gilt der Produktensatz für Polynome:

Sind $\lambda_1, \lambda_2, \dots, \lambda_n$ die Wurzeln des Polynoms $b_0K^n+b_1K^{n-1}+\dots+b_nK^0$, so

läßt sich dieses Polynom durch

$$(K-\lambda_1)(K-\lambda_2)\dots(K-\lambda_n) = b_0 K^n + b_1 K^{n-1} + \dots + b_n K^0 \quad (21.115)$$

darstellen.

Verwendet man diese Produktendarstellung für die Polynome in Gleichung (21.114), so erhält man

$$Y(t)[(K-\lambda_1)(K-\lambda_2)\dots(K-\lambda_n)] = [(K-\delta_1)(K-\delta_2)\dots(K-\delta_n)]E(t) \quad (21.116)$$

mit λ und δ als Wurzeln. Stimmen zwei Wurzeln aus den beiden Polynomen miteinander überein, so kann die Gleichung durch $(K-\lambda_*)$ geteilt werden, wenn λ_* diese gemeinsame Wurzel ist. Man erkennt, daß alle gemeinsamen Wurzeln der beiden Polynome das Systemverhalten nicht beeinflussen. Entscheidend für eine Verhaltensanalyse ist daher nur die Endgleichung, in der die Wurzeln ihrer Operatorpolynome verschieden sind. Von dieser Voraussetzung wurde bisher stillschweigend ausgegangen. Im Falle der Endgleichungsform (21.114) ist jedoch keine Garantie gegeben, daß die Wurzeln der charakteristischen Gleichung von

$$|P(K)| Y_v(t) = 0 \quad v=1,2,\dots,m$$

und den Operatorpolynomen der exogenen Variablen verschieden sind. Enthält eine Endgleichung daher eine exogene Variable mit unterschiedlichen Verzögerungen, so sind die Wurzeln des entsprechenden Operatorpolynoms mit den Wurzeln der charakteristischen Gleichung zu vergleichen. Erst wenn sich diese als verschieden erweisen, ist es zulässig, anhand der Wurzeln das dynamische Verhalten des Systems zu beurteilen.

γ) Zustandsraumdarstellung linearer Systeme und ihre Analysemethoden

Unsere bisherigen Betrachtungen basierten auf dem Konzept der Endgleichungsanalyse. Das dynamische Verhalten eines Systems wird in diesem Fall bezüglich bestimmter endogener Variablen anhand ihrer Endgleichungen beurteilt. Diese Untersuchungsmethode ist besonders dann angebracht, wenn man nur an dem dynamischen Verhalten einer ganz bestimmten endogenen Variablen interessiert ist und damit eine 'ver-

dichtete Aussage' über die Beziehungen zwischen dieser endogenen Variablen und den exogenen Größen gewinnen will.

Im folgenden wollen wir uns mit einer anderen Methode der Analyse dynamischer Systeme beschäftigen, die auf der Zustandsraumdarstellung eines linearen Systems beruht.

Jedes lineare dynamische System läßt sich durch ein äquivalentes Differenzengleichungssystem ersten Grades beschreiben. Die um eine Periode verzögerten endogenen Variablen einer solchen Darstellung können als ein Zustandsvektor $Z(t-1)$ angesehen werden, der zusammen mit den exogenen Variablen den Zustandsvektor $Z(t)$ bestimmt. Angesichts dieser Interpretationsmöglichkeit spricht man in diesem Fall von der Zustandsraumdarstellung linearer Systeme.

Diese Darstellungsform erlaubt eine übersichtliche Beurteilung der Abhängigkeiten der endogenen Variablen eines Systems. Die Zustandsraumdarstellung eines dynamischen Systems in Matrixform erhöht aber nicht nur die Übersichtlichkeit, sondern gestattet auch eine einfache numerische Analyse im Falle der Anwendung von EDV-Anlagen. Wie erwähnt, geht das heute viel verwendete Modellierungskonzept System Dynamics von der Grundkonzeption aus, daß die Welt durch Beziehungen zu beschreiben ist, die zum Ausdruck bringen, in welcher Weise ein System von bestimmten Systemzuständen in Periode $t-1$ in die Zustände in Periode t übergeht. Demzufolge bilden System-Dynamics-Modelle in ihrem Primäransatz stets ein System von Differenzengleichungen ersten Grades. Diese Konzeption soll zwar hier nicht besprochen werden, doch mag dieser Hinweis genügen, um deutlich zu machen, daß die bei einer Zustandsraumdarstellung eines primären Modellansatzes einzuführenden zusätzlichen 'künstlichen' endogenen Variablen durchaus einer empirisch sinnvollen Interpretation zugänglich sein können. Zur Einschätzung dieser Darstellungsform dürfte es auch von Interesse sein, daß die zur Analyse von Volkswirtschaften verwendeten dynamischen Input-Output-Modelle im Primäransatz bereits einer Zustandsraumdarstellung entsprechen.

Nach einer Beschreibung der Zustandsraumdarstellung wenden wir uns

den Methoden zu, mit denen primäre Modellansätze in eine Zustandsraumdarstellung überführt werden können. Daran anschließend werden bestimmte Analysemethoden im Rahmen der Zustandsraumdarstellung erörtert.

Als erstes wollen wir versuchen, ein allgemeines MA-Modell in einer Zustandsraumdarstellung zu formulieren.

Als Grundlage dient das Gleichungssystem

$$Y(t) = C(t) + I_i(t) + I_a(t) \quad (21.117)$$

$$I_i(t) = \beta[C(t) - C(t-1)] \quad (21.118)$$

$$C(t) = \alpha Y(t-1) \quad (21.119)$$

Setzt man Gleichung (21.119) in (21.118) und (21.117) ein, so erhält man

$$Y(t) = \alpha Y(t-1) + I_i(t) + I_a(t) \quad (21.120)$$

$$I_i(t) = \beta[\alpha Y(t-1) - C(t-1)] \quad (21.121)$$

$$C(t) = \alpha Y(t-1)$$

Mit Gleichung (21.121) in (21.120) folgt:

$$Y(t) = (\alpha + \alpha\beta)Y(t-1) - \beta C(t-1) + I_a(t)$$

$$I_i(t) = \alpha\beta Y(t-1) - \beta C(t-1)$$

$$C(t) = \alpha Y(t-1)$$

und in Matrizenform

$$\begin{bmatrix} Y(t) \\ I_i(t) \\ C(t) \end{bmatrix} = \begin{bmatrix} \alpha + \alpha\beta & 0 & -\beta \\ \alpha\beta & 0 & -\beta \\ \alpha & 0 & 0 \end{bmatrix} \begin{bmatrix} Y(t-1) \\ I_i(t-1) \\ C(t-1) \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} I_a(t) \quad (21.122)$$

In diesem Fall erübrigt sich die Einfügung einer künstlichen endogenen Variablen, d.h. alle endogenen Variablen entstammen dem Primäransatz. Als weiteres Beispiel soll das Gleichungssystem

$$Y_1(t) + 3Y_1(t-1) - 6Y_1(t-2) - 2Y_1(t-3) = E_1(t) \quad (21.123)$$

$$Y_2(t) - 2Y_2(t-1) + 3Y_2(t-1) = E_2(t) \quad (21.124)$$

in eine Zustandsraumdarstellung überführt werden.

Definieren wir

$$\begin{aligned} z_1(t) &= y_1(t) \\ z_2(t) &= y_2(t) \\ z_3(t) &= z_1(t-1) \\ z_4(t) &= z_3(t-1) \end{aligned}$$

und setzen diese Definitionen in Gleichungen (21.123) und (21.124) ein, dann erhalten wir:

$$\begin{aligned} z_1(t) + 3z_1(t-1) - 6z_3(t-1) - 2z_4(t-1) &= e_1(t) \\ z_2(t) - 2z_2(t-1) + 3z_1(t-1) &= e_2(t) \end{aligned}$$

In Matrizedarstellung ergibt sich:

$$\begin{bmatrix} z_1(t) \\ z_2(t) \\ z_3(t) \\ z_4(t) \end{bmatrix} = \begin{bmatrix} -3 & 0 & 6 & 2 \\ -3 & 2 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} z_1(t-1) \\ z_2(t-1) \\ z_3(t-1) \\ z_4(t-1) \end{bmatrix} + \begin{bmatrix} e_1(t) \\ e_2(t) \\ 0 \\ 0 \end{bmatrix}$$

Da uns im Prinzip nur die Beobachtungsgrößen y_1 und y_2 interessieren, definieren wir die Matrizengleichung

$$\begin{bmatrix} y_1(t) \\ y_2(t) \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} z_1(t) \\ z_2(t) \\ z_3(t) \\ z_4(t) \end{bmatrix}$$

Das angeführte Beispiel lässt anhand Abbildung 21.28 folgende Interpretation zu:

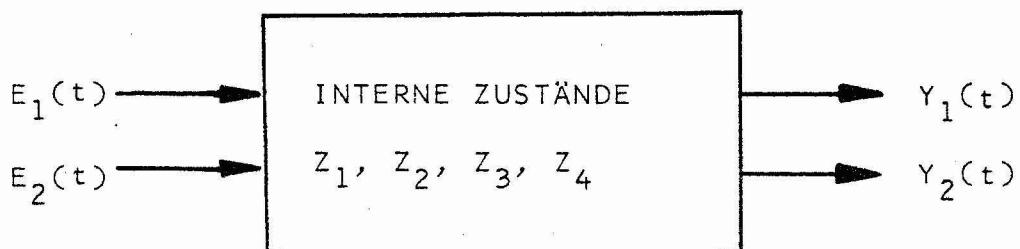


Abb. 21.28 Zustandsraumdarstellung eines dynamischen Systems

Man kann davon sprechen, daß ein System mit den Eingängen E_1 und E_2 über die internen Systemzustände Z_1 bis Z_4 die Ausgangsvariablen Y_1 und Y_2 bestimmt.

Definiert man

$$Z(t) = \begin{bmatrix} Z_1(t) \\ Z_2(t) \\ Z_3(t) \\ Z_4(t) \end{bmatrix} \quad E(t) = \begin{bmatrix} E_1(t) \\ E_2(t) \\ 0 \\ 0 \end{bmatrix}$$

$$M = \begin{bmatrix} -3 & 0 & 6 & 2 \\ -3 & 2 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \quad Y(t) = \begin{bmatrix} Y_1(t) \\ Y_2(t) \end{bmatrix}$$

und

$$B = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

so kann das angeführte Beispiel durch die Matrizengleichungen

$$\begin{aligned} Z(t) &= MZ(t-1) + E(t) \\ Y(t) &= BZ(t) \end{aligned} \tag{21.125}$$

beschrieben werden. Die Matrix M wird als Zustandsmatrix bezeichnet, $Z(t)$ als Zustandsvektor, $E(t)$ und $Y(t)$ als Eingangs- bzw. Ausgangsvektor und B als Beobachtungsmatrix. Verallgemeinernd kann das Gleichungssystem als die allgemeine Form der Zustandsraumsbeschreibung eines dynamischen Systems angesehen werden. Im folgenden wollen wir uns intensiver mit der Struktur der Zustandsmatrix M beschäftigen, welche alle Informationen über das dynamische Verhalten des Systems enthält. Dies erkennen wir bei dem Versuch, anhand von Gleichung (21.125) die Entwicklung des Zeitverlaufes $Z(t)$ zu bestimmen.

$$Z(1) = MZ(0) + E(1)$$

$$Z(2) = MZ(1) + E(2) = M^2Z(0) + ME(1) + E(2)$$

$$Z(3) = MZ(2) + E(3) = M^3Z(0) + M^2E(1) + ME(2) + E(3)$$

⋮

$$Z(t) = MZ(t-1) + E(t) = M^tZ(0) + M^{t-1}E(1) + M^{t-2}E(2) + \dots + M^0E(t)$$

Es wird klar, daß die Stabilität oder Instabilität des Systems bestimmt wird von der zeitlichen Entwicklung der Potenzmatrix M^t . Um die Struktur dieser Potenzmatrix besser beurteilen zu können, sind einige neue Begriffe und Sätze der Matrizentheorie einzuführen.

Als charakteristische Gleichung einer Matrix M bezeichnet man den Ausdruck

$$|M - \lambda I| = 0 \quad (21.126)$$

Handelt es sich um eine ($n \times n$)-Matrix, so ist die Gleichung (21.126) ein Polynom n -ten Grades. Angenommen, M sei durch

$$M = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -3 & 1 & 3 \end{bmatrix}$$

konkretisiert, dann wird

$$M - \lambda I = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -3 & 1 & 3 \end{bmatrix} - \begin{bmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{bmatrix} = \begin{bmatrix} -\lambda & 1 & 0 \\ 0 & -\lambda & 1 \\ -3 & 1 & 3-\lambda \end{bmatrix}$$

welches zu der folgenden Determinante führt

$$|M - \lambda I| = \lambda^3 - 3\lambda^2 - \lambda + 3$$

Es gilt der Satz:¹²

Sind die Wurzeln $\lambda_1, \lambda_2, \dots, \lambda_n$ der charakteristischen Gleichung einer Matrix M verschieden, so läßt sich eine Matrix S finden, die die Beziehung

$$S \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix} S^{-1} = M \quad (21.127)$$

erfüllt. Bezeichnen wir die Diagonalmatrix in (21.127) mit A , d.h.

$$A = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix}$$

¹² Zum Beweis siehe [227, S.167]

dann kann die Potenzmatrix von M in folgender Weise ermittelt werden:

$$\begin{aligned} M^1 &= SAS^{-1} \\ M^2 &= (SAS^{-1})(SAS^{-1}) = SA^2S^{-1} \\ M^3 &= (SAS^{-1})(SA^2S^{-1}) = SA^3S^{-1} \\ &\vdots \\ M^t &= (SAS^{-1})(SA^{t-1}S^{-1}) = SA^tS^{-1} \end{aligned} \quad (21.128)$$

Für die Berechnung der Potenzmatrix gilt, wie man leicht nachprüfen kann, der Satz:

Die Potenzmatrix A^t einer Diagonalmatrix

$$A = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \ddots & \lambda_n \end{bmatrix}$$

ist die Diagonalmatrix

$$A^t = \begin{bmatrix} \lambda_1^t & & 0 \\ & \ddots & \\ 0 & & \ddots & \lambda_n^t \end{bmatrix}$$

Unsere ursprüngliche Potenzmatrix M^t lässt sich daher durch

$$M^t = S \begin{bmatrix} \lambda_1^t & & 0 \\ & \ddots & \\ 0 & & \ddots & \lambda_n^t \end{bmatrix} S^{-1} \quad (21.129)$$

beschreiben. Zur vollständigen Beurteilung dieser Beziehung ist es notwendig, die Gestalt der Matrix S näher zu spezifizieren.

Sie wird als Modalmatrix von M bezeichnet und lässt sich in n Spaltenvektoren $S = [S_1, S_2, \dots, S_n]$ zergliedern, die als Eigenvektoren bezeichnet werden.

Sind $\lambda_1, \lambda_2, \dots, \lambda_n$ die (verschiedenen) Wurzeln der charakteristischen Gleichung der Matrix M, so bestimmen sich die Eigenvektoren S_1, S_2, \dots, S_n der Modalmatrix von M nach der Beziehung¹³

$$MS_i = \lambda_i S_i \quad i=1, 2, \dots, n \quad (21.130)$$

¹³ Zum Beweis siehe [227, S. 170]

Der Satz sei am Beispiel der folgenden Matrix demonstriert,

$$M = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -3 & 1 & 3 \end{bmatrix}$$

deren charakteristische Gleichung wir bereits auf Seite 254 mit

$$\lambda^3 - 3\lambda^2 - 1\lambda + 3 = 0$$

ermittelt haben. Die Wurzeln dieser Gleichung sind $\lambda_1=1$, $\lambda_2=-1$, $\lambda_3=3$. Entsprechend dem oben angeführten Satz gilt für den Eigenvektor S_1

$$MS_1 = \lambda_1 S_1$$

und stärker spezifiziert

$$\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -3 & 1 & 3 \end{bmatrix} \begin{bmatrix} S_{11} \\ S_{12} \\ S_{13} \end{bmatrix} = 1 \begin{bmatrix} S_{11} \\ S_{12} \\ S_{13} \end{bmatrix}$$

oder

$$\begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \\ -3 & 1 & 2 \end{bmatrix} \begin{bmatrix} S_{11} \\ S_{12} \\ S_{13} \end{bmatrix} = 0$$

Die Lösung dieses homogenen Gleichungssystems liefert

$$S_{11} = 1 \quad S_{12} = 1 \quad S_{13} = 1.$$

Die analoge Berechnung mit λ_2 und λ_3 liefert die Matrix

$$S = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 3 \\ 1 & 1 & 9 \end{bmatrix}$$

Ihre Inverse bestimmt sich durch

$$S^{-1} = \begin{bmatrix} 3/4 & 1/2 & -1/4 \\ 3/8 & -1/2 & 1/8 \\ -1/8 & 0 & 1/8 \end{bmatrix}$$

Damit kann gemäß (21.128) die Potenzmatrix M^t durch

$$M^t = \begin{bmatrix} 1 & -2 & 2 \\ 1 & 2 & 1 \\ 2 & 1 & 2 \end{bmatrix}^t = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 3 \\ 1 & 1 & 9 \end{bmatrix} \begin{bmatrix} 3^t & 0 \\ 1^t & 1^t \\ 0 & 1^t \end{bmatrix} \begin{bmatrix} 3/4 & 1/2 & -1/4 \\ 3/8 & -1/2 & 1/8 \\ -1/8 & 0 & 1/8 \end{bmatrix}$$

ausgedrückt werden.

Da die Modalmatrix S nur Elemente enthält, die nicht von t abhängen, wird deutlich, daß das dynamische Verhalten eines Systems ausschließlich durch die Wurzeln λ_1 bis λ_n der charakteristischen Gleichung in (21.126) bestimmt wird.

Im Rahmen der Analyse dynamischer Systeme beschränkt man sich bei großen Systemen oft nur darauf, die Zustandsmatrix M eines Systems zu entwickeln und mit Hilfe von EDV-Standardprogrammen die Wurzeln $\lambda_1, \lambda_2, \dots, \lambda_n$ der charakteristischen Gleichung von $|M-\lambda I|$ zu ermitteln. Da die Ermittlung der Wurzeln einer Matrix ein klassisches Problem der numerischen Mathematik darstellt, stehen eine Reihe von EDV-Programmen zur Verfügung, die auch zur Bestimmung der Wurzeln großer Matrizen verwendet werden können. So berechnete beispielsweise HOWERY mit Hilfe eines EDV-Programmes die Wurzeln einer (56x56)-Zustandsmatrix eines dynamischen Modells der US-Wirtschaft [92, S. 654].

2.1.2. Nichtlineare Modellformen

Nach der Beschreibung einiger in den Wirtschafts- und Sozialwissenschaften verwendeter Typen nichtlinearer Hypothesen wird ein kurzer Überblick über die einschlägigen Verfahren der deduktiven Implikationenaufdeckung nichtlinearer Modelle gegeben. Anhand der Phasendiagrammdarstellung nichtlinearer Modelle ersten Grades werden die von linearen Modellen abweichenden Verhaltensweisen exemplarisch demonstriert. Schließlich wird eine Kenngröße zur Charakterisierung nichtlinearer Modelle diskutiert.

A. Begriffliche Klärung und empirische Interpretation

Der Formenreichtum nichtlinearer dynamischer Modelle ist so groß, daß sich die formale Struktur eines nichtlinearen dynamischen Modells allein als negative Abgrenzung eines linearen dynamischen Modells definieren läßt. Ein großer Teil der bisher entwickelten dynamisch-ökonomischen Modelle ist linear. Geht man von der intuitiv einleuchtenden Feststellung aus, daß die 'reale Welt' nichtlinear sei, so liegt es nahe, die empirische Relevanz linearer Modelle anzuzweifeln.

Die Gründe, welche dazu führten, daß nicht ausschließlich nichtlineare dynamische Modelle Anwendung finden, liegen zum einen in dem Mangel an geeigneten analytischen Deduktionsmethoden zur Erschließung der Modellimplikationen nichtlinearer Zusammenhänge. Zum anderen sind die zur Entwicklung ökonometrischer Modelle notwendigen Parameterschätztechniken für nichtlineare Modelle weniger weit entwickelt.

In der reinen ökonomischen Theorie sind nichtlineare Modelle zum Beispiel von HICKS, GOODWIN, KALDOR, KRELLE und THALBERG entwickelt worden. [82], [70], [106], [118], [200]. Auch fast alle größeren ökonometrischen Modelle enthalten Nichtlinearitäten, sind jedoch in ihrem Grundcharakter linear. So verwendet TINBERGEN sogenannte Schwellenvariable als nichtlineare Modellglieder. Die Modelle von KLEIN, KLEIN-GOLDBÄRGER und CHRIST sind bis auf die Verwendung sogenannter 'compound variables' linear. [205], [206], [111], [110], [31].

Auf beide Arten von Nichtlinearitäten wird im folgenden eingegangen.

Für die Beschreibung mikroökonomischer Zusammenhänge ist die Verwendung nichtlinearer Modelle von besonderer Bedeutung. Wie TINBERGEN bemerkt, ist die Annäherung nichtlinearer Verläufe durch lineare Funktionen bei Gleichungen mit aggregierten Größen eher möglich als bei Gleichungen mit disaggregierten Größen. Denn die nichtlinearen Verläufe ökonomischer Beziehungen geringer Aggregation kompensieren sich in der Regel durch den Aggregationsprozeß. [207]. Da mikroökonomische Beziehungen ex definitione eine geringe Aggregation aufwei-

sen, und ein derartiger Linearisierungseffekt daher nicht zum Tragen kommt, erfordert ihre wirklichkeitsnahe Beschreibung auch nicht-lineare Modelle. Die Entwicklung von realitätskonformen Branchen- oder Firmenmodellen verlangt daher oft in besonders starkem Maße die Verwendung nichtlinearer Hypothesen.

Wie im Falle linearer Modelle bestehen nichtlineare Modelle im Normalfall aus einem System von Gleichungen, die die einzelnen empirischen Zusammenhänge beschreiben. Auch hier kann man eine bestimmte relevante endogene Variable des Modells auswählen und versuchen, durch algebraische Umformungen ihre nichtlineare Endgleichung abzuleiten. Dieses Verfahren wird jedoch nur dann sinnvoll sein, wenn analytische Techniken zur Verfügung stehen, um eine derartige nichtlineare Endgleichung zu lösen oder auch nur ihr Stabilitätsverhalten zu beurteilen. In der Regel ist es aber selten möglich, aus einem nichtlinearen Gleichungssystem überhaupt die Endgleichung einer bestimmten endogenen Variablen abzuleiten.

Im folgenden sollen einige wichtige Typen primärer nichtlinearer Hypothesen und Definitionsgleichungen erörtert werden.

(1) Strukturgleichungen, welche Compound-Variablen erklären

Als Compound-Variable bezeichnet man eine endogene Variable, die sich aus der Multiplikation oder Division zweier erklärender Variablen ergibt. Die beiden erklärenden Variablen müssen dabei endogene Variablen des Modells sein.¹⁴ Strukturelle Gleichungen, welche Compound-Variablen erklären, sind zumeist Definitionsgleichungen. Wird beispielsweise der Umsatz in einem Modell als Produkt aus Preis mal Menge definiert, und sind Preis und Menge endogene Variablen dieses Modells, so stellt der Umsatz eine Compound-Variable dar.

Compound-Variablen können aber auch durch Hypothesengleichungen erklärt werden. So wird beispielsweise in FORRESTERs Weltmodell die Todesrate der Weltbevölkerung aus dem Produkt von vier Multiplikatoren bestimmt, die den Einfluß des Lebensstandards, der Umweltver-

¹⁴ Wäre z.B. eine der erklärenden Variablen eine exogene Variable, so könnte ein lineares Modell mit zeitvariablen Koeffizienten vorliegen.

schmutzung, der Nahrungsmittelversorgung und der Überbevölkerung beschreiben. [Vgl. Seite 480]

Die durch Compound-Variablen gekennzeichnete multiplikative Verstärkung oder Abschwächung von erklärenden Variablen ist als Wirkungsprinzip nicht sehr plausibel, wenn sie menschliches Verhalten zum Ausdruck bringen soll.

(2) Monokausale Hypothesengleichungen mit abnehmendem Grenzzuwachs Hypothesen dieser Art werden im ökonomischen Bereich zumeist zur Beschreibung von Beziehungen benutzt, die bestimmte Konkretisierungen des Ertragsgesetzes ausdrücken. Dies bedeutet, daß es sich um Funktionsverläufe handelt, in denen der positive Grenzzuwachs der endogenen Variablen mit zunehmender erklärender Variablen abnimmt.

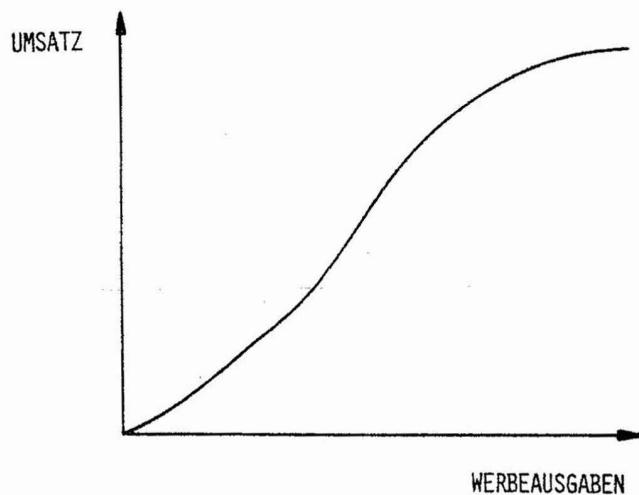


Abb. 21.29 Zusammenhang zwischen Werbeausgaben und Umsatz (Werbeertragsfunktion)

Als Beispiel sei auf den Zusammenhang zwischen den wirksamen Werbeausgaben und den durch sie induzierten Umsätzen hingewiesen, der durch eine 'Werbeertragsfunktion' wie in Abbildung 21.29 dargestellt wird.

Diese in der amerikanischen Literatur auch als 'sales response function' bezeichnete Funktion wird oft durch eine der folgenden Funktion-

nen dargestellt. [114, S.224]

$$U = a\sqrt{W}$$

$$U = a + bW + cW^2 + dW^3$$

mit U =Umsatz und W =Werbeausgaben

In vielen Fällen ist es jedoch nicht möglich, bestimmte nichtlineare Funktionen anhand derartiger elementarer Formulausdrücke zu beschreiben. Man behilft sich dann damit, die vorgegebenen nichtlinearen Verläufe durch sogenannte Tabellenfunktionen darzustellen. Tabellenfunktionen bewirken eine stückweise Linearisierung eines nichtlinearen Verlaufes mit Hilfe von Polygonzügen.

Abbildung 21.30 zeigt eine durch einen Polygonzug ersetzte nichtlineare Konsumfunktion eines MA-Modells.

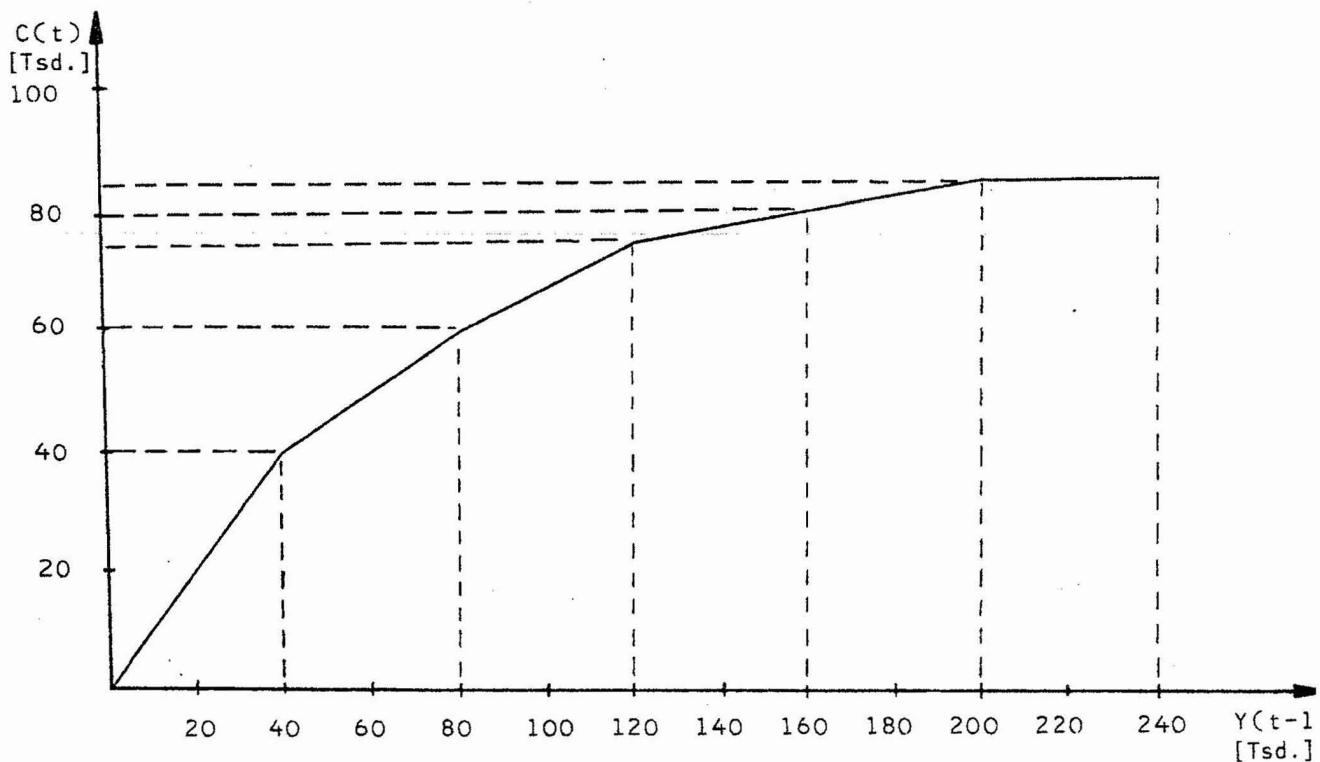


Abb. 21.30 Beispiel einer nichtlinearen Konsumfunktion, welche zur modellmäßigen Erfassung durch einen Polygonzug annähernd wiedergegeben wird

Die Formulierung von Tabellenfunktionen erfolgt fast ausschließlich direkt in einer ein dynamisches System beschreibenden Computersimulationssprache.¹⁵ Daher werden Modelle mit Tabellenfunktionen auch nicht analytisch untersucht, sondern nur anhand von Simulationsexperimenten.

(3) Hypothesengleichungen mit oberen und unteren Grenzen

In monokausalen Hypothesen wird oft der ansonsten positiv ansteigende Funktionsverlauf nach unten, oben oder beiderseits durch sogenannte Schwellen oder Plafonds begrenzt. Derartige Schwellen wurden bereits von TINBERGEN in seinen Modellen verwendet. [205, S.120]

Das nichtlineare Konjunkturmodell von GOODWIN gewinnt seine Nichtlinearität aus der Tatsache, daß die Höhe der Investitionen nach oben durch die Kapazitätsgrenze der Investitionsgüterindustrie, nach unten durch die normalen Abschreibungsbeträge begrenzt ist. [70, S.4] In dem Modell von HICKS entsteht die Nichtlinearität ebenfalls durch eine obere Begrenzung sowie stückweise aneinander anschließende Geraden mit unterschiedlich positivem Anstieg. (Vgl. [101, S.200])

Auch in einzelbetrieblichen Modellen kann die Kapazitätsgrenze die

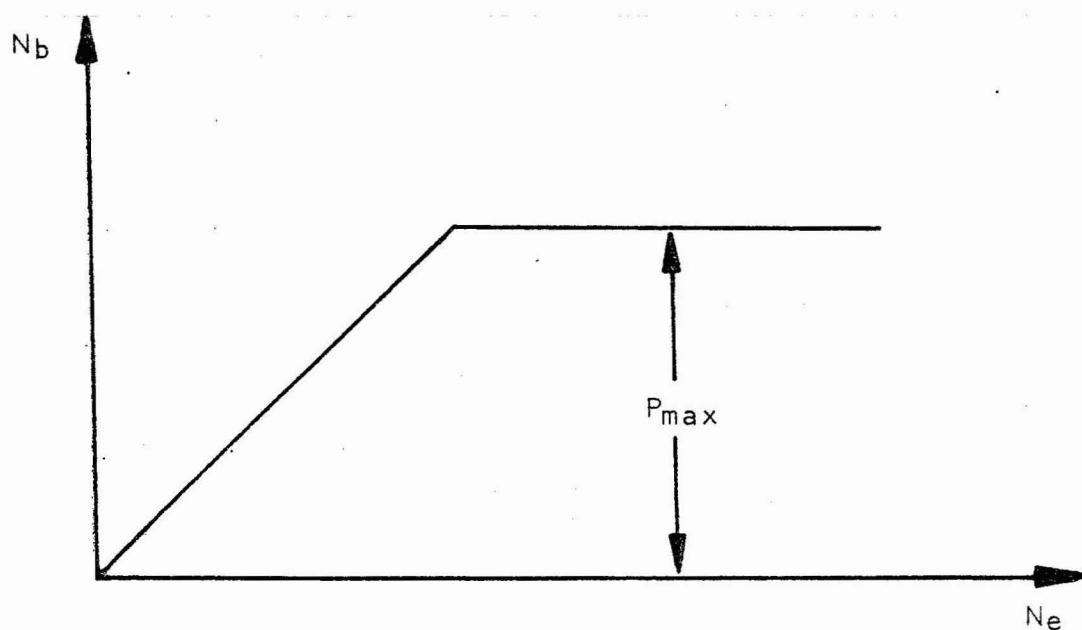


Abb. 21.31 Zusammenhang zwischen zu befriedigender und effektiv vorhandener Nachfrage

¹⁵ Vgl. Seite 419f.

obere Begrenzung einer linearen Funktion mit positivem Anstieg bilden. In Abbildung 21.31 beispielsweise ist ein solcher Zusammenhang zwischen der von einem Betrieb zu befriedigenden Nachfrage N_b , der effektiv vorhandenen Marktnachfrage N_e und der Produktionskapazität P_{\max} dieses Betriebes in einer Periode dargestellt.

Ist in einer multikausalen Hypothesengleichung der Funktionsverlauf für bestimmte Argumentenbereiche durch unterschiedliche Formulausdrücke definiert, dann ist es zweckmäßig, solche funktionalen Verknüpfungen konditional auszudrücken:

Beispielsweise kann der Zusammenhang zwischen der von einem Betrieb zu befriedigenden Nachfrage N_b , dem Lagerbestand L und der effektiv gegebenen Marktnachfrage N_e durch die in Abbildung 21.32 dargestellte Funktion beschrieben werden.

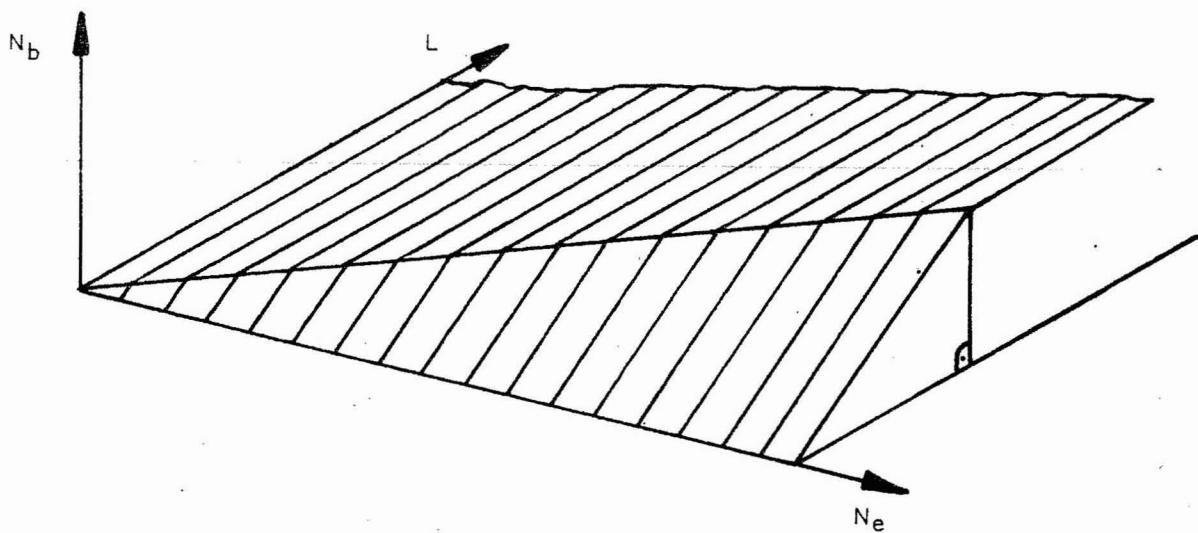


Abb. 21.32 Funktionale Beziehung zwischen dem Lagerbestand L , der effektiven Marktnachfrage N_e und der durch den Betrieb zu befriedigenden Nachfrage N_b

In einer konditionalen Darstellungsweise wird dieser Zusammenhang folgendermaßen gekennzeichnet:

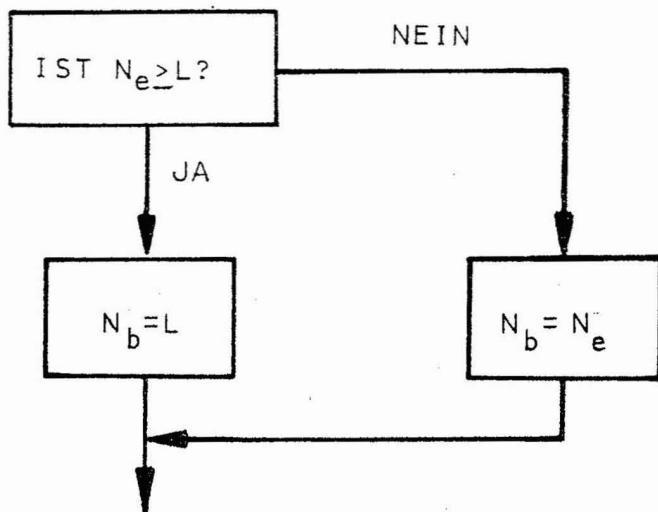


Abb. 21.33 Konditionalstruktur einer Entscheidung über die Höhe der zu befriedigenden Nachfrage N_b

Die Formulierung nichtlinearer Beziehungen in dynamischen Simulationsmodellen wird zu einem großen Teil in Form derartiger Konditionalstrukturen vorgenommen.

Der funktionale und nichtlineare Charakter dieser Konditionalstrukturen wird bei einer ersten Betrachtung nicht immer deutlich. Man sollte sich daher bewußt sein, daß bei der Untersuchung der nichtlinearen Eigenschaften bestimmter Simulationsmodelle derartige Konditionalstrukturen einen starken Anteil an der Nichtlinearität der Modellzusammenhänge besitzen.

B. Analyse nichtlinearer Modelle

Nach Erörterung der bedeutsamen Arten nichtlinearer Beziehungen in ökonomischen Ansätzen stellt sich die Frage nach den zur Verfügung stehenden analytischen Methoden zur Erschließung bestimmter Modellimplikationen. Für die analytische Untersuchung nichtlinearer Differenzengleichungssysteme ist es von Bedeutung, daß sich diese stets zu einem äquivalenten System von Differenzengleichungen ersten Grades umformen lassen. Dies läßt sich einfach zeigen:

Ist in einem beliebigen nichtlinearen Ansatz eine endogene Variable

$Y(t), Y(t-1), Y(t-2), \dots, Y(t-n)$ enthalten, so können folgende Definitionen eingeführt werden

$$\begin{aligned} Z_1(t) &= Y(t) \\ Z_2(t) &= Y(t-1) = Z_1(t-1) \\ Z_3(t) &= Y(t-2) = Z_2(t-1) \\ &\vdots \quad \vdots \quad \vdots \quad \vdots \\ &\vdots \quad \vdots \quad \vdots \quad \vdots \\ Z_n(t) &= Y(t-n+1) = Z_{n-1}(t-1) \end{aligned}$$

Unter Einbeziehung dieser Definitionen wird im primären Ansatz die Substitution $Y(t-i)=Z_i(t-1)$ für $i=1, \dots, n$ vorgenommen. Werden alle endogen verzögerten Variablen des primären Ansatzes in gleicher Weise behandelt, so erhält man letztlich ein nichtlineares System von Differenzengleichungen ersten Grades.

Im Gegensatz zum linearen Fall gibt es keine einheitliche Théorie nichtlinearer Differenzengleichungen. Analytische Verfahren sind zu meist nur für spezielle Typen nichtlinearer Differenzengleichungen bekannt und dienen zur Ermittlung der Funktionslösung oder bestimmter Kennzeichen des dynamischen Systemverhaltens.

Funktionslösungen nichtlinearer Ansätze sind nur in wenigen Fällen möglich. (Vgl. [105, S. 175], [191], [209])

Als Folge dessen stellt sich die Frage, ob zumindest gewisse analytische Methoden zur Verfügung stehen, die es erlauben, ohne Kenntnis der Funktionslösungen bestimmte dynamische Verhaltenscharakteristiken nichtlinearer Systeme zu ermitteln.

Eine über den Einzelfall hinausgehende mathematische Theorie stellt die sogenannte direkte Methode von Ljapunow dar, mit deren Hilfe das dynamische Verhalten nichtlinearer Ansätze beurteilt werden kann. Diese Methode wurde ursprünglich von LJAPUNOW zur Beurteilung der Stabilität von Differentialgleichungen entwickelt und ist von HAHN auf die Stabilitätsanalyse von Differenzengleichungen übertragen worden.¹⁶

Die Ljapunowsche Methode lässt sich folgendermaßen charakterisieren: In dem System $Y(t)=F[Y(t-1)]$ bilden $Y(t)$ und $Y(t-1)$ die Spaltenvek-

¹⁶ Zu einer ausführlichen Erörterung der Stabilitätstheorie von LJAPUNOW siehe [170], zur Arbeit von HAHN siehe [77].

toren der endogenen Modellvariablen. Besitzt das System eine Lösung $F[\bar{Y}] - \bar{Y} = 0$, so kann dieser Gleichgewichtspunkt im Falle von $\bar{Y} \neq 0$ durch die Koordinatentransformation $X = Y - \bar{Y}$ in den Ursprung gelegt werden, so daß sich die triviale Lösung $F[X] = 0$ mit $X = 0$ ergibt.

Bei Betrachtung eines bestimmten Anfangswertes $X_a \neq 0$ kann, wie erwähnt, immer dann von einer asymptotischen Stabilität gesprochen werden, wenn $X(t)$ mit wachsendem t gegen Null konvergiert. Dieser Fall liegt unter folgenden Voraussetzungen vor:¹⁷

- (1) Es sei $V[X(t)]$ eine positive definite Funktion, d.h. $V[X(t)] > 0$ mit Ausnahme von $X=0$, wo sie Null ist.
- (2) Die Funktion der ersten Differenz von $V[X(t)]$

$$\Delta V[X(t)] = V[X(t+1)] - V[X(t)]$$

sei negativ und im Ursprung Null.

Erfüllt eine Funktion die Bedingungen (1) und (2), so wird sie als diskrete Ljapunow-Funktion bezeichnet.

Als Beispiel sei die Differenzengleichung

$$X(t) + X^2(t-1) + X^2(t-2) = 0 \quad (21.131)$$

angeführt. (Vgl. [214, S.234]). Es wird die Ljapunow-Funktion

$$V[X(t)] = 2X^2(t-1) + X^2(t-2) \quad (21.132)$$

eingeführt. Damit wird

$$\Delta V[X(t)] = 2X^2(t) + X^2(t-1) - 2X^2(t-1) - X^2(t-2) \quad (21.133)$$

Aus (21.131) folgt

$$2X^2(t) = 2X^4(t-1) + 4X^2(t-1)X^2(t-2) + 2X^4(t-2)$$

Mit (21.133) folgt

$$\Delta V[X(t)] = [X^2(t-1) + X^2(t-2)][2X^2(t-1) + 2X^2(t-2) - 1] \quad (21.134)$$

Es zeigt sich, daß Gleichung (21.132) die Bedingung (1) erfüllt. Bedingung (2) gilt gemäß (21.134) im Falle:

$$2X^2(t-1) + 2X^2(t-2) - 1 < 0 \quad (21.135)$$

¹⁷ Zum Beweis [61, S.160]

Ist also Bedingung (21.135) erfüllt, so ist das durch die Differenzengleichung beschriebene System asymptotisch stabil.

Eine Schwierigkeit bei der Anwendung dieser Methode resultiert aus der Tatsache, daß es kein allgemeines Verfahren zur Ermittlung einer geeigneten Ljapunow-Funktion gibt.

Weiterhin ist zu bemerken, daß die auf diese Weise ermittelten Parameterbereiche nur hinreichend, aber nicht notwendig für das Vorliegen von Stabilität sind. Durch die Wahl einer anderen Ljapunow-Funktion ist es durchaus möglich, andere Stabilitätsbedingungen zu erhalten. Die größte praktische Schwierigkeit bei der Anwendung der Ljapunow-Methode ergibt sich aber aus der Forderung, daß die ermittelten Stabilitätsbedingungen, wie zum Beispiel die Bedingung (21.135), von allen zeitlichen Realisationen der endogenen Variablen erfüllt werden müssen. Der Zeitverlauf der endogenen Variablen $X(t)$ ist jedoch gar nicht bekannt. Vielmehr sind in der Regel nur die Anfangswerte und der Verlauf bestimmter exogener Variablen vorgegeben. VIDAL hat, aufbauend auf die Arbeiten von CETAEV und AJZERMAN, zwei Methoden angegeben, mit denen unter Verwendung von Ljapunowschen Funktionen hinreichende Bedingungen bezüglich der Anfangswerte für eine asymptotische Stabilität ermittelt werden können. Eine Anwendung ist jedoch nur in wenigen Fällen möglich. [214, S.229]

Als Ergebnis ist festzuhalten, daß nach dem gegenwärtigen Erkenntnisstand die praktische Anwendung analytischer Verfahren der Stabilitätsanalyse relativ beschränkt ist.

Weitere analytische Verfahren zur Beurteilung des dynamischen Verhaltens nichtlinearer Systeme sind im Rahmen der ökonomischen Wachstumstheorie zur Beurteilung des proportionalen Wachstums von Wirtschaftssystemen entwickelt worden.

Hierbei wird von einem nichtlinearen Differenzengleichungssystem

$$Y(t) = H[Y(t-1)] \quad (21.136)$$

ausgegangen, in dem Y den Vektor des sektoralen Einkommens oder des mengenmäßigen Ausstoßes einer Wirtschaft repräsentiert. (Vgl. [184], [195], [142], [140]). Hinsichtlich der funktionalen Verknüpfung $H[]$

werden bestimmte restriktive Annahmen wie Monotonität und Homogenität unterstellt. Im Rahmen dieser Restriktionen werden die Bedingungen eines proportionalen Wachstums aufgezeigt.

NIKAIDO hat den Ansatz (21.136) unter den gleichen Restriktionen für H auf die Form

$$Y(t) = H[Y(t-1)] + A(t-1)$$

erweitert. Hierbei repräsentiert Y wiederum den Vektor der sektoralen Einkommen, während A den Vektor der exogen bedingten sektoralen Ausgaben bildet. Dieser nichtlineare Prozeß der Einkommensentwicklung wird von NIKAIDO hinsichtlich der Bedingungen eines proportionalen Wachstums untersucht.

Die Beschreibung einiger analytischer Verfahren zur Behandlung nichtlinearer dynamischer Modelle soll nicht den Eindruck erwecken, daß diese Verfahren für die Analyse der heute zur Verfügung stehenden Modelle von sehr großer Bedeutung sind. Sämtliche derzeitigen relevanten nichtlinearen Modelle sind praktisch nur mit Hilfe von Computersimulationen untersuchbar. Die Beschränktheit der diskutierten Verfahren macht die Relevanz von Computersimulationen deutlich.

Neben den analytischen stehen auch bestimmte geometrische Verfahren zur Beurteilung nichtlinearer Systeme zur Verfügung. Auch diese Verfahren, die sich nur auf eine nichtlineare Endgleichung ersten oder höchstens zweiten Grades beziehen, sind aus der Anwendungssicht als irrelevant zu bezeichnen, da realistische Modelle ja stets komplexer sind. Das Studium dieser Verfahren liefert jedoch typische Einsichten in das (von linearen Systemen abweichende) Verhalten nichtlinearer Systeme, welche auch für die Beurteilung der Simulation komplexer nichtlinearer Systeme von Bedeutung sind.

Wir wenden uns daher der Analyse einer beliebigen nichtlinearen Endgleichung ersten Grades, d.h. einer Gleichung der Form

$$Y(t) = F[Y(t-1)]$$

zu. Der Zeitverlauf von $Y(t)$ kann hierbei durch eine einfache gra-

phische Darstellung in der Form eines sogenannten Phasendiagramms ermittelt werden.

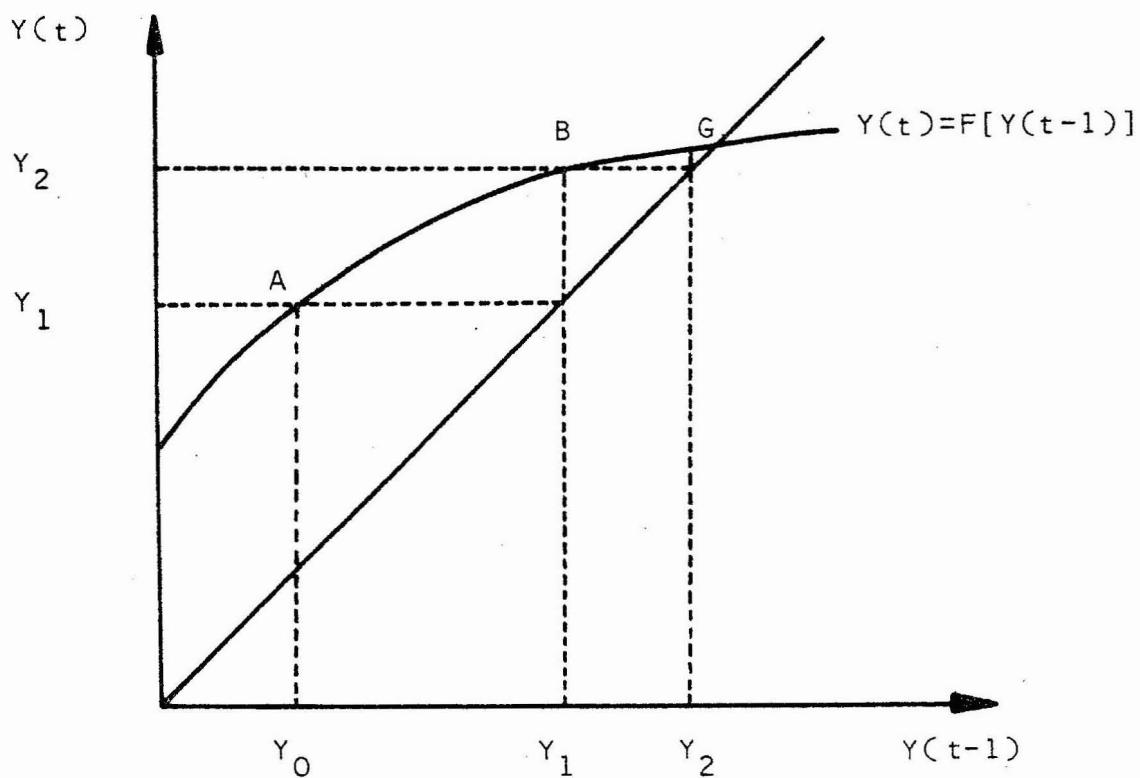


Abb. 21.34 Phasendiagramm einer nichtlinearen Differenzengleichung ersten Grades

In dem Koordinatensystem der Abbildung 21.34 sind die Funktionen $Y(t)=F[Y(t-1)]$ und $Y(t)=Y(t-1)$ eingetragen. Ist wie in Abbildung 21.34 ein bestimmtes $Y(0)$ vorgegeben, so ergibt sich wegen $Y(1)=F[Y(0)]$ die Größe $Y(1)$ als der Ordinatenwert des Punktes A.

Den Wert von $Y(2)$ erhält man, wenn der gefundene Ordinatenwert von $Y(1)$ auf die Abszisse übertragen wird und dann analog zum ersten Schritt $Y(2)$ aus dem Schnittpunkt der Senkrechten über $Y(1)$ mit der Funktion F ermittelt wird (Punkt B). Die Übertragung der Strecke $Y(1)$ von der Ordinate auf die Abszisse kann dadurch sehr einfach vorgenommen werden, indem man von dem Ordinatenwert $Y(1)$ eine waagerechte Linie zieht und vom Schnittpunkt dieser Linie mit der Funktion $Y(t)=Y(t-1)$,

d.h. der 45° -Linie nach unten lotet. Der Schnittpunkt des Lotes mit der Abszisse ergibt $Y(1)$.

Durch Fortsetzung dieses Verfahrens gelangt man zu $Y(3)$, $Y(4)$ usw. Man erkennt aus Abbildung 21.34, daß das System in diesem Fall einem Gleichgewicht im Punkt G zustrebt.

Die hier beschriebene Darstellungsform nichtlinearer Differenzengleichungen ersten Grades läßt erkennen, daß ihr dynamisches Verhalten sowohl vom Anstieg der Funktion F als auch ihrer Lage zur 45° -Linie beeinflußt wird. Das dynamische Verhalten läßt sich durch folgende Maximen kennzeichnen: [14, S. 265]

- (1) Wenn sich die Funktion $F[Y(t-1)]$ über der 45° -Linie befindet, dann ist $Y(t)$ stets größer als $Y(t-1)$, d.h. Y weist einen wachsenden Verlauf auf. Demgegenüber zeigt Y einen abnehmenden Verlauf, wenn sich F unter der 45° -Linie befindet.
- (2) Wenn der Anstieg von $F[Y(t-1)]$ positiv und kleiner als Eins ist, dann weist das System ein monoton gedämpftes Verhalten auf.
- (3) Ist der Anstieg von $F[Y(t-1)]$ positiv und größer als Eins, so besitzt das System ein monoton ungedämpftes Verhalten.
- (4) Weist $F[Y(t-1)]$ einen negativen Anstieg auf, dann zeigt das System ein oszillierendes Verhalten, das sich für $F'[Y(t-1)] < -1$ als ungedämpft oszillierend und für $0 > F'[Y(t-1)] > -1$, als gedämpft oszillierend erweist.

Eine Funktion $F[Y(t-1)]$ kann jedoch im Rahmen ihres Definitionsbereiches sowohl unterschiedliche Anstiege als auch Positionen zur 45° -Linie aufweisen, so daß in bestimmten Intervallen des Definitionsbereiches unterschiedliche Verhaltensweisen gemäß (1) bis (4) auftreten können. Abbildung 21.35 zeigt eine Funktion, für die dieses zutrifft.

Das Verhalten eines konkreten Systems wird in diesem Fall von der Lage des Anfangswertes $Y(0)$ entscheidend beeinflußt. Liegt der Anfangswert in dem Intervall A'C', so wird das System gegen den Gleichgewichtspunkt B' konvergieren. Liegt er dagegen im Intervall C'E', so strebt das System gegen den Punkt D'. Falls der Anfangswert nicht in diese beiden Intervalle fällt, weist das System ein instabiles Verhalten auf.

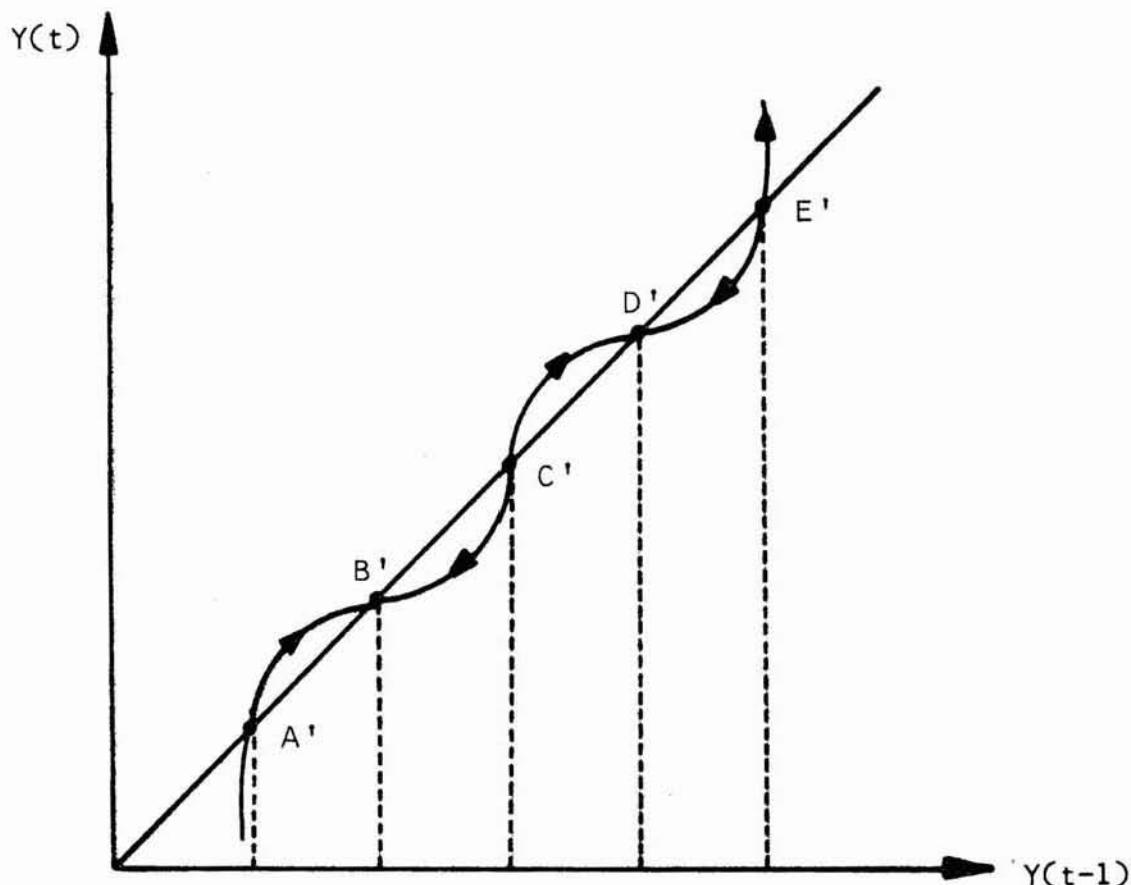


Abb. 21.35 Phasendiagramm einer nichtlinearen Differenzengleichung ersten Grades

Das Beispiel zeigt ebenfalls in anschaulicher Form, daß es - wie erörtert - im Rahmen nichtlinearer Systeme notwendig ist, zwischen lokaler und globaler Stabilität zu unterscheiden.¹⁸ Während das in Abbildung 21.35 beschriebene System nur lokale Stabilität aufweist, besitzt das in Abbildung 21.34 angeführte System eine globale Stabilität.

Untersuchungen nichtlinearer MZÄ-Modelle mit Hilfe von Phasendiagrammen liegen bisher kaum vor. Neben einem Ansatz von DORFMAN, SAMUELSON und SOLOW [39, S.333] untersuchte vor allem DAY die Bedingungen des Wachstums von Einzelbetrieben anhand einer quadratischen Differenzengleichung erster Ordnung unter Verwendung von Phasendiagrammen. [35]

Auch zur geometrischen Analyse nichtlinearer Differenzengleichungen

¹⁸ Vgl. im einzelnen Seite 77

zweiten Grades stehen bestimmte Diagrammtechniken zur Verfügung.

[214, S.112f.]

Im Gegensatz zu dem beschriebenen Verfahren sind diese Techniken jedoch relativ aufwendig und unübersichtlich, so daß es empfehlenswert erscheint, auch schon derartige Ansätze relativ geringer Komplexität mit Hilfe von Simulationsverfahren zu untersuchen.

Im Rahmen der Erörterung linearer Modelle wurde eine Reihe 'linear-modellspezifischer' Kennzeichen wie der Totalmultiplikator, die Einheitsimpulsantwort oder die Sprungantwort eines linearen Systems behandelt:

Es liegt nahe, nach der Übertragung dieser Begriffe auf nichtlineare Systeme zu fragen. Grundsätzlich gilt, daß eine solche Übertragung auf ein nichtlineares System nicht ohne weiteres möglich ist. Denn ein derartiges System (als negative Abgrenzung eines linearen Systems) läßt sich gerade nicht durch solche systemspezifischen Kenngrößen charakterisieren.

So besitzt ein nichtlineares System keine systemspezifische Einheitsimpulsantwort, die ja im linearen Fall den gesamten Transformationsmechanismus des Systems zum Ausdruck bringt. Bei nichtlinearen Systemen hängen die Verläufe der Impulsantworten aber von den Impulshöhen ab. Wählt man in einem linearen oder nichtlinearen Modell die Anfangswerte derart, daß sich das System im Gleichgewicht GW befindet, und bezeichnet man die Impulsantwort eines Systems bezüglich eines Impulses der Höhe IH als IA(t), so kann man die Kenngröße

$$SEI(t) = \frac{IA(t) - GW}{IH} \quad (21.137)$$

formulieren. SEI soll als standardisierte Einheitsimpulsantwort bezeichnet werden.

In einem linearen System verändert sich SEI(t) nicht bei Variation der Impulshöhe IH.

Im Gegensatz dazu ändert sich bei nichtlinearen Systemen der Verlauf von SEI(t) in Abhängigkeit von der Impulshöhe. Dieser Umstand kann dazu verwendet werden, die Stärke der Nichtlinearität eines Systems

zu beurteilen. Eine derartige Untersuchung soll am Beispiel eines MA-Systems demonstriert werden, welches eine nichtlineare Konsum- und Investitionshypothese besitzt.

$$Y(t) = C(t) + I_i(t) + 46000$$

$$I_i(t) = \begin{cases} 3,5[C(t)-C(t-1)] & \text{für } C(t)-C(t-1) \geq 0 \\ 0 & \text{für } C(t)-C(t-1) < 0 \end{cases}$$

$$C(t) = F_1[Y(t-1)]$$

Die Investitionshypothese besagt, daß bei einem Rückgang der Nachfrage die Unternehmer ihre induzierten Investitionen einstellen. Es handelt sich daher um eine nichtlineare 'Schwellenhypothese'. Weiterhin wird angenommen, daß die Konsumfunktion $F_1[Y(t-1)]$ durch die in Abbildung 21.30 dargestellte Tabellenfunktion beschrieben wird. Bezeichnet man die Investitionshypothese verallgemeinernd mit $I_i(t)=F_2[C(t)-C(t-1)]$, dann kann man analog zum linearen Fall die Endgleichung

$$Y(t) = F_1[Y(t-1)] + F_2[F_1[Y(t-1)] - F_1[Y(t-2)]] + 46000$$

aufstellen. Eine analytische Bestimmung des Gleichgewichtswertes von

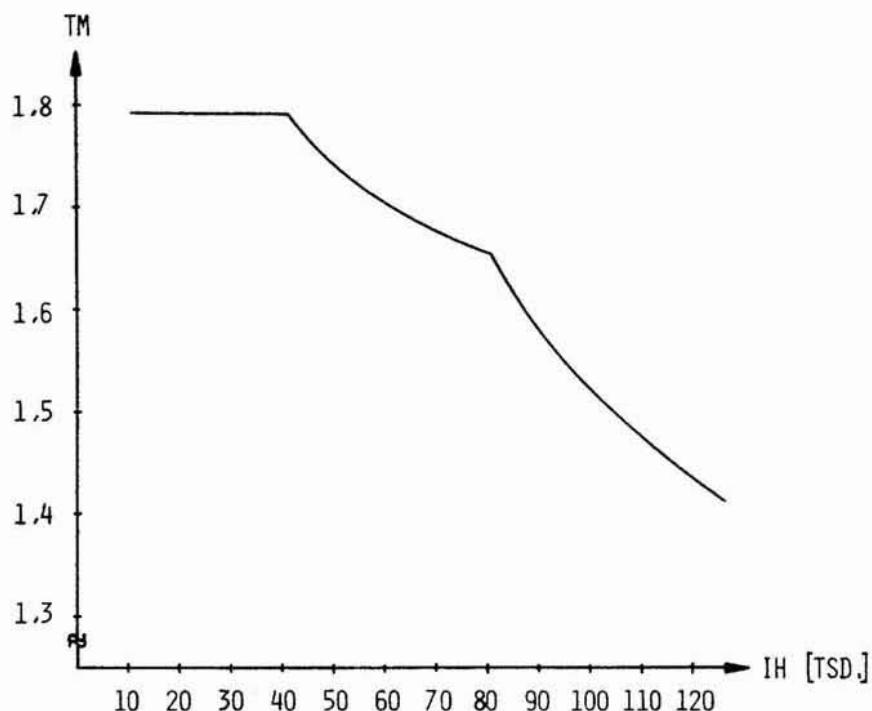


Abb. 21.36 Zusammenhang zwischen Totalmultiplikator TM und Impuls Höhe IH in einem nichtlinearen MA-System

Y wäre zwar grundsätzlich möglich, aber viel zu aufwendig. Anhand einer Simulation zeigt sich, daß das System bei $Y=120 \cdot 10^3$ ein Gleichgewicht aufweist. Dieser Gleichgewichtswert GW wird im folgenden zu Grunde gelegt (d.h. $Y(0)=Y(1)=120 \cdot 10^3$). Dem System werden nunmehr positive Impulse verschiedener Höhe IH aufgeprägt, und es werden die standardisierten Einheitsimpulsantworten SEI ermittelt.

Die Summe der Einheitsimpulsantworten ergibt eine Größe, die dem Totalmultiplikator eines linearen Systems entspricht. Weil es sich aber um ein nichtlineares System handelt, werden sich mit variierenden Impulshöhen unterschiedliche Totalmultiplikatoren ergeben.

In Abbildung 21.36 ist der ermittelte Totalmultiplikator verschiedenen Impulshöhen gegenübergestellt.

Während im linearen Fall der Totalmultiplikator nicht durch die Impulshöhe beeinflußt wird, zeigt sich hier eine Abhängigkeit, die als eine Charakteristik der Nichtlinearität des vorliegenden Modells aufgefaßt werden kann.

2.2. Offene und geschlossene Modellformen

Die Bezeichnungen 'offenes' und 'geschlossenes System' werden in so vielfältiger Bedeutung im Rahmen systemtheoretischer Betrachtungen verwendet, daß es vielleicht ratsam wäre, solche abgenutzten Worte nicht mehr zu verwenden. Wenn wir dennoch das Begriffspaar offengeschlossen zur Klassifikation dynamischer Systeme gebrauchen, so zeichnet sich diese begriffliche Festlegung dadurch aus, daß ihre Unterscheidungsmerkmale an syntaktischen und empirisch interpretierbaren Strukturelementen dynamischer Modelle ansetzen.

Am Beispiel eines MA-Modells sollen diese Unterschiede herausgearbeitet werden. Betrachten wir die Endgleichung

$$Y(t) = 1,89Y(t-1) - 0,99Y(t-2) + I_a(t) \quad (22.1)$$

dann hängt die Entscheidung, ob es sich um ein offenes oder geschlossenes System handelt, von dem konkreten Verlauf der exogenen Vari-

ablen $I_a(t)$ ab. Falls für den gesamten Betrachtungszeitraum $I_a(t)=0$, dann kann von jedem kalendarischen Zeitpunkt $n=0,1,2,\dots$ ein zeitliches Bezugssystem für t aufgebaut werden.

In Abbildung 22.1 erkennt man, daß ein Zeitsystem mit $t=0,1,\dots$ im Jahre 1952 ansetzt, während ein zweites im Jahre 1950 beginnt. Für beide Anfangszeitpunkte gilt im Falle $I_a(t)=0$ dieselbe Endgleichung mit unterschiedlichen Anfangswerten.

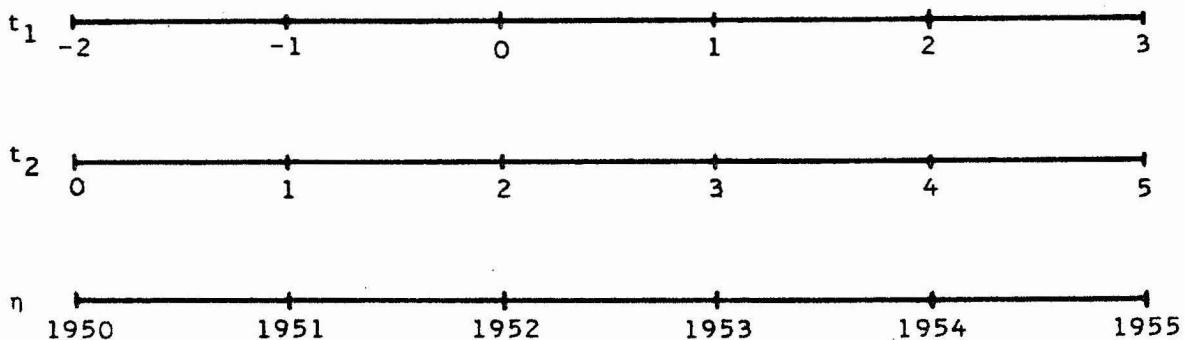


Abb. 22.1 Beispiel zur Unterscheidung von offenen und geschlossenen Modellen

Generell spricht man immer dann von einem geschlossenen Modell, wenn die Gültigkeit der Modellhypothesen von einer Verschiebung des zeitlichen Bezugssystems nicht berührt wird.

Nach dieser Feststellung liegt die Frage nahe, unter welchen Umständen ein dynamisches Modell eine derartige 'Verschiebbarkeit' seines Zeitsystems nicht mehr zuläßt.

Gehen wir davon aus, daß die Endgleichung in dem Zeitsystem T_{1952} , welches im Jahre 1952 beginnt, durch einen Verlauf der autonomen Investitionen der Form $I_a(t)=2^t$ ($t=0,1,2,\dots$) beschrieben wird. Damit wird für $I_a(t)$ die folgende Zeitreihe festgelegt

t	0	1	2	
Jahr:	1952	1953	1954	usw.
Wert von I_a :	1	2	4	

Die Werte für I_a im Jahre 50 und 51 sind in diesem Ansatz nicht bekannt. Prüfen wir nunmehr, ob die Endgleichung

$$Y(t) = 1,89Y(t-1) - 0,99Y(t-2) + 2^t$$

auch für das Zeitsystem T_{1950} Gültigkeit hat, welches im Jahre 1950 beginnt, so sehen wir, daß dieser Ansatz ab 1952 zu dem Verlauf

t	2	3	4	
Jahr:	1952	1953	1954	usw.
Wert von I_a :	4	8	16	

führt, d.h. die autonomen Investitionen nunmehr einen geänderten Verlauf aufweisen. Es zeigt sich damit, daß in Fällen, in denen sich der numerische Wert der exogenen Variablen verändert, eine Transformation des Zeitsystems nicht möglich ist, weil sich damit der Inhalt der Hypothesengleichungen ändern würde. Wir können nunmehr folgende Festlegung treffen: Wenn ein Modell exogene Variablen aufweist, die in den einzelnen Kalenderzeitpunkten ihres Auftretens unterschiedliche Werte annehmen, dann handelt es sich um ein offenes, im andern Fall um ein geschlossenes Modell. Im Fall linearer Systeme mit konstanten Koeffizienten wird ein offenes System durch eine inhomogene Endgleichung der Form

$$Y(t) + a_1Y(t-1) + \dots + a_nY(t-n) = E(t)$$

mit $E(t) \neq 0$ für $t = \dots, -1, 0, 1, \dots$ beschrieben, ein geschlossenes System dagegen durch eine homogene Endgleichung

$$Y(t) + a_1Y(t-1) + \dots + a_nY(t-n) = 0$$

Neben der hier verwendeten Bezeichnung des Begriffspaares offen und geschlossen werden eine Fülle unterschiedlicher Bezeichnungen, wie kausal - historisch, inhomogen - homogen, gezwungen - ungezwungen, nicht autonom - autonom oder semi-endogen - endogen, absolut isoliert - relativ isoliert verwendet. (Vgl. [61, S. 27], [123, S. 205], [208, S. 3], [72, S. 9])

In welchem Umfang diese Begriffspaaare mit der hier entwickelten Differenzierung in offen und geschlossen übereinstimmen, ist jedoch

nicht immer klar zu erkennen, weil viele Autoren mit diesen Begriffen zwar im Rahmen bestimmter Untersuchungen operieren, die ihren Begriffsbildungen zugrunde liegenden Abgrenzungskriterien aber nicht eindeutig definieren, sondern mit exemplarischen Definitionen arbeiten.

Zur empirischen Interpretation offener und geschlossener Modelle sei folgender Aspekt angeführt. Es ist ein Ziel naturwissenschaftlicher Forschung, zeitinvariante Naturgesetze festzustellen. [169, S.401] Analog hierzu kann man fordern, daß es ein Ziel der sozialwissenschaftlichen Forschung sein sollte, zeitinvariante strukturelle Gleichungen, d.h. Verhaltensgleichungen sowie technologische und institutionelle Gleichungen aufzufinden. Die Forderung nach der Ermittlung zeitinvarianter Beziehungen bedeutet, daß die strukturellen Gleichungen bildenden Beziehungen nicht durch Größen verändert werden dürfen, die explizit von der Zeit t abhängen. Daraus folgt, daß eine Formulierung eines gegebenen Zusammenhangs immer dann nicht als befriedigend angesehen werden kann, wenn es notwendig wird, die Zeit t explizit in diese Formulierung aufzunehmen. Würde man ein offenes Modell als adäquate Repräsentation einer bestimmten Theorie ansehen, so müßte man zugestehen, daß es zeitveränderliche empirische Gesetze gibt. Eine solche Konsequenz ist zwar logisch denkbar, würde aber einem Grundpostulat der heutigen Naturwissenschaften widersprechen, welches behauptet, daß sich eine Ursache (hier repräsentiert durch den Wert einer exogenen Variablen) nicht allein deswegen ändern kann, weil die Zeit voranschreitet, sondern, weil sie zugleich die Wirkung einer anderen Ursache ist.

Offene Systeme können daher als eine Art unvollkommener geschlossener Systeme interpretiert werden. [171, S.319]

Soll beispielsweise ein empirisch vorliegender Zusammenhang untersucht werden, so ist man fast immer gezwungen, eine bestimmte Teilmenge von Variablen aus dem Gesamtzusammenhang als Sub-system zu isolieren. Die Abgrenzung des Subsystems wird soweit wie möglich unter dem Gesichtspunkt erfolgen, seinen Umfang so festzulegen, daß mög-

lichst wenig zeitveränderliche Variablen das System von außen beeinflussen.¹ Eine völlige Isolierung läßt sich aber fast nie erreichen. Die Variablen, die das Subsystem von außen zeitveränderlich beeinflussen, sind bei Betrachtung des gesamten Systemzusammenhangs sehr häufig aber von den endogenen Variablen des Subsystems abhängig. Es besteht in diesen Fällen eine Beziehung zwischen den endogenen Variablen des Subsystems und seinen zeitveränderlichen exogenen Variablen. Da diese Beziehungen nicht bekannt sind, können die Verläufe der exogenen Variablen des Subsystems in einer ersten Annäherung als Funktionen der Zeit angesetzt werden. Gelingt es, durch weitere Informationen auch diese Beziehungen durch strukturelle Gleichungen zu fassen und die in diesen auftretenden zeitveränderlichen exogenen Variablen wiederum als Verknüpfungen der nunmehr erweiterten Menge der endogenen Variablen zu definieren, so gelangt man letztlich zu einem geschlossenen Modell. Dieses geschlossene Modell setzt sich hierbei ausschließlich aus zeitinvarianten Strukturgleichungen zusammen und bildet daher - wie schon erwähnt - aus wissenschaftstheoretischer Sicht die wünschenswerte Repräsentation einer Theorie. Wenn man jedoch von der Annahme einer allgemeinen Interdependenz aller Größen der empirischen Wirklichkeit ausgeht, so kann es im strengen Sinne kein empirisch gültiges geschlossenes Modell bestimmter empirischer Teilbereiche geben. Ein geschlossenes dynamisches Modell ist unter dieser Annahme nur in Gestalt eines Totalmodells im Sinne der Laplace'schen Weltformel denkbar. Nur in diesem Fall ist jegliche ceteris-paribus-Klausel aufgehoben.

Akzeptiert man diese Annahme, so folgt daraus, daß geschlossene Modelle gewissermaßen schon a priori eine beschränkte empirische Gültigkeit aufweisen, da die immer vorhandenen exogenen zeitveränderlichen Variablen im Modellansatz durch implizite oder explizite ceteris-paribus-Klauseln als unveränderlich angenommen werden.

Die in den Sozialwissenschaften entwickelten dynamischen Modelle sind fast ausnahmslos offene Modelle, d.h. sie besitzen zeitveränderliche exogene Variablen. KLEIN und GOLDBERGERs Modell der amerikani-

¹ Zu den Problemen einer zulässigen Subsystemabspaltung im Rahmen eines geschlossenen Modells, siehe Abschnitt 2.5.3.

schen Wirtschaft enthält beispielsweise in zwanzig Gleichungen achtzehn zeitveränderliche exogene Variablen.

Den Extremfall eines offenen Modells stellen die anfangs erwähnten zeitverzögerten (d.h. Bestandsgleichungen enthaltende) Definitionssysteme dar, die keine zeitinvarianten Hypothesen enthalten, sondern deren 'empirische Füllung' allein anhand der Schätzung ihrer exogenen Variablen erfolgt.² Eines der ersten geschlossenen Modelle im Bereich der Ökonomie wurde von KALECKI in Form einer gemischten Differential-Differenzengleichung zur Erklärung von Konjunkturverläufen entwickelt [107]. In den letzten Jahren wurden geschlossene Modelle vorwiegend von FORRESTER und seinen Schülern im Rahmen des von ihm entwickelten System-Dynamics-Konzeptes erstellt. Sie werden in diesem Zusammenhang diskutiert.³

Ob es gelingt, die Realität mit Hilfe eines geschlossenen Modells in adäquater Weise einzufangen, kann nur von Einzelfall zu Einzelfall beurteilt werden.

2.3. Zyklische und kaskadierende Modellformen

Die Unterscheidung zwischen zyklischen und kaskadierenden Modellen führt zu bestimmten Typen von Hypothesen, die sich als zyklische und kaskadierende Hypothesen bezeichnen lassen. In diesem Abschnitt werden die formale Struktur und empirische Interpretation dieser Hypothesenarten sowie die speziellen Probleme ihrer Implikationaufdeckung erörtert. Eng mit den zyklischen Hypothesen ist der Begriff eines Feedback verbunden. Dieser Begriff wird formal expliziert und kritisch diskutiert. Es folgt eine ausführliche Darstellung der Beziehungen zwischen zyklischen und den sogenannten sequentiellen Hypothesen, welche primär dem Ziel dient, einige konstruktive Einsichten in den Funktionsmechanismus linearer Modelle zu erhalten. Mit dieser Darstellung, welche die Beziehungen zwischen den Gewichts- und Impulsfunktionen linearer Modelle zum Inhalt hat, wird eine wichtige Grundlage für das Verständnis linearer Modelle geschaffen.

2 Siehe Seite 47

3 Siehe Seite 399ff.

2.3.1. Begriffliche Klärung und empirische Interpretation

A. Zyklische und kaskadierende Hypothesen

Unter einem zyklischen Modell soll ein Modell verstanden werden, in welchem zumindest eine endogene Variable von ihren eigenen verzögerten Realisationen abhängt. Ein kaskadierendes Modell zeichnet sich dagegen durch das Fehlen der verzögernden Beeinflussung einer endogenen Variablen auf sich selbst aus.

Zur Präzisierung der beiden Modellbegriffe ist es notwendig, zwischen zyklischen und kaskadierenden Hypothesen zu unterscheiden.

Kennzeichen einer (primären oder sekundären) zyklischen Hypothese ist es, daß die zu erklärende endogene Variable in ihrer eigenen Hypothesengleichung verzögert auftritt. So ist die bereits erwähnte demographische Hypothese der Weltbevölkerungsentwicklung

$$B(t) = 1,02B(t-1) \quad (23.1)$$

eine zyklische Hypothese, da die zu erklärende Variable B sich verzögert selbst erklärt. Primäre zyklische Hypothesen sind jedoch relativ selten. Das MA-Modell enthält beispielsweise keine primäre zyklische Hypothese. So zählt die Investitionshypothese

$$I_i(t) = \beta[C(t)-C(t-1)] \quad (23.2)$$

nicht zu den zyklischen Hypothesen, während die bereits früher beschriebenen sekundären Hypothesen des Investitionsverhaltens

$$I_i(t) = \alpha\beta[C(t-1)-C(t-2)+I_i(t-1)-I_i(t-2)+I_a(t-1)-I_a(t-2)] \quad (23.3)$$

und

$$I_i(t) = (\alpha+\alpha\beta)I_i(t-1) - \alpha\beta I_i(t-2) + \alpha\beta I_a(t-1) - \alpha\beta I_a(t-2) \quad (23.4)$$

in die Klasse der zyklischen Hypothesen fallen, weil I_i in verzögter Form zur Erklärung seiner 'zukünftigen' Ausprägung beiträgt.

Eine kaskadierende Hypothese zeichnet sich dadurch aus, daß die zu erklärende endogene Variable ausschließlich durch (verzögerte und unverzögerte) Variablen bestimmt wird, die nicht Verzögerungen der zu

erklärenden Variablen sind. Anders ausgedrückt: eine kaskadierende Hypothese liegt vor, wenn die Gruppe der vorherbestimmten Variablen, welche in einer Hypothesengleichung die endogene Variable Y bestimmen, selbst keine Verzögerungen dieser endogenen Variablen enthält. Kaskadierende Hypothesen lassen sich als eine negative Abgrenzung zur Klasse der zyklischen Hypothesen auffassen. Damit wird auch deutlich, daß die (primären) Hypothesen des MA-Modells kaskadierende Hypothesen sind, da sie, wie erwähnt, nicht zu den zyklischen Hypothesen zählen.

Die Unterscheidung zwischen zyklischen und kaskadierenden Hypothesen gibt einen Einblick in die Wirkungsrichtungen der kausalen Verknüpfungen dynamischer Modelle. Eine zyklische Hypothese kennzeichnet einen Zusammenhang, der oft als Feedback bezeichnet wird, weil die Ausprägung einer endogenen Variablen einen verzögerten Einfluß auf sich selbst ausübt. Eine völlige Gleichsetzung ist jedoch problematisch. Andernfalls müßte jede Bestandsfortschreibungsgleichung als Feedback gedeutet werden. Meist pflegt man erst dann von einem Feedback zu sprechen, wenn diese verzögerte Rückwirkung auf sich selbst über mehrere endogene Variablen erfolgt, was dadurch zum Ausdruck kommt, daß nicht die primären Hypothesen dieser endogenen Variablen, sondern erst die (durch Schleifenreduktion abgeleiteten) sekundären Hypothesen in Form von zyklischen Hypothesen auftreten.

Sind die primären oder auch sekundären Hypothesen eines Modells keine zyklischen Hypothesen, so kann man daraus nicht den zwingenden Schluß ziehen, das Modell enthalte keine Feedbacks. Dieses Urteil ist erst möglich, wenn man die primären oder sekundären minimalen Hypothesen (oder in mathematischer Ausdrucksweise die Endgleichungsform) des Modells ermittelt hat.

Enthält keine der Endgleichungen eine endogene verzögerte Variable, dann besitzt das Modell keine Feedbacks, und man kann von einem kaskadierenden Modell sprechen. Der dynamische Charakter eines solchen Modells resultiert in diesem Fall aus den Verzögerungen der exogenen Variablen. Als Beispiel sei ein Modell angeführt, durch welches der Zusammenhang zwischen den Werbeausgaben $W(t)$ eines Unternehmens, sei-

nem Umsatz $U(t)$ und den aus dem Umsatzprozeß resultierenden Kassen-eingängen $K(t)$ beschrieben wird.

$$\begin{aligned} U(t) &= 1,5W(t) + 6W(t-1) + 4,5W(t-2) + 3W(t-3) \\ K(t) &= 0,1U(t-1) + 0,4U(t-2) + 0,4U(t-3) + 0,1U(t-4) \end{aligned} \quad (23.5)$$

Die Werbeausgaben beeinflussen verzögert den Umsatz, und die aus dem Umsatzprozeß folgenden Geldeinnahmen treffen wegen der unterschiedlichen Zahlungsgewohnheiten der Schuldner ebenfalls verzögert ein. Es handelt sich um ein kaskadierendes Modell, denn die Gleichung für $U(t)$ ist bereits eine Endgleichung und enthält nur exogene verzögerte Variablen. Durch Einsetzung der Endgleichung von $U(t)$ in die Gleichung für $K(t)$ kann die Endgleichung für $K(t)$ ermittelt werden, die, wie man leicht erkennt, nur Verzögerungen von W enthält.

Die meisten Modelle gehören der Gruppe der zyklischen Modelle an, deren eindeutige Identifizierung auf der Grundlage der minimalen Hypothesen (bzw. Endgleichungen) möglich ist.

Im Falle eines MA-Modells, dessen primäre Hypothesen keine Feedbacks aufweisen, besitzen die minimalen Hypothesen oder Endgleichungen die folgende Form:

$$\begin{aligned} Y(t) &= (\alpha+\alpha\beta)Y(t-1) - \alpha\beta Y(t-2) + I_a(t) \\ I_i(t) &= (\alpha+\alpha\beta)I_i(t-1) - \alpha\beta I_i(t-2) + \alpha\beta I_a(t-1) - \alpha\beta I_a(t-2) \\ C(t) &= (\alpha+\alpha\beta)C(t-1) - \alpha\beta C(t-2) + \alpha I_a(t-1) \end{aligned} \quad (23.6)$$

In diesen Endgleichungen des MA-Modells schlagen sich die Einflüsse sämtlicher in dem System wirkender Feedbacks nieder. Es bietet sich daher an, Endgleichungen als die formalen Repräsentanten anzusehen, an denen sich oft verwendete Systemattribute wie positiver oder negativer Feedbackkreis explizieren und ihre empirischen Konsequenzen darstellen lassen. Ein solcher Versuch der formalen Aufweisung des Feedbackbegriffes im Rahmen linearer Systeme wird im folgenden unternommen.

Der Begriff des 'Feedback' ist im sozial- und wirtschaftswissenschaftlichen Bereich weit verbreitet. Von einem positiven Feedbackkreis wird gemeinhin gesprochen, wenn ein Anwachsen des Wertes einer Größe dazu führt, daß auf diese Größe ein verzögter Einfluß zum ei-

genen weiteren Wachstum ausgeübt wird. Ein negativer Feedbackkreis dagegen liegt nach allgemeiner Auffassung dann vor, wenn die Erhöhung einer Größe verzögert zu einem diesem Anwachsen gegenläufigen Effekt führt.¹

Zur Diskussion des Feedbackbegriffes betrachten wir das in Abbildung 23.1 dargestellte Flußdiagramm eines MA-Modells. Es lassen sich zwei Feedbackkreise unterscheiden, die durch die Zahlen 1 und 2 gekennzeichnet werden.

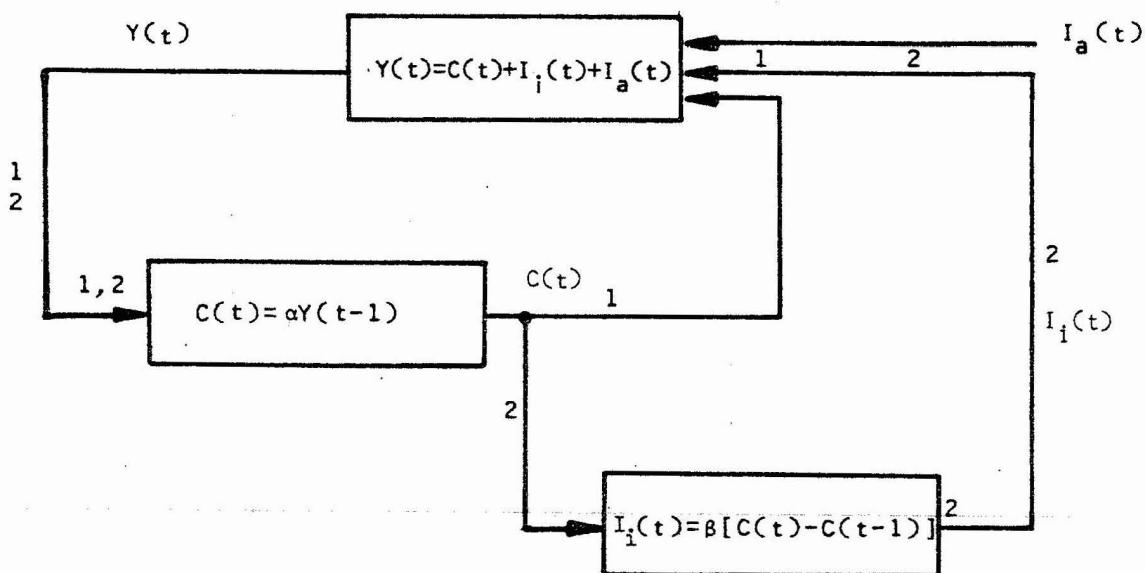


Abb. 23.1 Feedbackinterpretation eines MA-Systems

Wir wollen uns als erstes die Frage stellen: Kann man die zwei Feedbackkreise in 'positive' und 'negative' Feedbacks klassifizieren. Diese Frage wurde bereits bei der Besprechung komparativer Schaubildmodelle bejaht: im Falle des ersten Kreises wächst C mit Y , und ein wachsendes C erhöht wiederum Y . Für den zweiten Kreis gilt: wenn Y wächst, dann wächst auch C , wenn C wächst, wächst auch die Differenz $C(t) - C(t-1)$, was zur Folge hat, daß auch I_i wächst, welches wiederum Y erhöht.

Da das Modell keine nichtkomparativen Hypothesen enthält, erweisen sich beide Feedbackkreise als positiv.

¹ Vgl. Seite 59f.

Wir haben gesehen, daß die Existenz von Feedbacks in der Endgleichung einer endogenen Variablen durch autoregressive Beziehungen, d.h. der Abhängigkeit der endogen unverzögerten von ihren eigenen verzögerten Ausprägungen, zum Ausdruck kommt. Es liegt daher die Frage nahe, ob es möglich ist, zwischen der Existenz bestimmter positiver oder negativer Feedbackkreise und der Endgleichung Korrespondenzen festzustellen. Da die Endgleichung einer Variablen als die verdichtete Zusammenfassung der Auswirkung aller Feedbackkreise auf eine Variable anzusehen ist, fragt man sich vor allem, ob anhand einer Endgleichung zu erkennen ist, daß ein System ausschließlich negative oder positive oder beide Arten von Feedbackkreisen enthält. Ließen sich solche Beziehungen aufweisen, dann hieße dies, daß das Vorliegen einer dieser drei Konstellationen von Feedbackkreisen eine bestimmte strukturelle Beschaffenheit der betreffenden Endgleichung zur Folge hätte. Beispielsweise böte es sich an, zu überprüfen, ob im Falle nur positiver Feedbacks alle Koeffizienten der Endgleichung positiv sind, während sie im Falle negativer Feedbacks nur negative Werte annehmen und im gemischten Fall sowohl negative als auch positive Koeffizientenausprägung besitzen. Eine solche Feststellung würde dazu führen, daß der empirische Gehalt derartiger Feedbackbehauptungen durch die in einer Endgleichung 'verbotenen' Parameterausprägungen expliziert wird und mit dem empirischen Gehalt der angestrebten parametrisch-singulären Endgleichung verglichen werden kann.²

Ein Blick auf die Endgleichungen des MA-Modells zeigt jedoch, daß die angeführten Hypothesen nicht zutreffen, denn obgleich im System nur positive Feedbacks festzustellen waren, existiert ein negativer Koeffizient. Man kann zeigen, daß es nicht möglich ist, aus der Kenntnis bestimmter Konstellationen positiver und negativer Feedbackkreise in zwingender Weise auf irgendeine strukturelle Beschränkung der Koeffizienten einer Endgleichung zu schließen.³

Eine abschließende Betrachtung im Lichte dieser Formalisierung soll der manchmal vertretenen Ansicht gelten, ein positives Feedback führe zur Explosion eines Systems, während im Falle eines negativen Feed-

² Zur Beurteilung des empirischen Gehalts, siehe Seite 38ff.

³ Mit Ausnahme einer Differenzengleichung ersten Grades

backs dagegen ein bestimmter Gleichgewichtszustand realisiert werden würde.

In dieser undifferenzierten Form kann eine solche Auffassung (, der man allgemein in den Sozialwissenschaften nicht widersprechen soll, weil sie niemandem schadet) bei ihrer 'Rückübertragung' auf dynamische Systemmodelle zu Fehleinschätzungen führen.

Ein positiver Feedback ist zwar notwendig, aber nicht hinreichend für das explosive Wachstum einer Variablen. Betrachten wir den einfachsten Fall eines im Gleichgewicht befindlichen linearen Systems ersten Grades, welchem ein Einheitsimpuls $E^*(t)$ aufgeprägt wird.

$$Y(t) = \alpha Y(t-1) + 100 + E^*(t) \quad Y(-1) = \frac{100}{(1-\alpha)} \quad (23.7)$$

Das System besitzt ein positives Feedback, wenn $\alpha > 0$, aber nur im Fall $\alpha > 1$ tritt ein unbegrenztes Wachstum auf. Im Gegensatz dazu kann man im Falle $\alpha < 0$ von einem negativen Feedback sprechen. Eine Stabilität dieses Systems, d.h. eine Rückführung zum Gleichgewicht ist aber dennoch nur im Falle $0 < \alpha < 1$ gegeben, während das System für $\alpha < -1$ fluktuierend explodiert.

Zusammenfassend kann man sagen: Die Begriffe eines positiven und negativen Feedbacks können zur Kennzeichnung bestimmter rückführender Beeinflussungsketten von Variablen verwendet werden, wenn alle Hypothesen in diesen Ketten komparativ sind. Über die (unter Umständen mögliche) Klassifizierung von Feedbackkreisen hinausgehend, erlaubt die Feststellung solcher Eigenschaften von Feedbackkreisen keine Schlußfolgerung, die zur Einschränkung des potentiell möglichen Verhaltens von Systemen führt. In stärker formaler Sprachweise ausgedrückt heißt dies: aus den komparativen Hypothesen eines Systemmodells kann man keine Aussage über strukturelle Beschränkungen der minimalen parametrisch-generellen Hypothesen desselben Systemmodells ableiten. Die vorangegangene Veranschaulichung dieser recht einsichtigen Folgerung erscheint mir notwendig zu sein, um den Stellenwert solcher in der systemtheoretischen Literatur behandelter paradigmatisch überfrachteter Begriffe wie positiver und negativer Feedbackkreis klarer beurteilen zu können.

Die vorangehenden Betrachtungen über Feedbackkreise betrafen Hypothesensysteme, welche durch zyklische Modelle beschrieben werden, d.h. Modelle, deren Endgleichungen oder minimale Hypothesen Autoregressio-nen enthalten. Ein potentieller Opponent wird einwenden, daß die ver-wendete Definition eines zyklischen Modells nicht als operationales Kriterium geeignet ist, um zu entscheiden, ob ein vorliegendes Mo-dell zur Klasse der zyklischen Modelle gehört oder nicht. Denn zum einen ist es praktisch unmöglich, die Endgleichung nichtlinearer Mo-delle zu ermitteln, und zum anderen ist die Ermittlung der Endglei-chung linearer Modelle ein mühseliges Unterfangen.

Diesem Einwand ist zuzustimmen. Die entwickelte Definition lieferte zwar einen fruchtbaren Ansatzpunkt zur Diskussion bestimmter Arten von Feedbackbeziehungen, sie ist jedoch als Kriterium zur begriffli-chen Klassifizierung ungeeignet. Statt dessen soll im folgenden die geänderte Definition verwendet werden, daß ein zyklisches Modell im-mer dann vorliegt, wenn gezeigt werden kann, daß wenigstens eine en-dogene unverzögerte Variable des Modells zumindest von einem ihrer eigenen Variablenwerte verzögert beeinflußt wird, wobei nichts über den Grad der Verzögerung und die konkrete funktionale Form der Ab-hängigkeit gesagt wird. Eine solche weniger scharfe Definition ist jedoch für die Kennzeichnung von zyklischen Modellen voll ausreichend. Denn die in einer Endgleichung zusätzlich zum Ausdruck kommenden In-formationen über den Grad der Verzögerung und die konkrete funktiona-le Form sind für die Klassifizierung ohne Belang. Zur Beurteilung ei-nes konkreten Modells läßt sich die Existenz von Feedbackkreisen da-durch feststellen, daß man mit einem formalen Verfahren auf 'Schlei-fensuche' geht, d.h. die Modellgleichungen daraufhin untersucht, ob sie Beeinflussungsketten enthalten, die sich zu einer vorgegebenen Ausgangsvariablen wieder zurückführen lassen. Eine derartige Schlei-fensuche kann daher für lineare und nichtlineare Modelle gleicher-maßen durchgeführt werden. Die Beschreibung dieses Verfahrens erfolgt erst, nachdem die auch für andere Fragestellungen notwendigen Elemen-te der Modellstrukturanalyse entwickelt wurden.⁴

⁴ Vgl. Seite 360f.

B. Sequentielle Hypothesen

Die Abgrenzung zwischen zyklischen und kaskadierenden Modellen diente nicht nur dazu, die zyklischen Modelle und ihre empirischen Kennzeichen hervorzuheben. Unter den vielfältigen Formen von kaskadierenden Hypothesen (, deren Name sich daraus herleitet, daß die vorherbestimmten Variablen wie Kaskaden verzögert die zu erklärenden endogenen Variablen beeinflussen) verdient eine Hypothesenart, die als sequentielle Hypothese bezeichnet werden soll, besondere Beachtung.

In einer sequentiellen Hypothese hängt die zu erklärende endogene Variable allein von den mit einem Gewichtungsfaktor versehenen additiv miteinander verknüpften verzögerten und unverzögerten Einwirkungen einer Variablen ab.

Die allgemeine Form dieser Hypothese folgt mit $n=0$ aus (12.11)

$$Y(t) = g_0 E(t) + g_1 E(t-1) + g_2 E(t-2) + \dots + g_s E(t-s) \quad (23.8)$$

Neben den sequentiellen Hypothesen mit endlich vielen Koeffizienten kann man auch von der Annahme einer unendlichen Gewichtungsfolge g_0, g_1, \dots ausgehen. Entsprechend soll zwischen finiten und infiniten sequentiellen Hypothesen unterschieden werden.

Als finite sequentielle Hypothese kann die bereits erwähnte Werbeertragsfunktion (23.5) angeführt werden. Eine infinite sequentielle Hypothese kann beispielsweise im analogen Fall einer Werbeertragshypothese die Form

$$U(t) = \sum_{n=0}^{\infty} 15 \left[\left(\frac{1}{36} \right) \left(1+n \right) \left(\frac{5}{6} \right)^n W(t-n) \right] \quad n=0,1,2,\dots \quad (23.9)$$

besitzen. Die Gewichtungskoeffizienten sequentieller Hypothesen können sowohl negative als auch positive Werte umfassen. Auch ist es denkbar, daß die (unendliche) Summe der Koeffizienten infiniter Modelle unendlich ist.

Im Rahmen der finiten und infiniten primären sequentiellen Hypothesen spielen die sogenannten verteilten Verzögerungshypothesen die wichtigste Rolle. Der Charakter dieser Hypothesenform soll im folgenden gekennzeichnet werden.

Von einer verteilten Verzögerungshypothese kann gesprochen werden,

wenn eine sequentielle Hypothese die folgende Form annimmt,

$$Y(t) = M[w_0 E(t) + w_1 E(t-1) + w_2 E(t-2) + \dots] \quad (23.10)$$

wobei folgende Restriktionen gefordert werden:

- 1.) $w_n \geq 0$ für $n=0,1,2,\dots$
- 2.) $\sum_{n=0}^{\infty} w_n = 1$ (23.11)
- 3.) $M > 0$

Die erste Forderung verlangt nichtnegative Gewichtungsfaktoren, während durch die zweite Forderung zum Ausdruck kommt, daß die Summe der Gewichtungsfaktoren Eins sein soll. Mit der dritten Forderung wird ein positives M gefordert.

Die Werbeertragshypothese (23.5) läßt sich durch entsprechende Normierung auf die in (23.10) verlangte Form überführen, d.h.

$$U(t) = 15[0,1W(t) + 0,4W(t-1) + 0,3W(t-2) + 0,2W(t-3)] \quad (23.12)$$

Verteilte Verzögerungshypothesen bringen bildhaft gesprochen einen Vorgang zum Ausdruck, in welchem eine Eingangsgröße E nach einem festen Schlüssel erst verzögert zur Wirkung gelangt, und ihre ausgelöste Wirkung das M -fache des Betrages ihrer verzögerten Komponenten darstellt.

Die Größe M kann als der Totalmultiplikator der verteilten Verzögerungshypothese bezeichnet werden, während die Ausdrücke Mw_i Verzögerungsmultiplikatoren genannt werden.

Es liegt nahe, nach Kenngrößen zu fragen, die den Verlauf der Gewichtungskoeffizienten zum Ausdruck bringen. Da die Gewichtungskoeffizienten den Bedingungen 1) und 2) von (23.11) genügen, entsprechen sie formal einer Wahrscheinlichkeitssverteilung. Kenngrößen einer Wahrscheinlichkeitsverteilung wie das arithmetische Mittel und die Varianz sind daher zur Charakterisierung des Verlaufes der Gewichtungskoeffizienten analog anwendbar. Die durchschnittliche Verzögerungsdauer, oder kürzer, die durchschnittliche Verzögerung einer verteilten Verzögerung ist mit

$$D = 0w_0 + 1w_1 + 2w_2 + \dots + nw_s = \sum_{i=0}^s iw_i \quad (23.13)$$

definiert. Sie entspricht dem Erwartungswert einer Verteilung der Elemente der Zahlenmenge $0,1,2,\dots,s$, welche mit einer Wahrscheinlichkeit von $w_0, w_1, w_2, \dots, w_s$ auftreten. Die durchschnittliche Verzögerung liefert ein Maß für die Schnelligkeit, mit der Änderungen der Eingangsgröße auf die Ausgangsgröße durchschlagen. Daraus folgt: Je geringer die durchschnittliche Verzögerung, umso geringer ist der zum Tragen kommende Verzögerungseffekt.

Wenden wir die eingeführten Begriffe auf die in Gleichung (23.12) beschriebene verteilte Verzögerungshypothese an, dann zeigt sich: Der totale Multiplikator ist 15, es gibt vier Verzögerungsmultiplikatoren nullten bis dritten Grades 1,5; 6,0; 4,5 und 3,0. Die durchschnittliche Verzögerung beläuft sich auf

$$D = 0,1 \cdot 0 + 0,4 \cdot 1 + 0,3 \cdot 2 + 0,2 \cdot 3 = 1,6 \quad [\text{Perioden}]$$

Verteilte Verzögerungshypothesen können nicht direkt beobachtet werden. Daher sind sowohl die Folge der Gewichtungskoeffizienten als auch der Totalmultiplikator anhand der vorliegenden Zeitreihen für den Ein- und Ausgang der Verzögerung statistisch zu schätzen. Geht man von der Annahme einer endlichen Gewichtungsfolge aus, so können die Koeffizienten der verteilten Verzögerungshypothese (23.10) mit Hilfe von Regressionen statistisch geschätzt werden. Man unterstellt in diesem Fall, daß ein stochastischer Prozeß vorliegt, der durch

$$Y(t) = g_0 E(t) + g_1 E(t-1) + \dots + g_s E(t-s) + \varepsilon(t) \quad (23.14)$$

beschrieben wird. Von diesem stochastischen Prozeß wird weiter angenommen, daß die stochastische Variable $\varepsilon(t)$ den Erwartungswert Null besitzt, und ihre Realisationen in den einzelnen Perioden stochastisch unabhängig sind.

Für die Durchführung der Schätzung ist es notwendig, daß die Variablen $E(t-n)$ keine Multikolinearität untereinander aufweisen. Dies ist, wie die Erfahrung zeigt, bei einer Zeitreihe oft nicht zutreffend. Aus diesem Grund führt die Bestimmung der Parameter verteilter Verzögerungshypothesen mit Hilfe von Kleinstquadratschätzungen oft zu unzureichenden Schätzergebnissen.

Zur Vereinfachung des Schätzprozesses hat man sich bemüht, Verfahren zu entwickeln, die die Anzahl der zu schätzenden Parameter reduzieren.⁵

Dies versucht man dadurch zu erreichen, daß von vornherein unterstellt wird, die betreffende verteilte Verzögerungshypothese gehöre einer bestimmten Klasse von Verzögerungsformen an, die sich dadurch auszeichnet, daß relativ wenige Parameter geschätzt zu werden brauchen.

Eine oft verwendete Klasse von verteilten Verzögerungshypothesen bilden die Almon-Verzögerungen. [4, S.179,196]

Man geht von der Annahme aus, daß die Gewichtskoeffizienten

w_0, w_1, \dots, w_s in

$$Y(t) = M[w_0 E(t) + w_1 E(t-1) + \dots + w_s E(t-s)] \quad (23.15)$$

durch ein Polynom niedrigen Grades beschrieben werden, d.h. man unterstellt

$$w_i = b_0 + b_1 i + b_2 i^2 + \dots + b_p i^p \quad i=1,2,\dots,s$$

p ist hierbei ein vorzugebender Parameter. Setzt man diesen Ausdruck in (23.15) ein, so erkennt man, daß das Schätzproblem auf die Bestimmung der $p+1$ Parameter b_0, b_1, \dots, b_p reduziert ist. Das Verfahren zielt darauf ab, daß die Parameterzahl $p+1$ wesentlich kleiner sein soll als s . In der praktischen Anwendung muß man verschiedene Schätzungen mit variierenden Parametern p und s durchführen und die geeignete Schätzung auswählen. Eine andere Gruppe von Verfahren beruht darauf, infinit sequentielle Hypothesen in zyklische umzuformen. Auf diese Verfahren wird in Abschnitt B dieses Kapitels eingegangen.

Werden verteilte Verzögerungshypothesen in einem primären Ansatz verwendet, dann ist es oft schwierig, eine theoretische Rechtfertigung für den Gebrauch einer bestimmten Verzögerungsform oder zumindest für die nicht diskutierte Einschränkung auf eine bestimmte Klasse von Verzögerungen zu liefern. Während ein Modellentwickler in der Regel von plausiblen parametrisch-generellen Hypothesen ausgeht, deren

⁵ Zur weitergehenden Erörterung siehe DHRYMES [37].

numerische Spezifizierung anhand statistischer Schätzungen durchgeführt wird, läßt sich die Wahl der generellen Verlaufsform einer verteilten Verzögerung nur selten plausibel begründen.

Im folgenden soll eine besondere Form einer verteilten Verzögerungshypothese beschrieben werden, die zur Beschreibung der Beziehungen zwischen den Zu- und Abgängen von Beständen verwendet werden kann.

Unter einem Bestand versteht man eine bestimmte Menge gleichartiger, in der Realität aufweisbarer Elemente. Im allgemeinen zeichnen sich Bestände durch zumindest einen Zu- und Abgang aus, so daß ihre Entwicklung durch die übliche Bestandsfortschreibungsgleichung

$$B(t) = B(t-1) + Z(t) - A(t) \quad (23.16)$$

beschrieben werden kann. Im Hinblick auf das Ziel, die Welt mit Hilfe dynamischer Modelle zu beschreiben, lassen sich Bestände als eine Art 'Verzögerungsspeicher' interpretieren, denn ihre Existenz resultiert aus dem Umstand, daß die ihnen zufließenden Elemente den Bestand verzögert wieder verlassen.

Als Beispiel sei der Lagerbestand eines Betriebes angeführt: Über die Bestellungen werden dem Lager Güter zugeführt, die dort eine gewisse Zeit auf Lager bleiben, um dann zur Fertigung weitergeleitet zu werden. Auch die Entwicklung des Bestandes an TBC-Patienten ist ein Beispiel für den verzögerten Durchlauf eines Stroms von Elementen. Hypothesen, die die Beziehungen zwischen dem Zu- und Abgang eines Bestandes beschreiben, werden als Verweilzeithypothesen bezeichnet.

Primäre sequentielle Verweilzeithypothesen gehen von der Annahme aus, daß die einem Bestand während einer Periode zugehenden Elemente nach festen Anteilen w_0, w_1, \dots in unterschiedlichen Perioden den Bestand wieder verlassen. Unter dieser Annahme erweisen sich die Verweilzeithypothesen als Teilklasse der durch (23.11) gekennzeichneten verteilten Verzögerungshypothesen mit dem Totalmultiplikator $M=1$.

Diese formale Definition reicht jedoch allein nicht aus, um von einer Verweilzeithypothese sprechen zu können. Zusätzlich muß es stets

möglich sein, die Variable Y in (23.10) als den Ausgang und die Variable E als den Eingang eines Bestandes interpretieren zu können.

Neben der Verweilzeithypothese (auch Bestandsübergangsfunktion genannt) kann daher stets eine Bestandsgleichung formuliert werden, welche bei nur einem Zu- und Abgang durch (23.16) dargestellt wird. Die verzögert wirkenden Beziehungen zwischen dem Zugang eines Bestandes Z und seinem Abgang A kann durch

$$A(t) = w_0 Z(t) + w_1 Z(t-1) + w_2 Z(t-2) + \dots$$

mit $w_i > 0$ und $\sum_{i=0}^{\infty} w_i = 1$

beschrieben werden.

Eine Verweilzeithypothese braucht nicht notwendigerweise in der Form einer primären sequentiellen Hypothese aufzutreten. Sie kann auch durch Ableitungen aus einem bestimmten primären Hypothesensystem als sekundäre zyklische oder auch sekundäre sequentielle Verweilzeithypothese gewonnen werden.

Ein solcher Fall liegt immer dann vor, wenn es möglich ist, durch bestimmte Verhaltengleichungen die 'Behandlung' der in einen Bestand eintretenden Elemente detaillierter zu modellieren. So könnte man versuchen, das Schicksal der in den 'Bestand an Aufträgen in der Fertigung' eintretenden Elemente zu erfassen. Wenn es dann gelingt, die modellierten Abläufe zur Plazierung, Zwischenlagerung und Ausführung der Aufträge durch eine einzige Beziehung zwischen dem Lagerzu- und -abgang zusammenzufassen, dann erhielte man die sekundäre Verweilzeithypothese dieses Bestandes.

Die Entwicklung einer derartigen sekundären Verweilzeithypothese soll an einem einfachen Lagerhaltungsmodell demonstriert werden.

Der Lagerbestand dieses Systems wird durch die übliche Definitionsgleichung

$$L(t) = L(t-1) + Z(t) - A(t) \quad (23.17)$$

beschrieben. Es wird angenommen, der Lagerdisponent bestimme die Lagerabgänge nach folgender Entscheidungsregel:

$$A(t) = \alpha L(t-1) + \beta [S(t) - L(t-1)] \quad \alpha, \beta > 0 \quad (23.18)$$

Sie besagt, daß im Falle der Übereinstimmung von Soll- und Istdlagerbestand $S(t)$ und $L(t-1)$ der Lagerabgang das α -fache des Bestandes der Vorperiode betragen soll. Existiert dagegen eine Abweichung zwischen dem Soll- und Istdlager, so wird der Lagerabgang um das β -fache dieser Differenz korrigiert.

Der Solllagerbestand $S(t)$ wird auf das dreifache des prognostizierten Lagerzuganges $P(t)$ festgelegt, d.h.

$$S(t) = 3P(t) \quad (23.19)$$

Der prognostizierte Lagerzugang $P(t)$ wird nach der Methode der exponentiellen Glättung bestimmt und damit nach

$$P(t) = P(t-1) + \gamma [Z(t-1) - P(t-1)] \quad (23.20)$$

Die sekundäre Verweilzeithypothese des Lagers ist ermittelt, wenn es gelingt, die Endgleichung von $A(t)$ zu berechnen.

Diese Berechnung liefert die zyklische Verweilzeithypothese

$$A(t) = a_1 A(t-1) + a_2 A(t-2) + b_1 Z(t-1) + b_2 Z(t-2) \quad (23.21)$$

$$\text{mit } a_1 = 2 - \gamma - \alpha + \beta \quad b_1 = \alpha - \beta + 3\beta\gamma \\ a_2 = -1 + \gamma + \alpha - \beta - \alpha\gamma + \beta\gamma \quad b_2 = -\alpha + \beta + \alpha\gamma - 4\beta\gamma$$

welche, wie sich nachprüfen läßt, mit $M = (b_1 + b_2) / (1 - a_1 - a_2)$ den Totalmultiplikator von 1 besitzt.

Das beschriebene Beispiel zeigt, daß es grundsätzlich zwei Möglichkeiten gibt, um die Verweilzeithypothese eines Bestandes zu ermitteln: entweder aus einem primären sequentiellen Ansatz oder durch Ableitung aus einem System von Verhaltengleichungen. Wenn man sich die Frage stellt, welches der beiden Hypothesengewinnungsverfahren das 'bessere' sei, so wird mit Recht der Einwand erhoben, daß es für die Beurteilung einer Verweilzeithypothese völlig unmaßgeblich sei, auf welche Weise sie gewonnen wurde, entscheidend sei nur ihre empirische Gültigkeit. Dieser Einwand verkennt aber, daß mit 'besser' nicht die empirische Gültigkeit angesprochen wird, sondern ein Kriterium,

welches man als den Explikationsgrad der kontrollierbaren Parameter bezeichnen könnte. Wird die Verweilzeithypothese (23.20) aus den Gleichungen (23.17) bis (23.19) gewonnen, so sind die in diesen Gleichungen auftretenden Parameter als Größen anzusehen, die prinzipiell von bestimmten Personen kontrolliert werden. Der Explikationsgrad der kontrollierbaren Parameter wäre in diesem Falle 1. Dies hätte zur Folge, daß man im Rahmen einer (hier allein anstehenden) explikativen Modellverwendung Sensitivitätsanalysen im Hinblick auf bestimmte kontrollierbare Parameter vornehmen kann.

Im Falle eines primären sequentiellen Ansatzes hätte man nach einer statistischen Schätzung nur die numerischen Werte der Parameter w_1, w_2, w_3, \dots zur Verfügung, über deren auch nur indirekt aufweisbare Kontrollierbarkeit durch Personen aber nichts bekannt ist, so daß in diesem Fall der Explikationsgrad der kontrollierbaren Parameter Null ist. Man kann auch bezüglich dieser Parameter bestimmte Sensitivitätsanalysen durchführen, aber ihre Aussagekraft ist weniger informativ, da sie keine eindeutige Verknüpfung zu bestimmten Handlungsträgern und damit keine kontextuale Verankerung gestatten.

Da die Gewichtungskoeffizienten von primären sequentiellen Verweilzeithypothesen nur in seltenen Fällen eine Interpretation als kontrollierbare Parameter gestatten, ist es meistens erstrebenswerter, sie aus einem Hypothesensystem mit einem höheren Explikationsgrad der kontrollierbaren Parameter abzuleiten.

Primäre sequentielle Hypothesen werden daher auch oft nur verwendet, wenn die Ableitung von Verweilzeithypothesen nicht möglich ist. Denkt man zum Beispiel an die Verweilzeithypothese für den Bestand an TBC-Kranken, so würde ihre Ableitung aus einem Hypothesensystem mit einem hohen Explikationsgrad erfordern, daß ein System von Hypothesen zur Verfügung stände, in welchem die unterschiedliche Krankheitsdauer von Patientengruppen in Abhängigkeit von bestimmten prinzipiell kontrollierbaren Parametern zum Ausdruck käme. Da solche Hypothesen unbekannt sind, behilft man sich mit Hypothesen, die zwar einen Explikationsgrad der kontrollierbaren Parameter von Null besitzen, den primär angestrebten prognostischen Zwecken aber voll genügen.

2.3.2. Beziehungen zwischen linear zyklischen und infinit sequentiellen Hypothesen

Zyklische und sequentielle lineare Hypothesen sind nicht generell miteinander unverträglich. Es ist vielmehr unter Umständen möglich, eine primäre infinit sequentielle Hypothese in eine sekundäre zyklische Hypothese zu überführen. Die primäre infinite Werbeertragshypothese

$$U(t) = \sum_{\eta=0}^{\infty} 5 \left[\left(\frac{1}{12}\right) (1+\eta) \left(\frac{5}{6}\right)^{\eta} W(t-\eta) \right] \quad (23.22)$$

lässt sich beispielsweise in die sekundäre zyklische Hypothese

$$U(t) = \frac{5}{3} U(t-1) - \frac{25}{36} U(t-2) + \frac{5}{12} W(t) \quad t=0,1,2,\dots \quad (23.23)$$

umwandeln. (W : Werbeausgaben, U : Umsatz)

Gehen wir von zyklischen Hypothesen aus, dann gilt sogar, daß jede zyklische Hypothese in eine infinit sequentielle umgewandelt werden kann. So ist es beispielsweise möglich,

$$C(t) = (\alpha + \alpha\beta) C(t-1) - \alpha\beta C(t-2) + \alpha l_a(t-1)$$

mit $\alpha=0,972$ und $\beta=0,714$ in die infinite sequentielle Hypothese

$$C(t) = \sum_{\tau=0}^{\infty} (1+\tau) \left(\frac{5}{6}\right)^{\tau} 0,972 l_a(t-\tau-1)$$

zu transformieren. Der Transformationsprozeß zwischen den beiden Modellformen ist Gegenstand der folgenden Darstellung.

A. Überführung zyklischer in sequentielle Hypothesen

a) Gewichtsfunktion und Einheitsimpulsantwort

Jeder Einheitsimpuls, der einem zyklischen System aufgeprägt wird, bewirkt eine bestimmte Impulsantwort, die in Form einer unendlichen Folge c_0, c_1, c_2, \dots dargestellt werden kann.⁶ Einem derartigen System mit Null werdenden Anfangswerten soll im Zeitpunkt k ein beliebiger Eingang $E(\tau)$ mit $\tau=k, k+1, \dots$ aufgeprägt werden.

Der Eingang $E(k)$ bewirkt dabei aufgrund des Postulats der ungestörten Überlagerung linearer Systeme eine Antwort

⁶ Vgl. Seite 202f.

$$c_0 E(k), c_1 E(k), c_2 E(k), \dots$$

Der Eingang $E(k+1)$ bewirkt die Antwort

$$c_0 E(k+1), c_1 E(k+1), c_2 E(k+1), \dots$$

Ermittelt man in gleicher Weise die Antworten der folgenden Eingänge und ordnet die entsprechenden Komponenten der Antworten den Zeitpunkten ihres Auftretens entsprechend in einer Zeile an, so ergibt sich die folgende Matrix:

Zeit	k	k+1	k+2	...	k+s
	$c_0 E(k)$	$c_1 E(k)$	$c_2 E(k)$...	$c_s E(k)$
0		$c_0 E(k+1)$	$c_1 E(k+1)$
.		0	$c_0 E(k+2)$
.		.	0
.	
0		0	0	...	$c_0 E(k+s)$

Aufgrund der Beziehung $E(k-1)=E(k-2)=\dots=E(k-s)=0$ kann man die Felder unter der Hauptdiagonalen der Matrix mit Null werdenden Ausdrücken besetzen.

Zeit	k	k+1	k+2	...	k+s
	$c_0 E(k)$	$c_1 E(k)$	$c_2 E(k)$...	$c_s E(k)$
	$c_1 E(k-1)$	$c_0 E(k+1)$	$c_1 E(k+1)$...	$c_{s-1} E(k+1)$
	$c_2 E(k-2)$	$c_2 E(k-1)$	$c_0 E(k+2)$
.		$c_3 E(k-2)$	$c_3 E(k-1)$
.		.	$c_4 E(k-2)$
.	
	$c_s E(k-s)$	$c_s E(k-s+1)$	$c_s E(k-s+2)$...	$c_0 E(k+s)$
	$Y(k)$	$Y(k+1)$	$Y(k+2)$...	$Y(k+s)$

Die Summe jeder Spalte der Matrix stellt den Wert der endogenen Variablen $Y(k+i)$ mit $i=0,1,2,\dots,s$ dar. Sie ergibt sich aus

$$Y(k+i) = \sum_{\tau=0}^s c_\tau E[(k+i)-\tau]$$

Substituiert man $k+i=t$ und läßt s gegen unendlich laufen, folgt

$$Y(t) = \sum_{\tau=0}^{\infty} c_{\tau} E(t-\tau) \quad (23.24)$$

Die Folge c_{τ} kann als Gewichtsfunktion des Systems bezeichnet werden, weil sie angibt, in welcher Weise die um τ verzögerten Eingänge zur Bestimmung der Ausgangsgröße gewichtet werden. Wir erinnern uns jedoch, daß c_{τ} zugleich die Einheitsimpulsantwort des Systems darstellt. Damit zeigt sich, daß die Gewichtsfunktion eines linearen Systems mit dessen Einheitsimpulsantwort identisch ist.

Jedes lineare dynamische System läßt sich somit als ein Mechanismus interpretieren, durch den Eingangsgrößen mittels der Gewichtsfunktion in Ausgangsgrößen umgewandelt werden.

Mit diesen Ausführungen wird die Behauptung einsichtig, daß jede lineare zyklische Hypothese der Form

$$Y(t) = a_1 Y(t-1) + a_2 Y(t-2) + \dots + a_n Y(t-n) + E(t) \quad (23.25)$$

einer sequentiellen Hypothese der Form

$$Y(t) = \sum_{n=0}^{\infty} G(n) E(t-n) \quad (23.26)$$

äquivalent ist, wobei $G(n)$, d.h. die Gewichtsfunktion des sequentiellen Ansatzes (23.19) mit der Einheitsimpulsantwort des zyklischen Ansatzes (23.25) identisch ist.

Abbildung 23.2 zeigt beispielhaft den Ablauf dieses Transformationsprozesses.

Zur Illustration der sequentiellen Darstellung einer zyklischen Hypothese greifen wir auf unser Standardbeispiel eines MA-Systems zurück. Betrachten wir die Version $\alpha=0,972$, $\beta=0,714$, welche mit

$$\lambda_{1,2} = \frac{\alpha(1+\beta)}{2} \pm \sqrt{\frac{\alpha^2(1+\beta)^2}{4} - \alpha\beta}$$

zu einer gleichen Wurzel

$$\lambda^* = \frac{5}{6}$$

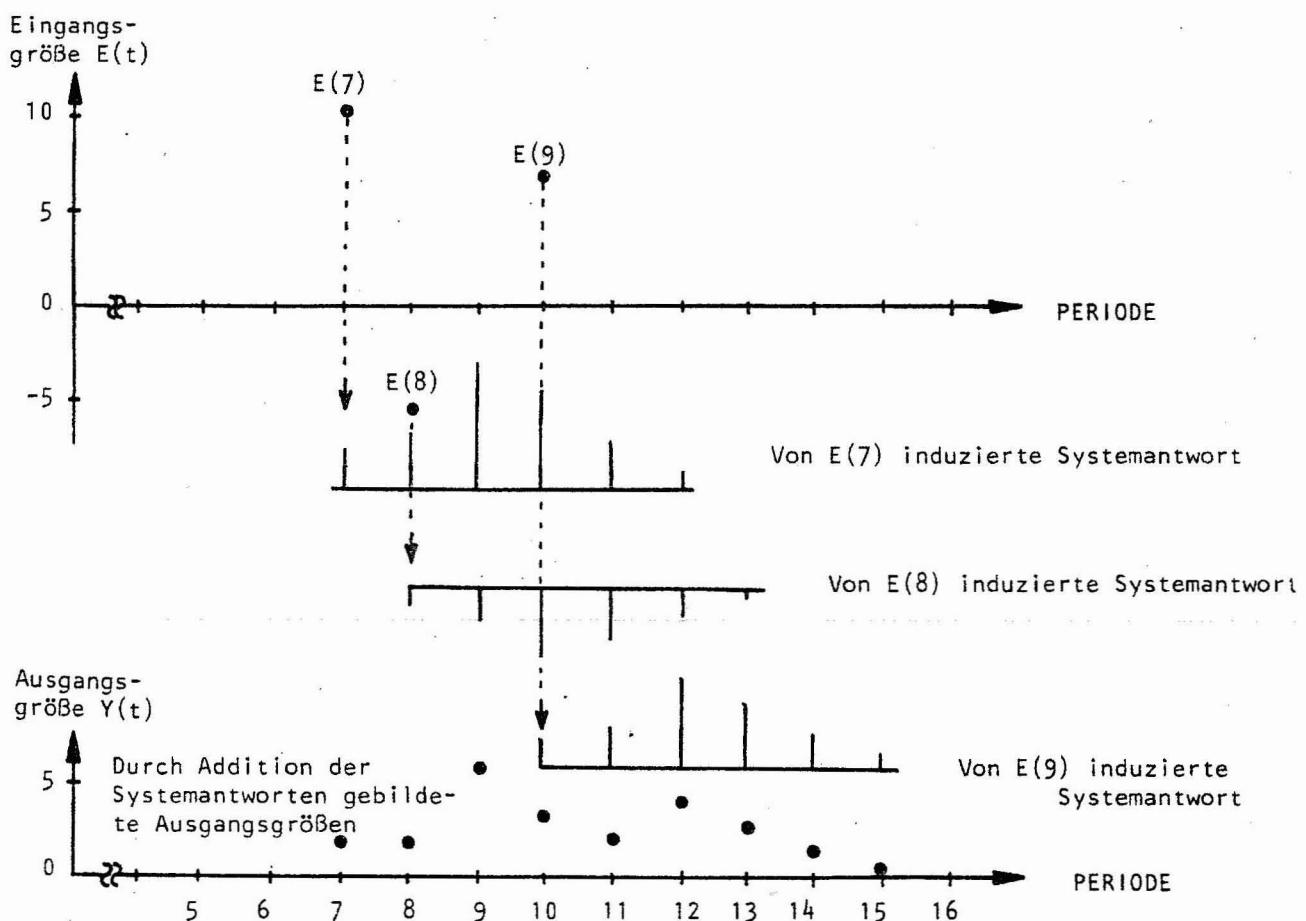


Abb. 23.2 Schematische Darstellung der Umwandlung einer Eingangsgröße in eine Ausgangsgröße mittels der Gewichtsfunktion eines linearen dynamischen Systems

und der Endgleichung

$$Y(t) = 1,666Y(t-1) - 0,69Y(t-2) + I_a(t) \quad (23.27)$$

führt. Die Funktionslösung der Einheitsimpulsantwort bestimmt sich

nach (21.76) durch

$$G(t) = (1+t)\lambda^t$$

und damit besitzt das System die Gewichtsfunktion:

$$G(\tau) = (1+\tau) \left(\frac{5}{6}\right)^{\tau} \quad \tau=0,1,2,\dots$$

Das MA-System wird daher im Falle verschwindender Anfangswerte durch die sequentielle Form:

$$Y(t) = \sum_{\tau=0}^{\infty} (1+\tau) \left(\frac{5}{6}\right)^{\tau} l_a(t-\tau) \quad t=0,1,2,\dots \quad (23.28)$$

beschrieben.

Die Behauptung der Äquivalenz zwischen einem zyklischen und sequentiellen Ansatz der Form (23.25) und (23.26), welche anhand des MA-Modells durch den Übergang von (23.27) auf (23.28) demonstriert wurde, erfolgt unter einer stillschweigenden Prämisse, welche im folgenden expliziert werden soll. Diese Prämisse bezieht sich auf den Fall, daß eine zyklische Hypothese der Form (23.25) Anfangswerte besitzt, die ungleich Null sind. Es läßt sich der Einwand erheben, daß in diesem Fall die Äquivalenzbehauptung zwischen (23.25) und (23.26) nicht gilt, weil (23.26) nur den Einfluß der Eingangsvariablen E auf das System berücksichtigt, nicht aber die Auswirkungen der von Anfang an im System 'gespeicherten' Anfangswerte.

Dieser Einwand läßt sich ausräumen, wenn man zeigen kann, daß durch die Vornahme einer Retrodiktion eine Folge von Eingängen $E(-1), E(-2), \dots, E(-m)$ rückprognostiziert werden kann, die exakt zu den in der zyklischen Hypothese vorgegebenen Anfangswerten $Y(-1), Y(-2), \dots, Y(-m)$ führt.

Dies ist gerade der Fall. Durch die Auflösung des Gleichungssystems

$$Y(-i) = \sum_{j=1}^m a_j Y(-j-i) + E(-i) \quad i=1,2,\dots,m$$

mit $Y(-m-i)=0$

nach $E(-i)$ erhält man eine exogene Variablensequenz, deren Verwendung im sequentiellen Ansatz die uneingeschränkte Gültigkeit der Äqua-

lenzbehauptung zur Folge hat.

Sind somit in der Endgleichung Y eines MA-Modells die Anfangswerte $Y(-1)$ und $Y(-2)$ vorgegeben, so ermittelt man die in ihrem sequentiellen Äquivalent zu verwendenden autonomen Investitionen $I_a(-1)$ und $I_a(-2)$ anhand des Gleichungssystems

$$\begin{aligned} Y(-1) &= (\alpha + \alpha\beta)Y(-2) - \alpha\beta Y(-3) + I_a(-1) \\ Y(-2) &= (\alpha + \alpha\beta)Y(-3) - \alpha\beta Y(-4) + I_a(-2) \end{aligned}$$

Als Ergebnis erhält man mit $Y(-3)=Y(-4)=0$

$$\begin{aligned} I_a(-1) &= Y(-1) - (\alpha + \alpha\beta)Y(-2) \\ I_a(-2) &= Y(-2) \end{aligned}$$

Wir haben bisher nur den Fall der Gewichtsfunktion eines zyklischen Modells betrachtet. Dieselbe sequentielle Darstellung eines Systems ist jedoch auch bei gemischt zyklisch sequentiellen Hypothesen möglich.

Als Beispiel unterstellen wir, daß die autonomen Investitionen $I_a(t)$ eines MA-Systems von den geplanten Investitionen der Unternehmer $I_p(t)$ durch die Beziehung

$$I_a(t) = 0,3I_p(t) + 0,7I_p(t-1)$$

beschrieben werden, denn naturgemäß existieren zeitliche Verzögerungen zwischen Planung und Realisierung einer Investition. Man erhält unter dieser Annahme die gemischt zyklisch sequentielle Minimalhypothese

$$Y(t) = (\alpha + \alpha\beta)Y(t-1) - \alpha\beta Y(t-2) + 0,3I_p(t) + 0,7I_p(t-1)$$

Für $\alpha=0,96$ und $\beta=0,667$ zeigt der Kurvenverlauf a in Abbildung 23.3 die Einheitsimpulsantwort des Systems bezüglich $0,3I_p(t)$, während der Verlauf b die Einheitsimpulsantwort bezüglich $0,7I_p(t-1)$ zum Ausdruck bringt. Die durch c gekennzeichnete Kurve, welche wegen des Prinzips der ungestörten Überlagerung durch die Ordinatenaddition der Verläufe a und b zustande kommt, ist die Gewichtsfunktion des Systems, und ihre Funktionslösung führt zu der sequentiellen Hypothese

$$Y(t) = \sum_{\tau=0}^{\infty} [1,175(1+\tau)(\frac{4}{5})^\tau - 0,875(\frac{4}{5})^\tau] I_p(t-\tau) \quad (23.29)$$

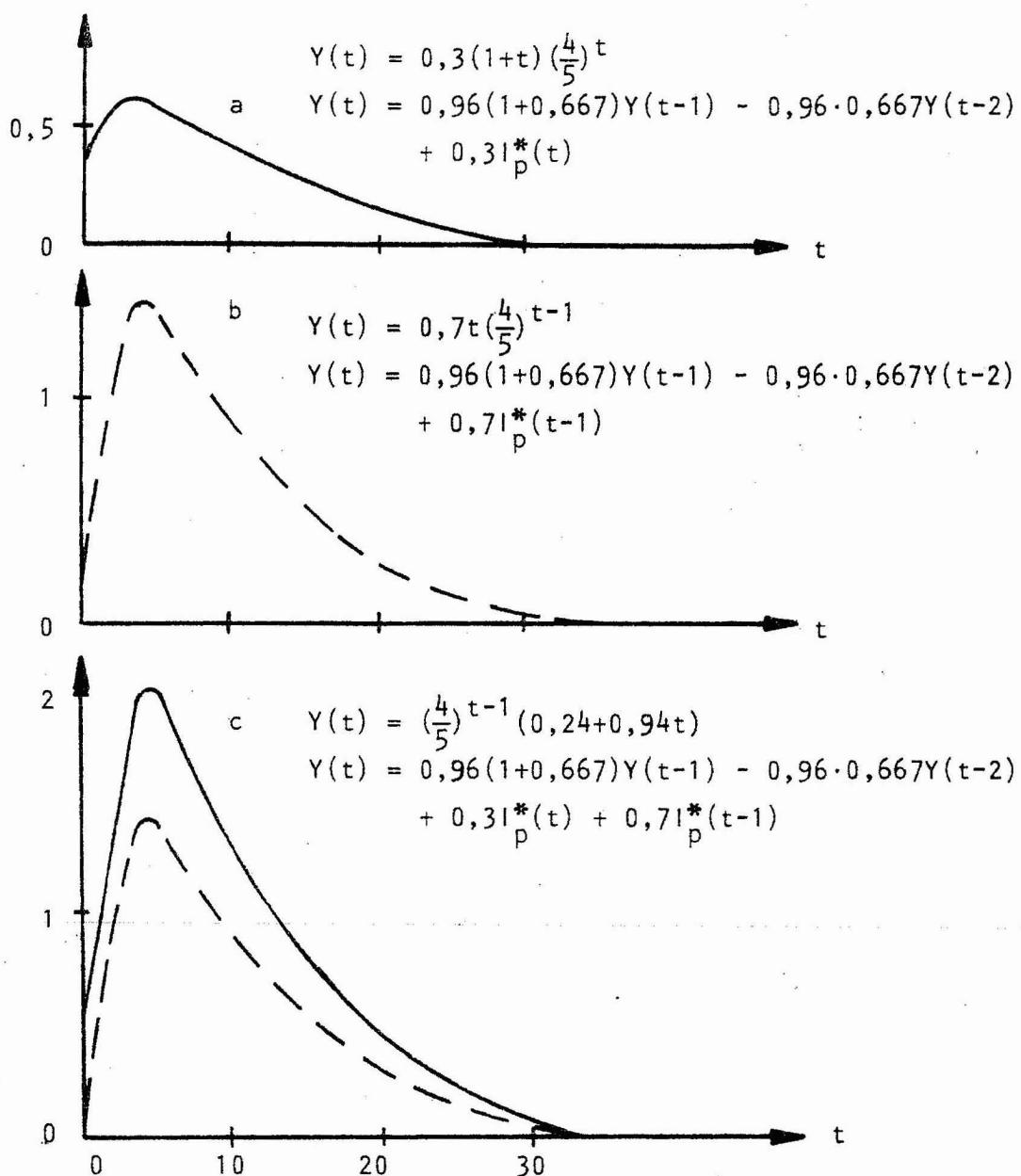


Abb. 23.3 Darstellung zur Ermittlung der Einheitsimpulsantwort oder Gewichtsfunktion eines gemischt zyklisch sequentiellen Modells

Wie erwähnt, ist es zur analytischen und computertechnischen Behandlung oft wünschenswert, dynamische Systeme in Form einer Zustandsraumdarstellung auszudrücken.⁷ Dies bedeutet, daß im Rahmen von Matrixengleichungen mit Verzögerungen ersten Grades eine für die gemischt zyklisch sequentielle Hypothese

⁷ Vgl. Seite 250

$$Y(t) + a_1 Y(t-1) + \dots + a_n Y(t-n) = g_0 E(t) + g_1 E(t-1) + \dots + g_s E(t-s)$$

äquivalente Zustandsraumdarstellung gefunden werden muß.

Eine solche Darstellung des gemischt zyklisch sequentiellen Ansatzes wird im Fall $n > s$ durch

$$\begin{bmatrix} z_1(t) \\ z_2(t) \\ \vdots \\ z_n(t) \end{bmatrix} = \begin{bmatrix} -a_1 & -a_2 & -a_3 & \dots & -a_{n-1} & -a_n \\ 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & 0 & \dots & 0 & 0 \\ \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\ 0 & 0 & 0 & \dots & 1 & 0 \end{bmatrix} \begin{bmatrix} z_1(t-1) \\ z_2(t-1) \\ \vdots \\ z_n(t-1) \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} E(t-1)$$

$$\text{mit } Y(t) = c_1 z_1(t) + c_2 z_2(t) + \dots + c_n z_n(t) + b_0 E(t)$$

$$\text{und } c_i = b_i - a_i b_0 \quad (i=1,2,\dots,n)$$

zum Ausdruck gebracht.

Als Beispiel wählen wir das eben erörterte MA-Modell mit $\alpha=0,96$ und $\beta=0,667$, d.h.

$$Y(t) = 1,6 Y(t-1) - 0,64 Y(t-2) + 0,3 I_p(t) + 0,7 I_p(t-1)$$

Die entsprechende Zustandsraumdarstellung dieses gemischt zyklisch sequentiellen Modells ergibt

$$\begin{bmatrix} z_1(t) \\ z_2(t) \end{bmatrix} = \begin{bmatrix} -1,6 & 0,64 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} z_1(t-1) \\ z_2(t-1) \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} I_p(t-1)$$

$$Y(t) = 1,18 z_1(t) + 0,192 z_2(t) + 0,3 I_p(t)$$

b) Ermittlung der Gewichtungsmatrix sequentieller Matrizenmodelle

Die Betrachtungen zur Umformung eines zyklischen in ein sequentielles Modell können auch im Rahmen der Matrizendarstellung linearer Modelle vorgenommen werden. Sie erhalten damit eine höhere Allgemeinheit, weil die gefundenen Erkenntnisse unabhängig von der Anzahl der auftretenden Variablen gelten.

Ausgehend von der Matrizengleichung

$$Y(t) = M Y(t-1) + E(t) \quad (23.30)$$

soll gezeigt werden, daß der Spaltenvektor $Y(t)$ der endogenen Varia-

blen eines linearen Modells als Funktion der Spaltenvektoren der exogenen verzögerten Variablen $E(t-\eta)$ dargestellt werden kann. Diese Darstellungsform soll als sequentielle Matrizenmodell bezeichnet werden.

Durch zeitliche Verzögerung um eine Periode erhält man aus Gleichung (23.30)

$$Y(t-1) = MY(t-2) + E(t-1) \quad (23.31)$$

Setzt man Gleichung (23.31) in (23.30) ein, dann folgt

$$Y(t) = M^2 Y(t-2) + ME(t-1) + E(t) \quad (23.32)$$

Aus (23.30) folgt

$$Y(t-2) = MY(t-3) + E(t-2) \quad (23.33)$$

Die Einsetzung von Gleichung (23.33) in Gleichung (23.32) liefert

$$Y(t) = M^3 Y(t-3) + M^2 E(t-2) + ME(t-1) + E(t)$$

bei w -facher Wiederholung dieses Verfahrens erhält man

$$Y(t) = M^{w+1} Y(t-w-1) + M^w E(t-w) + M^{w-1} E(t-(w-1)) + \dots + M^0 E(t)$$

Nehmen wir an, daß das System eine unendliche Vergangenheit hat, d.h. $w \rightarrow \infty$, dann strebt im Falle von Systemstabilität der Ausdruck $M^{w+1} Y(t-w-1)$ gegen Null. Denn gemäß (21.129) ist M^w (im Falle ungleicher Wurzeln) durch

$$M^{w+1} = S \begin{bmatrix} \lambda_1^{w+1} \\ \lambda_2^{w+1} \\ \vdots \\ \lambda_m^{w+1} \end{bmatrix} S^{-1}$$

darstellbar. Sind sämtliche Absolutwerte von λ kleiner als Eins, so streben die Elemente der Diagonalmatrix gegen Null. Unter diesen Umständen läßt sich $Y(t)$ durch

$$Y(t) = \sum_{\eta=0}^{\infty} M^{\eta} E(t-\eta) \quad (23.34)$$

formulieren. Diese sequentielle Darstellungsform eines linearen dynamischen Systems wird auch als Finalform bezeichnet. [203, S.138]

Die Matrix M^{η} soll als die Gewichtungsmatrix eines linearen Systems

bezeichnet werden. Zur Aufdeckung der mit dieser Gewichtungsmatrix eines Systems verbundenen Implikationen wäre es vorteilhaft, wenn man bei einem vorliegenden Modell die Elemente von M^n als Funktionen von n formulieren kann. Es stellt sich daher die Frage nach der Berechnung von M^n .

Zur Entwicklung einer einfachen Berechnungsmethode für M^n stützen wir uns auf das Cayley-Hamilton-Theorem. Es besagt:⁸

Satz 23.1: Ist

$$f(\lambda) = |M - \lambda I| = a_0 + a_1 \lambda + \dots + a_n \lambda^n = 0$$

und wird jede Potenz λ^i durch die Matrix M gleicher Potenz, d.h. M^i ersetzt, so ergibt sich eine Nullmatrix, d.h.

$$f(M) = a_0 + a_1 M + a_2 M^2 + \dots + a_n M^n = 0$$

Die Gültigkeit dieses Theorems sei an einem Beispiel demonstriert:

Die Matrix

$$M = \begin{bmatrix} 4 & 3 \\ 7 & 6 \end{bmatrix}$$

besitzt die charakteristische Gleichung

$$|M - \lambda I| = \begin{bmatrix} 4-\lambda & 3 \\ 7 & 6-\lambda \end{bmatrix} = (4-\lambda)(6-\lambda) - 21$$

oder

$$|M - \lambda I| = \lambda^2 - 10\lambda + 3$$

Aus dem Cayley-Hamilton-Theorem folgt

$$M^2 - 10M + 3I = 0$$

Mit

$$M^2 = \begin{bmatrix} 4 & 3 \\ 7 & 6 \end{bmatrix} \begin{bmatrix} 4 & 3 \\ 7 & 6 \end{bmatrix} = \begin{bmatrix} 37 & 30 \\ 70 & 57 \end{bmatrix}$$

ergibt sich

$$\begin{bmatrix} 37 & 30 \\ 70 & 57 \end{bmatrix} - 10 \begin{bmatrix} 4 & 3 \\ 7 & 6 \end{bmatrix} + 3 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} = 0$$

⁸ Zum Beweis siehe [227, S. 178f.]

Bevor wir uns der Ermittlung von M^n zuwenden, müssen wir uns mit dem Problem der Division von Polynomen befassen.

Betrachten wir den Ausdruck λ^t , so ist dieser als ein Polynom t -ten Grades anzusehen. Was passiert, wenn dieses Polynom durch das Polynom

$$f(\lambda) = a_0 + a_1\lambda + a_2\lambda^2 + \dots + a_n\lambda^n$$

dividiert wird? Für $t > n$ erhalten wir stets ein Quotientenpolynom $g_t(\lambda)$ und ein Restpolynom $h_t(\lambda)/f(\lambda)$, d.h.

$$\lambda^t/f(\lambda) = g_t(\lambda) + \frac{h_t(\lambda)}{f(\lambda)} \quad (23.35)$$

Wählen wir beispielsweise $t=3$ und $f(\lambda)=1+2\lambda+3\lambda^2$, so ergibt die Division

$$\begin{array}{r} \lambda^3 : (1+2\lambda+3\lambda^2) = \frac{1}{3}\lambda - \frac{2}{9} + \frac{\frac{1}{9}\lambda + \frac{2}{9}}{(1+2\lambda+3\lambda^2)} \\ - (\lambda^3 + \frac{2}{3}\lambda^2 + \frac{1}{3}\lambda) \\ \hline - \frac{2}{3}\lambda^2 - \frac{1}{3}\lambda \\ - (-\frac{2}{3}\lambda^2 - \frac{4}{9}\lambda - \frac{2}{9}) \\ \hline \frac{1}{9}\lambda + \frac{2}{9} \end{array}$$

womit $g_3(\lambda) = \frac{1}{3}\lambda - \frac{2}{9}$ und $h_3(\lambda) = \frac{1}{9}\lambda + \frac{2}{9}$ ist.

Es läßt sich zeigen, daß man bei einem beliebigen Polynom λ^i ($i=n+1, n+2, \dots$) und einem Polynom der Form $f(\lambda)=a_0+a_1\lambda+a_2\lambda^2+\dots+a_n\lambda^n$ stets zu einem Zähler $h_t(\lambda)$ des Restpolynoms gelangt, der höchstens ein Polynom $n-1$ -ten Grades bildet. Denn in dem Quotienten $h_t(\lambda)/f(\lambda)$ darf $h_t(\lambda)$ nicht mehr durch $f(\lambda)$ geteilt werden können. Diese Feststellung ist wichtig für das weitere Vorgehen.

Die Multiplikation von Gleichung (23.35) mit $f(\lambda)$ liefert

$$\lambda^t = g_t(\lambda)f(\lambda) + h_t(\lambda) \quad (23.36)$$

Schränkt man den Definitionsbereich von $f(\lambda)$ auf die Wurzeln $\lambda_1, \lambda_2, \dots, \lambda_n$ von $f(\lambda)$ ein, dann gilt

$$g_t(\lambda)f(\lambda) = 0$$

und aus Gleichung (23.36) folgt

$$\lambda^t = h_t(\lambda) \quad (23.37)$$

Der Zähler des Restpolynoms kann, wie erwähnt, nur ein Polynom $(n-1)$ -ten Grades sein, d.h.

$$h_t(\lambda) = b_0(t) + b_1(t)\lambda^1 + b_2(t)\lambda^2 + \dots + b_{n-1}(t)\lambda^{n-1}$$

Mit (23.37) wird damit

$$\lambda^t = b_0(t) + b_1(t)\lambda^1 + b_2(t)\lambda^2 + \dots + b_{n-1}(t)\lambda^{n-1} \quad (23.38)$$

und wegen des Cayley-Hamilton-Theorems

$$M^t = b_0(t)I + b_1(t)M^1 + b_2(t)M^2 + \dots + b_{n-1}(t)M^{n-1} \quad (23.39)$$

Die Beziehung (23.39) dient uns zur Berechnung von M^t . Es zeigt sich, daß jede Potenzmatrix M^t für alle $t=n, n+1, \dots$ durch eine Linearkombination der Einheitsmatrix und der ersten $n-1$ Matrixpotenzen beschrieben werden kann.

Da sämtliche Wurzeln der charakteristischen Gleichung $\lambda_1, \lambda_2, \dots, \lambda_n$ die Gleichung (23.38) erfüllen, kann man die Koeffizienten $b_0(t), b_1(t), \dots, b_{n-1}(t)$ durch die Auflösung des Gleichungssystems

$$\begin{aligned} \lambda_1^t &= b_0(t) + b_1(t)\lambda_1^1 + \dots + b_{n-1}(t)\lambda_1^{n-1} \\ \lambda_2^t &= b_0(t) + b_1(t)\lambda_2^1 + \dots + b_{n-1}(t)\lambda_2^{n-1} \\ &\vdots \\ \lambda_n^t &= b_0(t) + b_1(t)\lambda_n^1 + \dots + b_{n-1}(t)\lambda_n^{n-1} \end{aligned} \quad (23.40)$$

gewinnen, welche in Gleichung (23.39) eingesetzt M^t bestimmen.

Das ganze Verfahren soll am Beispiel des MA-Modells demonstriert werden. Die Zustandsraumdarstellung des MA-Modells bestimmte sich nach Gleichung (21.122) auf Seite 251.

$$\begin{bmatrix} Y(t) \\ I_i(t) \\ C(t) \end{bmatrix} = \begin{bmatrix} \alpha + \alpha\beta & 0 & -\beta \\ \alpha\beta & 0 & -\beta \\ \alpha & 0 & 0 \end{bmatrix} \begin{bmatrix} Y(t-1) \\ I_i(t-1) \\ C(t-1) \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} I_a(t)$$

Wählen wir $\alpha=0,72$ und $\beta=0,25$, so wird die Zustandsmatrix durch

$$M = \begin{bmatrix} 0,9 & 0 & -0,25 \\ 0,18 & 0 & -0,25 \\ 0,72 & 0 & 0 \end{bmatrix}$$

bestimmt. Die Gewichtungsmatrix M^t wird nach Gleichung (23.39) durch

$$M^t = b_0(t)I + b_1(t)M + b_2(t)M^2$$

ermittelt. Die Wurzeln der charakteristischen Gleichung von M berechnen sich mit $\lambda_1=0,3$, $\lambda_2=0,6$ und $\lambda_3=0$. Die mit t variierenden Koeffizienten $b_0(t)$, $b_1(t)$ und $b_2(t)$ bestimmen sich nach (23.40) durch die Lösung des Gleichungssystems

$$\begin{aligned}\lambda_1^t &= (0,3)^t = b_0(t) + 0,3b_1(t) + 0,09b_2(t) \\ \lambda_2^t &= (0,6)^t = b_0(t) + 0,6b_1(t) + 0,36b_2(t) \\ \lambda_3^t &= 0 = b_0(t)\end{aligned}$$

Die Auflösung dieses Gleichungssystems liefert

$$\begin{aligned}b_0(t) &= 0 \\ b_1(t) &= (-0,6)^t + 4(0,3)^t / 0,6 \\ b_2(t) &= ((0,6)^t - 2(0,3)^t) / 0,18\end{aligned}$$

Nach (23.39) folgt

$$\begin{aligned}M^t &= (-0,6)^t + 4(0,3)^t / 0,6 \begin{bmatrix} 0,9 & 0 & -0,25 \\ 0,18 & 0 & -0,25 \\ 0,72 & 0 & 0 \end{bmatrix} + \\ &+ ((0,6)^t - 2(0,3)^t) / 0,18 \begin{bmatrix} 0,63 & 0 & -0,0225 \\ -0,018 & 0 & -0,045 \\ 0,648 & 0 & -0,18 \end{bmatrix}\end{aligned}$$

Eine Zusammenfassung der Matrizen ergibt die endgültige Form der Gewichtungsmatrix

$$M^t = \begin{bmatrix} 2(0,6)^t - (0,3)^t & 0 & -\frac{5}{6}(0,6)^t + \frac{5}{6}(0,3)^t \\ -\frac{2}{5}(0,6)^t + \frac{7}{5}(0,3)^t & 0 & \frac{1}{6}(0,6)^t - \frac{7}{6}(0,3)^t \\ \frac{12}{5}(0,6)^t - \frac{12}{5}(0,3)^t & 0 & -(0,6)^t + 2(0,3)^t \end{bmatrix}$$

Das allgemeine Glied der unendlichen Reihe (23.34), d.h. $M^\eta E(t-\eta)$, wird durch

$$\begin{bmatrix} 2(0,6)^n - (0,3)^n & 0 & -\frac{5}{6}(0,6)^n + \frac{5}{6}(0,3)^n \\ -\frac{2}{5}(0,6)^n + \frac{7}{5}(0,3)^n & 0 & \frac{1}{6}(0,6)^n - \frac{7}{6}(0,3)^n \\ \frac{12}{5}(0,6)^n + \frac{12}{5}(0,3)^n & 0 & -(0,6)^n + 2(0,3)^n \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = I_a(t-n)$$

dargestellt. Die Ausdrücke in der ersten Spalte bilden die Gewichtsfunktionen von Y , I_i und C bezüglich des Eingangs I_a . Die Gewichtsfunktion (oder Einheitsimpulsantwort) von Y wurde für dieselben Werte für α und β bereits auf Seite 206 berechnet und stimmt, wie man sich überzeugen kann, mit der hier ermittelten überein.

B. Überführung infinit sequentieller in zyklische Hypothesen

Jede Endgleichung der Form

$$Y(t) + a_1 Y(t-1) + \dots + a_n Y(t-n) = E(t) \quad (23.41)$$

läßt sich, wie wir gesehen haben, in ihr sequentielles Äquivalent umwandeln, d.h. auf die Form

$$Y(t) = M[w_0 E(t) + w_1 E(t-1) + \dots] \quad (23.42)$$

Würde man zur Modellierung eines Zusammenhangs eine infinit sequentielle Hypothese verwenden, die sich in eine sekundäre zyklische Hypothese umwandeln ließe, dann hätte dies den Vorteil, daß die Bestimmung der Endgleichungen einfacher würde. Im Falle einer fast immer durchzuführenden Simulation, würde eine zyklische Hypothese zudem den Programmier- und Speicheraufwand vermindern, weil im Gegensatz zu einer sequentiellen Hypothese weniger verzögernde Variablen zu speichern wären.

Primäre sequentielle Hypothesen gehören fast ausschließlich zur Familie der verteilten Verzögerungshypothesen. Will man diese in zyklische Hypothesen umwandeln, so müssen die primären Hypothesen infinite Gewichtsfunktionen besitzen, die zum einen die in (23.11) formulierten Restriktionen erfüllen und zum anderen auch Gewichtsfunktionen einer zyklischen Hypothese sind.

Im folgenden soll, ausgehend von der allgemeinen Form einer zyklischen Hypothese, eine Klasse von Gewichtsfunktionen abgeleitet werden, die die Restriktionen (23.11) einer verteilten Verzögerungshypothese erfüllen und damit als Gewichtungsfolge eines sequentiellen Ansatzes dienen können. Die Folge der Verzögerungsmultiplikatoren Mw_0, Mw_1, \dots einer verteilten Verzögerungshypothese hat mit (23.11) folgenden Forderungen zu genügen:

- (1) Jedes Glied der Gewichtungsfolge w_0, w_1, \dots soll positiv sein.
- (2) Die Gewichtungsfolge w_0, w_1, \dots soll eine Summe von 1 ergeben.
- (3) Der Totalmultiplikator M soll frei wählbar und größer Null sein.

Die Gewichtsfunktion einer zyklischen Hypothese erfüllt die erste Forderung, wenn sämtliche Wurzeln ihrer Funktionslösung zwischen Null und Eins liegen. Wir wollen uns bei unserer Suche nach einer geeigneten Gewichtsfunktion von vornherein auf diese Teilmenge beschränken, die sich wiederum in die Untermenge gleicher und ungleicher Wurzeln zerlegen lässt.

In einer weiteren Einschränkung sollen ausschließlich Minimalhypothesen oder Endgleichungen mit gleichen Wurzeln betrachtet werden.

Im ersten Schritt wollen wir aus der generellen Formulierung einer Endgleichung

$$Y(t) + a_1 Y(t-1) + \dots + a_n Y(t-n) = gE(t) \quad (23.43)$$

durch Spezialisierung eine Endgleichung mit identischen Wurzeln und einem Multiplikator von Eins entwickeln. Anhand dieser Endgleichung wird in einem zweiten Schritt das entsprechende sequentielle Äquivalent abgeleitet. Auf diese Weise gewinnen wir eine Zusammenstellung der primären sequentiellen Hypothese und ihres sekundären zyklischen Äquivalents.

Unter Verwendung des Operators $K^{n-\eta} Y(t) = Y(t-\eta)$ erhalten wir die Operatorengleichung

$$Y(t)[K^n + a_1 K^{n-1} + \dots + a_n K^0] = gE(t) \quad (23.44)$$

Das in eckigen Klammern stehende Polynom kann in seine Produktenform überführt werden, d.h.⁹

⁹ Siehe Seite 249

$$K^n + a_1 K^{n-1} + \dots + a_n K^0 = (K-\lambda_1)(K-\lambda_2)\dots(K-\lambda_n)$$

Da die Wurzeln $\lambda_1, \lambda_2, \dots, \lambda_n$ gleich sein sollen, folgt

$$K^n + a_1 K^{n-1} + \dots + a_n K^0 = (K-\lambda)^n$$

Der Ausdruck $(K-\lambda)^n$ läßt sich in seine sogenannte Binomialform überführen, d.h. [127, S.551].

$$(K-\lambda)^n = K^n - \binom{n}{1}\lambda K^{n-1} + \binom{n}{2}\lambda^2 K^{n-2} - \dots + (-1)^n \lambda^n K^0$$

Ersetzen wir das Polynom in Gleichung (23.44) durch diesen Ausdruck, so erhalten wir

$$Y(t)[K^n - \binom{n}{1}\lambda K^{n-1} + \binom{n}{2}\lambda^2 K^{n-2} - \dots + (-1)^n \lambda^n K^0] = gE(t) \quad (23.45)$$

Die Rücktransformation $Y(t-n)=K^{n-n}Y(t)$ führt zur generellen Form einer inhomogenen Endgleichung n-ten Grades mit gleichen Wurzeln λ .

$$Y(t) - \binom{n}{1}\lambda Y(t-1) + \binom{n}{2}\lambda^2 Y(t-2) - \dots + (-1)^n \lambda^n Y(t-n) = gE(t) \quad (23.46)$$

Der Totalmultiplikator der Endgleichung (23.44) bestimmt sich nach (21.82) mit

$$M = \frac{g}{1+a_1+a_2+\dots+a_n}$$

Der Vergleich der allgemeinen Endgleichungsform (23.43) mit dem Spezialfall gleicher Wurzeln λ ergibt:

$$1 + a_1 + a_2 + \dots + a_n = 1 - \binom{n}{1}\lambda + \binom{n}{2}\lambda^2 - \dots + (-1)^n \lambda^n = (1-\lambda)^n$$

Man erkennt, daß die Endgleichung (23.44) in diesem Fall immer dann einen Totalmultiplikator von 1 besitzt, wenn der freie Parameter g

$$g = (1-\lambda)^n$$

gewählt wird.

Mit dieser Festlegung wird die zweite an die Gewichtungsfolge einer verteilten Verzögerung zu stellende Forderung erfüllt. Da jedoch entsprechend der dritten Forderung in einer Familie der Gewichtungsfolgen von verteilten Verzögerungen der Totalmultiplikator M frei wählbar sein soll, muß g endgültig mit

$$g = M(1-\lambda)^n \quad (23.47)$$

gewählt werden.

Mit (23.47) in (23.46) erhält man mit (23.48) eine Familie von zyklischen Hypothesen, deren Gewichtsfunktionen den in (23.11) erhobenen Forderungen genügen.

$$Y(t) - \binom{n}{1} \lambda Y(t-1) + \binom{n}{2} \lambda^2 Y(t-2) - \dots + (-1)^n \lambda^n Y(t-n) = M(1-\lambda)^n E(t) \quad (23.48)$$

In einem zweiten Schritt wollen wir nunmehr das sequentielle Äquivalent der zyklischen Hypothese (23.48) entwickeln.

Verwenden wir den Operator $K^{-n}Y(t)=Y(t-n)$, so folgt aus (23.48)

$$Y(t)[1 - \binom{n}{1} \lambda K^{-1} + \binom{n}{2} \lambda^2 K^{-2} - \dots + (-1)^n \lambda^n K^{-n}] = M(1-\lambda)^n E(t)$$

oder

$$\sum_{\eta=0}^n \binom{n}{\eta} (-\lambda K^{-1})^\eta Y(t) = M(1-\lambda)^n E(t) \quad (23.49)$$

Aufgrund des Binomiallehrsatzes gilt

$$\sum_{\eta=0}^n \binom{n}{\eta} (-\lambda K^{-1})^\eta = (1-\lambda K^{-1})^n \quad (23.50)$$

(23.50) in (23.49) ergibt

$$(1-\lambda K^{-1})^n Y(t) = M(1-\lambda)^n E(t)$$

$$Y(t) = M(1-\lambda)^n (1-\lambda K^{-1})^{-n} E(t) \quad (23.51)$$

Der Ausdruck $(1-\lambda K^{-1})^{-n}$ kann als unendliche Reihe dargestellt werden

$$(1-\lambda K^{-1})^{-n} = \sum_{\eta=0}^{\infty} \binom{n+\eta-1}{\eta} (\lambda K^{-1})^\eta \quad (23.52)$$

Aus (23.51) und (23.52) folgt

$$Y(t) = (1-\lambda)^n \sum_{\eta=0}^{\infty} \binom{n+\eta-1}{\eta} (\lambda K^{-1})^\eta M E(t) \quad (23.53)$$

Die Rücktransformation $Y(t-\eta)=K^{-\eta}Y(t)$ liefert das sequentielle Äquivalent

$$Y(t) = (1-\lambda)^n \sum_{\eta=0}^{\infty} \binom{n+\eta-1}{\eta} \lambda^\eta M E(t-\eta)$$

Die Gewichtsfunktion ergibt sich mit

$$w_n = (1-\lambda)^n \binom{n+n-1}{n} \lambda^n \quad (23.54)$$

In Tabelle 23.1 sind die Gewichtsfunktionen der Differenzengleichungen ersten, zweiten, dritten und n-ten Grades einander gegenübergestellt.

Grad	Gewichtsfunktion
	Zyklisches Äquivalent
1	$w_n = (1-\lambda) \lambda^n$
	$Y(t) - \lambda Y(t-1) = (1-\lambda) E^*(t)$
2	$w_n = (1-\lambda)^2 (1+\eta) \lambda^n$
	$Y(t) - 2\lambda Y(t-1) + \lambda^2 Y(t-2) = (1-\lambda)^2 E^*(t)$
3	$w_n = (1-\lambda)^3 [(1+\eta)(2+\eta)/2] \lambda^n$
	$Y(t) - 3\lambda Y(t-1) + 3\lambda^2 Y(t-2) - \lambda^3 Y(t-3) = (1-\lambda)^3 E^*(t)$
n	$w_n = \binom{n+n-1}{n} (1-\lambda)^n \lambda^n$
	$Y(t) - \binom{n}{1} \lambda Y(t-1) + \binom{n}{2} \lambda^2 Y(t-2) - \dots + (-1)^n \lambda^n Y(t-n) = (1-\lambda)^n E^*(t)$

Tab. 23.1 Gewichtsfunktion sequentieller Ansätze und ihre zyklischen Äquivalente

Die auf diese Weise ermittelten Gewichtsfunktionen entsprechen einer Pascalverteilung. Dies ist eine in der Wahrscheinlichkeitstheorie verwendete Verteilung für nichtnegative ganzzahlige Zufallsvariablen.

Der Erwartungswert einer Pascalverteilung n-ten Grades berechnet sich nach¹⁰

$$\bar{w}_n = n\lambda / (1-\lambda) \quad (23.55)$$

Dieses Maß kann zugleich als eine Kenngröße für den Verlauf der Gewichtsfunktion dienen. Abbildung 23.4 zeigt den Verlauf der Gewichtsfunktion ersten bis dritten Grades mit einem Mittelwert von $\bar{w}=10$. Wie wir später sehen, werden die beschriebenen Gewichtsfunktionen dritten Grades von FORRESTER in seinem dynamischen Modellierungskon-

¹⁰ Vgl. [50, S. 202]

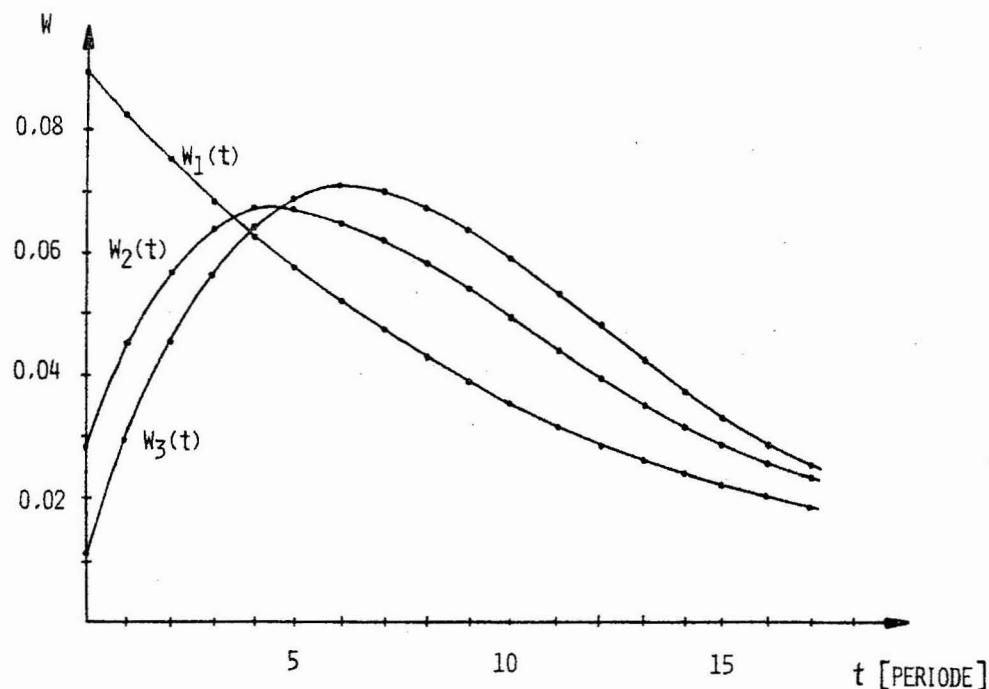


Abb. 23.4 Gewichtsfunktionen linearer Differenzengleichungen ersten bis dritten Grades bei gleichen Wurzeln und $\bar{w}=10$

zept System Dynamics in großem Umfang verwendet.

Die bisherigen Erörterungen lieferten uns eine durch (23.54) gekennzeichnete Familie von Gewichtsfunktionen, die den an verteilte Verzögerungshypothesen gestellten Forderungen (23.11) genügen und durch die zyklische Hypothese (23.48) zum Ausdruck kommen. Wir wollen uns im folgenden wieder der Frage zuwenden, ob und auf welche Weise eine Familie der Gewichtsfunktionen zyklischer Verzögerungshypothesen zur Modellierung primärer verteilter Verzögerungshypothesen verwendet werden kann. Als Ausgangspunkt dieser Erörterung dient uns die durch (23.54) gekennzeichnete Familie der Gewichtsfunktionen. Für die eigentliche Fragegestellung gehen wir von einer Klasse von Gewichtsfunktionen aus, die wir durch zwei Einschränkungen auf eine Teilklassen von (23.48) und eine Erweiterung dieser Teilklassen über (23.48) hinaus erreichen.

Die erste Einschränkung beruht darauf, daß wir nur die durch die Spezifizierung von $n=3$ sich in (23.48) ergebende Teilklassen von Gewichtsfunktionen untersuchen wollen. Es handelt sich um Gewichtsfunktionen

mit einem eingipfligen Verlauf, deren Verlaufsform für viele empirische Zusammenhänge plausibel erscheint. Abbildung 23.4 zeigt ein Exemplar dieser Teilklasse mit einer durchschnittlichen Verzögerung von 10 Perioden.

Die zweite Einschränkung bezieht sich auf die empirische Interpretation der verteilten Verzögerungshypothesen: es soll nur die Modellierung von Verweilzeithypothesen untersucht werden. Diese Einengung folgt aus der Absicht, im Rahmen der nachfolgenden Betrachtungen über die Simulation dynamischer Systeme ein Modellkonzept zu entwickeln, in welchem die Modellierung primärer sequentieller Verweilzeithypothesen eine wichtige Stellung einnimmt. Formal drückt sich diese Einschränkung durch die Wahl eines Totalmultiplikators von $M=1$ aus. Die Erweiterung des Repertoires an Gewichtsfunktionen ergibt sich durch die Einfügung von Totzeiten. Ersetzt man in (23.48) $E(t)$ durch $E(t-T)$, so hat dies zur Folge, daß die Variation von T zu unterschiedlichen Gewichtsfunktionen führt.

Die Auswirkung einer Totzeit läßt sich am anschaulichsten anhand der einer Gewichtsfunktion entsprechenden Impulsantwort eines Systems demonstrieren. Schickt man in ein System einen Einheitsimpuls, so wird die Zeit T , die vergeht bis der erste positive Wert der Einheitsimpulsantwort auftritt, als Totzeit bezeichnet. Die in Abbildung 23.4 dargestellten Gewichtsfunktionen besitzen eine Totzeit von 0 Perioden. Würde man diese Gewichtsfunktionen um eine Totzeit von T Perioden modifizieren, so hieße dies, daß die Verläufe um T Perioden nach rechts verschoben werden müßten:

Unter diesen Einschränkungen und Erweiterungen erhalten wir aus (23.48) die zyklische Verweilzeithypothese

$$A(t) = 3\lambda A(t-1) - 3\lambda^2 A(t-2) + \lambda^3 A(t-3) + (1-\lambda)^3 Z(t-T) \quad (23.56)$$

Aus mnemotechnischen Gründen wurden dabei die Variablen Y durch A und E durch Z ersetzt, so daß Z als Zugang und A als Abgang einer Bestandsgröße anzusehen sind.

Die Gewichtsfunktion von (23.56) ergibt sich mit (23.54) aus

$$w_n^3 = \begin{cases} 0 & \text{für } n=0,1,\dots,T-1 \\ (1-\lambda)^3 \binom{n-T+2}{n-T} \lambda^{n-T} & \text{für } n=T, T+1, \dots \end{cases} \quad (23.57)$$

Zur besseren Beurteilung von Hypothesen ist es erstrebenswert, daß ihre Parameter eine sinnvolle empirische Interpretation erlauben.

Aus diesem Grunde soll der Parameter λ in (23.57) durch zwei Parameter definiert werden, die sich als die durchschnittliche Verzögerung eines Elementes im Bestand (D) und als die Totzeit (T) der Gewichtsfunktion oder Systemantwort deuten lassen.

Besitzt eine Verweilzeithypothese die in (23.54) dargestellte Gewichtsfunktion w_n , dann kann die durchschnittliche Verzögerung eines den Bestand durchlaufenden Elementes mit

$$D^* = \sum_{n=0}^{\infty} w_n \cdot n \quad (23.58)$$

beschrieben werden. Im Falle der Erweiterung des Ansatzes (23.48) durch die Einführung einer Totzeit T kann die durchschnittliche Verzögerung D mit

$$D = \sum_{n=0}^{\infty} w_n (n+T)$$

berechnet werden. Hieraus folgt

$$D = \sum_{n=0}^{\infty} w_n \cdot n + T \sum_{n=0}^{\infty} w_n$$

und mit (23.11) sowie (23.58) ergibt sich

$$D = D^* + T \quad (23.59)$$

Da das in (23.55) angegebene arithmetische Mittel der Gewichtsfunktion mit der hier erörterten durchschnittlichen Verzögerung D^* identisch ist, gilt für $n=3$

$$D^* = 3\lambda/(1-\lambda) \quad (23.60)$$

Aus (23.59) und (23.60) folgt

$$\lambda = \frac{D-T}{D-T+3} \quad (23.61)$$

Da λ positiv und kleiner als 1 sein muß, gilt

$$D - T > 0$$

Mit (23.61) ist die geforderte Reduzierung von λ auf die durchschnittliche Verzögerung und die Totzeit vorgenommen.

Unter Berücksichtigung von (23.61) wird (23.57)

$$w_n^3 = \left\{ \begin{array}{ll} 0 & \text{für } n=0, 1, 2, \dots, T-1 \\ \left(1 - \frac{D-T}{D-T+3}\right)^2 \binom{n-T+2}{n-T} \left(\frac{D-T}{D-T+3}\right)^{n-T} & \text{für } n=T, T+1, \dots \end{array} \right. \quad (23.62)$$

2.4. Rekursive und interdependente Modellformen

Nach einer ersten intuitiven Klärung der formalen Struktur und der empirischen Deutungsmöglichkeit von rekursiven und interdependenten Modellen werden wir im zweiten Abschnitt Methoden diskutieren, mit denen man interdependente und rekursive dynamische Modelle beliebiger Größe voneinander unterscheiden und in übersichtlicher Weise gliedern kann. Abschließend wenden wir uns den besonderen Problemen der Zeitpfadermittlung im Rahmen interdependenten Modells zu.

2.4.1. Begriffliche Klärung und empirische Interpretation

Der grundsätzliche Unterschied zwischen einem rekursiven und einem interdependenten Modell kann anhand des Pfeilschemas in Abbildung 24.1 demonstriert werden.

In einem interdependenten Modell lassen sich zumindest zwei unverzögerte endogene Variablen finden, die eine wechselseitige Beeinflussung aufeinander ausüben. In Abbildung 24.1 beeinflußt Y_1 über Y_2 die Variable Y_3 und wird von dieser wiederum direkt beeinflußt. Eine derartige wechselseitige Beeinflussung kann durchaus über beliebig viele 'Stationen', d.h. andere endogene unverzögerte Variablen laufen. Entscheidend ist allein, daß eine geschlossene Beeinflussungskette vorhanden ist.

Anschaulich formuliert heißt dies, daß ein durch ein Pfeildiagramm gekennzeichnetes dynamisches System immer dann interdependent ist, wenn man für mindestens zwei unverzögerte endogene Variablen einen durch diese beiden Variablen führenden geschlossenen Beeinflussungspfad findet. Im Gegensatz dazu zeichnen sich rekursive dynamische Modelle durch die Abwesenheit eines solchen Beeinflussungspfades aus.

Als Folge davon kann man die Pfeildiagrammdarstellung eines rekursiven Systems durch Austauschen der Zeilen immer so umgestalten, daß alle Beeinflussungspfeile zwischen den endogenen unverzögerten Variablen in eine Richtung zeigen.

Tauscht man beispielsweise in Abbildung 24.1 die Anordnung der Variablen Y_1 und Y_2 miteinander aus, so wird eine solche 'Standardisierung' bewirkt.

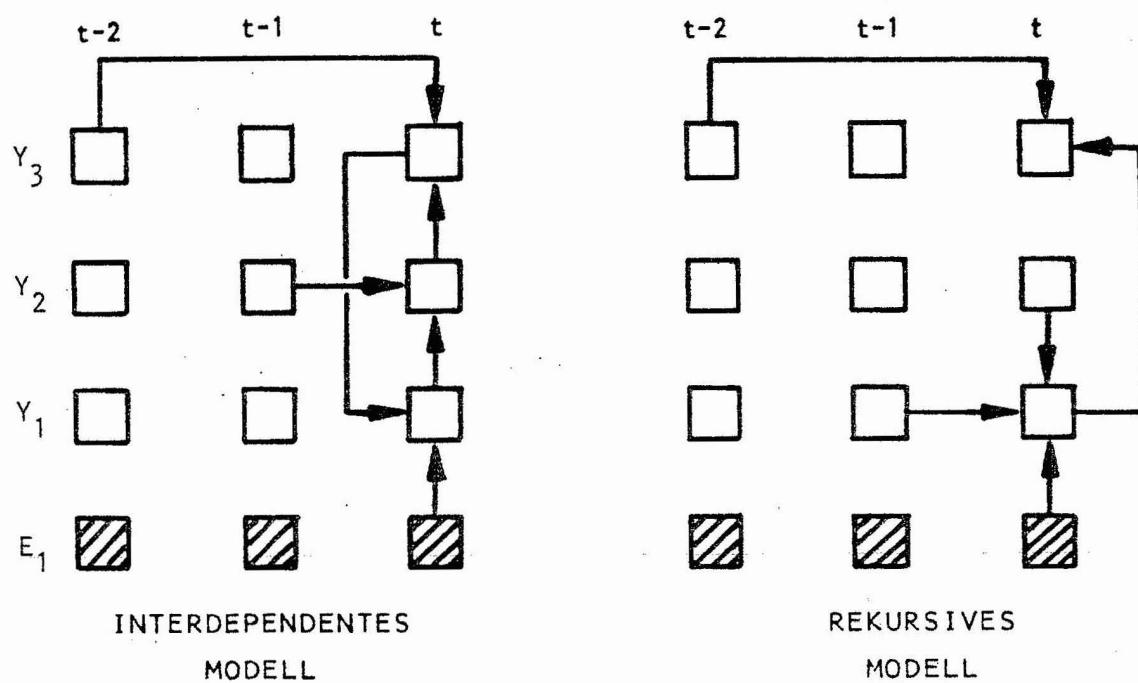


Abb. 24.1 Pfeildiagramm eines interdependenten und rekursiven dynamischen Modells

Bisher haben wir nur die formalen Unterschiede zwischen rekursiven und dynamischen Modellen erörtert. Es liegt jedoch nahe, daß auch die empirische Deutbarkeit der in den beiden Modelltypen auftretenden Hypothesen von unterschiedlicher Art sein wird.

Um dieser Frage nachzugehen, betrachten wir zwei modifizierte Versionen eines MA-Modells, welche durch

$$Y(t) = C(t) + I_a(t) + I_i(t) \quad (24.1)$$

$$I_i(t) = \beta[C(t) - C(t-2)] \quad (24.2)$$

$$C(t) = \alpha Y(t-1) \quad (24.3)$$

und

$$Y(n) = C(n) + I_a(n) + I_i(n) \quad (24.4)$$

$$I_i(n) = \beta[C(n) - C(n-1)] \quad (24.5)$$

$$C(n) = \alpha Y(n) \quad (24.6)$$

beschrieben werden.

Wie man leicht nachprüfen kann, stellt die erste Version ein rekursives, die zweite dagegen ein interdependentes Modell dar. Anhand dieser Beispiele wollen wir der Frage nachgehen, ob es bestimmte 'interdependente Modellhypotesen' gibt, die einen grundsätzlich anderen Charakter besitzen als die ihnen gegenüberzustellenden 'rekursiven Modellhypotesen'.

Hypothesen sind Wenn-Dann-Aussagen. Dem intuitiven Verständnis derartiger Wenn-Dann-Aussagen kommt es sehr entgegen, daß sie eine kausale Interpretation zulassen, d.h. die 'Wenn-Komponente' als Ursache einer Wirkung gedeutet werden kann, die in Form der 'Dann-Komponente' zum Ausdruck kommt. Rekursive Modelle werfen bei der Anwendung dieser Interpretation keine Probleme auf: die endogenen unverzögerten Variablen einer Hypothesengleichung können stets als 'Wirkung' der auf der rechten Seite der Gleichung stehenden 'Ursachen' gedeutet werden. Die Möglichkeit, in einem rekursiven Modell die endogenen unverzögerten Variablen in Form einer offenen Beeinflussungskette anzurufen, führt dazu, daß es bei rekursiven Modellen nie möglich ist, eine bestimmte endogene unverzögerte Variable als eigene Ursache ihrer Wirkung zu deuten. Gerade dies ist jedoch bei interdependenten Modellen der Fall.

Die Transitivität kausaler Beziehungen erlaubt die Behauptung, daß die Variable A(t) als die Ursache von C(t) bezeichnet werden kann, falls A(t) die Ursache von B(t) und B(t) die Ursache von C(t) ist. Betrachten wir unter dieser Deutung das interdependente Modell in Abbildung 24.1, so ist die Wirkung $Y_1(t)$ ihre eigene Ursache, weil $Y_1(t)$ die Ursache von $Y_2(t)$, $Y_2(t)$ die Ursache von $Y_3(t)$ und $Y_3(t)$ die Ursache von $Y_1(t)$ ist.

Die Version (24.4) bis (24.6) des MA-Modells zeigt die Schwierigkeiten einer kausalen Interpretation im Falle der Konsumfunktion.

Setzt man die Gleichung für $Y(\eta)$ in die Konsumfunktion ein, so ergibt sich:

$$C(\eta) = \alpha C(\eta) + \alpha I_a(\eta) + \alpha I_i(\eta)$$

d.h. die Dann-Komponente $C(\eta)$ auf der linken Seite der Gleichung ist zugleich Bestandteil der Wenn-Komponente.

Die Schwierigkeiten einer Kausalinterpretation führen zu der Frage, warum Modellentwickler überhaupt auf derartige Hypothesen kommen.

Diese Frage ist um so berechtigter, wenn sich zeigt, daß auch die mit interdependenten Modellen verbundenen Schätz- und Analyseverfahren aufwendiger als im Falle rekursiver Modelle sind. Gerade dies ist jedoch der Fall. Wenn dennoch interdependente Modelle verwendet werden, dann liegt der Grund darin, daß manche empirischen Zusammenhänge wegen Datenmangels allenfalls mit derartigen Modellformen erfaßt werden können.

Zur Verdeutlichung betrachten wir das folgende rekursive, aus zwei Hypothesen bestehende Modell

$$A(\eta) = aB(\eta-1) + bA(\eta-4) + cA(\eta-1)$$

$$B(\eta) = dA(\eta-1) + eA(\eta-2) + S$$

wobei der Zeitindex η Halbjahresperioden beschreiben soll. Die Variable A besitzt die Endgleichung

$$A(\eta) = cA(\eta-1) + adA(\eta-2) + aeA(\eta-3) + bA(\eta-4) + aS \quad (24.7)$$

Ein angenommener 'historischer Verlauf' von A für 50 Halbjahresperioden wird durch den Zeitverlauf mit dem Symbol + in Abbildung 24.2 dargestellt. Die Kleinstquadratschätzung der Parameter in Gleichung (24.7) anhand dieser Zeitreihe führt zu der parametrisch-singulären Hypothese

$$A(\eta) = 2,837A(\eta-1) - 2,831A(\eta-2) + 1,107A(\eta-3) - 0,116A(\eta-4) + 61,226 \quad (24.8)$$

deren simulierter Zeitverlauf durch das Symbol * beschrieben wird.

Gehen wir nunmehr davon aus, daß die Beobachtungswerte für $A(\eta)$ und $B(\eta)$ Flußgrößen darstellen, die nur als kumulierte Jahreswerte zur Verfügung stehen, so ist es nicht möglich, eine Parameterschätzung, wie sie sich in Gleichung (24.8) dokumentiert, vorzunehmen.

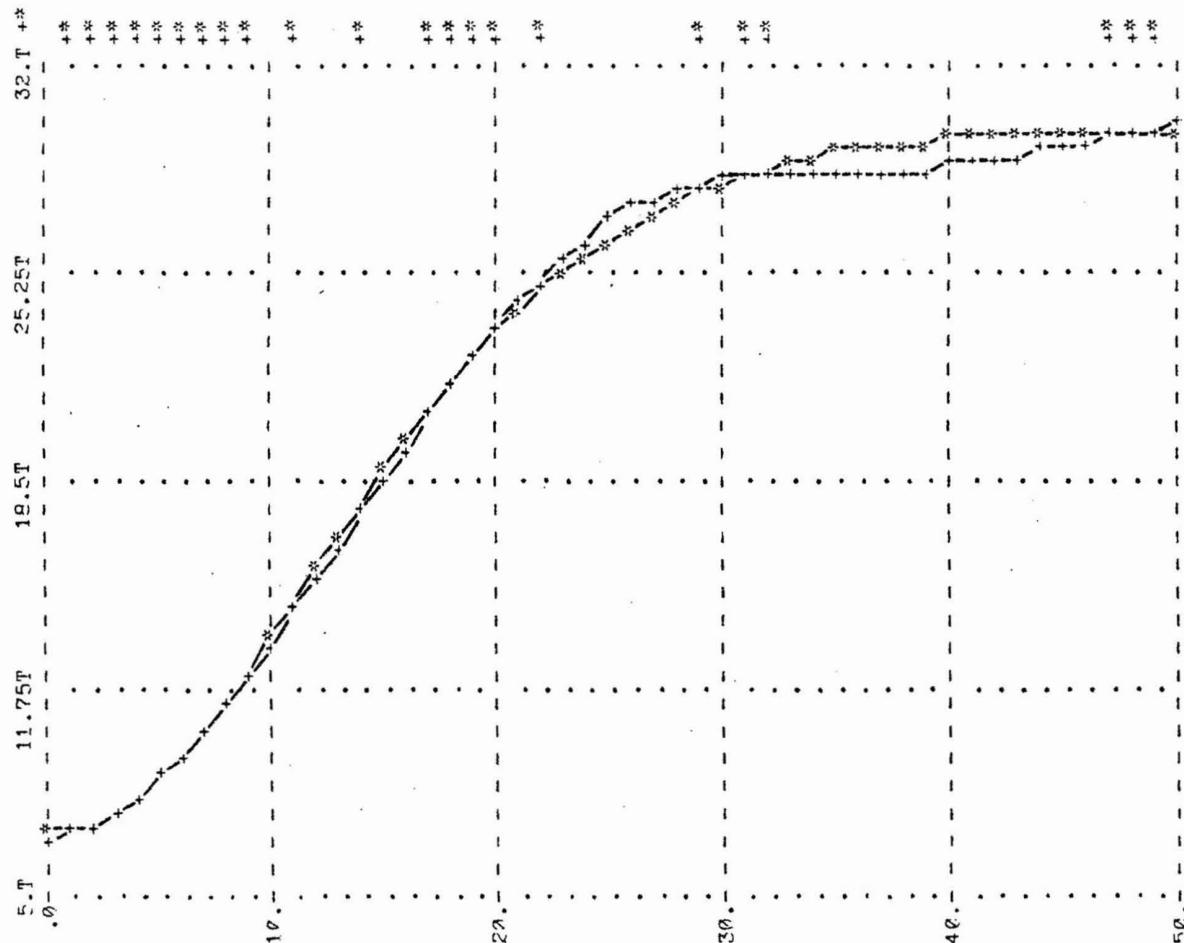


Abb. 24.2 Histogramm der beobachteten Zeitreihe der Variablen A (+) sowie der Zeitreihe für A, welche von der Gleichung (24.8) erzeugt wird (*) [Einheit T: Tausend]

Es fragt sich aber, ob es möglich ist, einen Modellansatz des zu beschreibenden Systems zu entwickeln, mit dem man eine Zeitreihe der Jahreswerte gewinnt, deren Verlauf gegenüber dem Zeitverlauf des angestrebten aber nicht realisierbaren Ansatzes (24.8) nur geringfügig abweicht.

In Abbildung 24.3 ist auf der linken Seite ein Pfeilschema des rekursiven Zusammenhangs angeführt.

Denken wir uns dieses Pfeilschema nunmehr so zusammengeschoben, daß sich die Quadrate der Perioden $\eta-1$, $\eta-3$, $\eta-5$ usw. mit denen der Periode η bzw. $\eta-2$, $\eta-4$ usw. decken, dann erhält man aufgrund dieser zeit-

lichen Aggregation das auf der rechten Seite dargestellte Schema, welches ein interdependentes Jahresmodell zeigt.

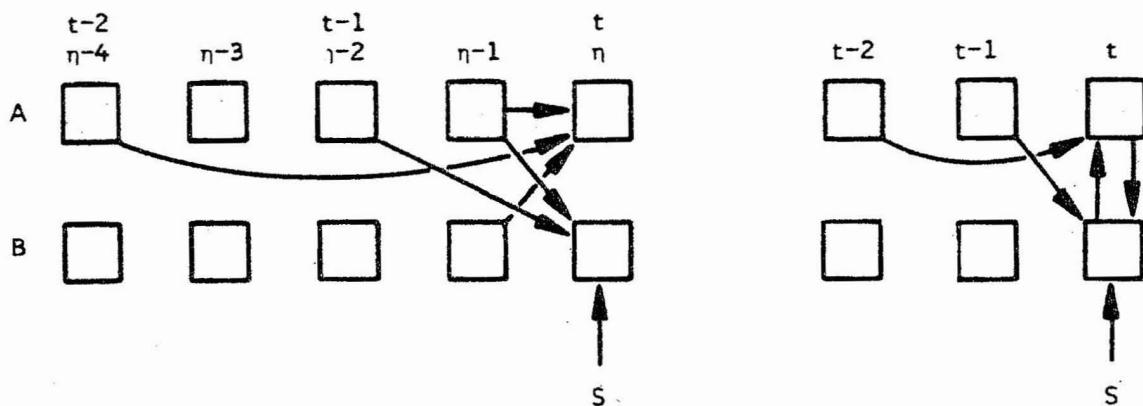


Abb. 24.3 Beispiel der Gewinnung eines interdependenten Modells durch zeitliche Aggregation eines rekursiven Modells

Dieses interdependente Jahresmodell wird durch

$$\begin{aligned} A(t) &= \alpha B(t) + \beta A(t-2) \\ B(t) &= \gamma A(t) + \delta A(t-1) + s \end{aligned} \quad (24.9)$$

beschrieben. Es liegt nahe, anhand der bekannten Jahresbeobachtungswerte die Parameter dieser 'Hilfskonstruktion aus Datenmangel' zu schätzen. Legt man die aus (24.9) ermittelte Endgleichung von $A(t)$

$$A(t) = \frac{\alpha\delta}{1-\alpha\gamma} A(t-1) + \frac{\beta}{1-\alpha\gamma} A(t-2) + \frac{\alpha}{1-\alpha\gamma} s$$

zugrunde, so führt eine Kleinstquadratschätzung der Parameter anhand der Jahresbeobachtungswerte zu der Gleichung

$$A(t) = 1,84507 A(t-1) - 0,86857 A(t-2) + 1259,56551 \quad (24.10)$$

Diese Jahresbeobachtungswerte ergeben sich jeweils aus der Addition von zwei Halbjahresbeobachtungswerten der Variablen, deren Verlauf in Abbildung 24.2 beschrieben wurde.

Abbildung 24.4 zeigt den Zeitverlauf der Jahreswerte A im Falle des rekursiven Ansatzes (24.8) sowie des interdependenten Ansatzes (24.10) im Vergleich mit den Beobachtungswerten von A , die der Schätzung der

Parameter beider Modelle zugrunde lagen.¹

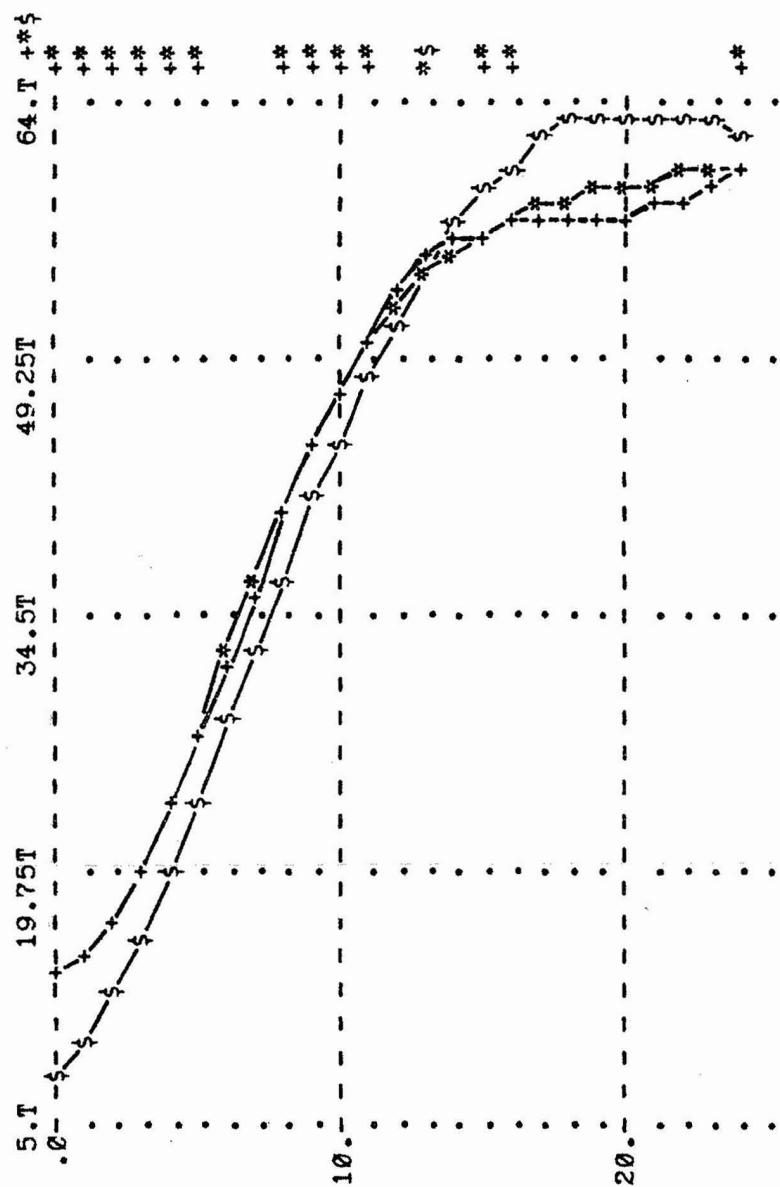


Abb. 24.4 Zeitlicher Verlauf der Variablen A im Falle des rekursiven Halbjahresmodells (*), des interdependenten Jahresmodells (\$) sowie der Jahresbeobachtungswerte von A(+) [Einheit T: Tausend]

Man erkennt, daß das rekursive Modell zu einer wesentlich stärkeren Übereinstimmung mit dem Verlauf der Beobachtungswerte führt als das interdependente Modell.

1 Die Jahreswerte des rekursiven Modells ergeben sich aus der Addition von jeweils zwei Halbjahreswerten des rekursiven Ansatzes.

Da es unter den geschilderten Umständen nur möglich ist, ein interdependentes Modell zu entwickeln, fragt es sich aber, ob die Abweichungen zwischen den durch das interdependente Modell erzeugten Verläufen und den Beobachtungswerten zu tolerieren sind. Diese Frage hängt von den Kriterien ab, mit denen die Akzeptierbarkeit statistisch geschätzter Modelle beurteilt wird.

Die Schilderung der Überlegungen, die zu einem interdependenten Modell führen können, läßt die Anwendungsberechtigung derartiger Modelle erkennen: Sie stellen Hilfskonstruktionen aus Datenmangel dar, deren Berechtigung aus ihrer praktischen Bewährung folgt.

Das geschilderte Verfahren macht jedoch deutlich, daß interdependenten Modelle nur im Rahmen von Überlegungen entwickelt werden können, die auf eine statistische Schätzung der Parameter des 'Ersatzmodells' hinauslaufen. Aus diesem Grunde treten in Modellierungskonzeptionen, die auf statistische Schätzungen verzichten und auf der Basis subjektiv geschätzter Parameter entwickelt werden, keine interdependenten Gleichungssysteme auf: Als Folge davon gibt es Simulationssysteme, wie DYNAMO oder CSMP, die von vornherein nur für rekursive dynamische Modelle angelegt sind.

Die Schätzung der Parameter interdependenten Modells mit Hilfe der Methode der Kleinstquadrate führt zu einer verzerrten Schätzung, weil die erwähnte vierte Voraussetzung einer effizienten Kleinstquadrat-schätzung nicht erfüllt ist.² Es handelte sich um die Forderung, daß zwischen einer Schockvariablen und den erklärenden Variablen einer strukturellen Gleichung keine Abhängigkeiten existieren dürfen.

Zur Illustration, daß solche Abhängigkeiten aber in interdependenten Modellen existieren, greifen wir auf das erörterte interdependenten MA-Modell

$$Y(t) = C(t) + I_i(t) + I_a(t) \quad (24.11)$$

$$I_i(t) = \beta[C(t) - C(t-1)] + \varepsilon(t) \quad (24.12)$$

$$C(t) = \alpha Y(t) + \mu(t) \quad (24.13)$$

zurück und wollen den Parameter β in der Gleichung (24.12) schätzen. Eine verzerrungsfreie Schätzung setzt voraus, daß die vorherbestimm-

² Vgl. Seite 115

ten Variablen $C(t)$ und $C(t-1)$ nicht von der Schockvariablen ε abhängen: Mit (24.12) in (24.11) erhalten wir

$$Y(t) = C(t) + \beta[C(t)-C(t-1)] + I_a(t) + \varepsilon(t) \quad (24.14)$$

(24.14) in (24.13) liefert

$$C(t) = (\alpha + \alpha\beta)C(t) - \alpha\beta C(t-1) + \alpha I_a(t) + \alpha\varepsilon(t) + \mu(t)$$

Wir erkennen, daß die vorherbestimmte Variable $C(t)$ von ε beeinflußt wird, und eine Unabhängigkeit im Sinne der vierten Voraussetzung einer effizienten Kleinstquadratschätzung nicht vorliegt.

Um im Fall interdependenten Modells zu besseren Schätzungen zu gelangen, werden verschiedene Schätzmethoden wie zweistufige Kleinstquadratschätzungen, Limited Information Maximum Likelihood Estimation, verwendet. Ihre Anwendung erfordert einen höheren theoretisch-statistischen Wissensstand als im Fall von Kleinstquadratschätzungen und ist auch in der technischen Durchführung aufwendiger.

Die Entwicklung interdependenten Modells ist so eng mit den Problemen ihrer statistischen Schätzung verbunden, daß sie heute von hauptberuflichen Ökonometrikern und, seltener, ökonometrisch geschulten Fachwissenschaftlern aus dem Bereich der Wirtschaftswissenschaften betrieben wird. Diese Situation wird sich ändern, sobald in größerem Umfang computergestützte Schätz- und Simulationssysteme zur Verfügung stehen, welche die komplizierteren Schätzverfahren für interdependenten Modelle auf einfache Weise zu handhaben erlauben. Als fortgeschrittenes Beispiel derartiger Schätz- und Simulationssysteme wird später das TROLL-System erörtert.³

2.4.2. Analyse der Verknüpfungsstruktur rekursiver und interdependenten Modelle

Die erste formale Charakterisierung rekursiver und interdependenten Modelle anhand eines Pfeilschemas erlaubt es nicht ohne weiteres, ein vorliegendes größeres Modell als rekursiv oder interdependent einzutragen.

³ Vgl. Seite 569 ff

Evident wird diese Behauptung im Fall der heute schon vorliegenden Modelle mit drei- bis vierhundert endogenen Variablen. Hier kann eine Entscheidung über den vorliegenden Modelltyp nur anhand formaler Überprüfungsverfahren durchgeführt werden. Zur Beurteilung der Verknüpfungsstruktur von Modellen bedient man sich sogenannter Strukturmatrizen.

A. Strukturmatrizen rekursiver Modelle

Da die Verknüpfung zwischen den endogenen unverzögerten Variablen zu beurteilen ist, können einem Modell aus Gründen einer besseren Übersicht alle Informationen entzogen werden, die sich nicht auf die Abhängigkeit zwischen den endogenen unverzögerten Variablen beziehen.

Die Gleichungen des MA-Modells

$$\begin{aligned} Y(t) &= C(t) + I_a(t) + I_i(t) \\ C(t) &= \alpha Y(t-1) \\ I_i(t) &= \beta [C(t) - C(t-1)] \end{aligned}$$

können daher auf die Form

$$\begin{aligned} Y(t) &= F_1[C(t), I_i(t)] \\ C(t) &= F_2[\text{const}] \\ I_i(t) &= F_3[C(t)] \end{aligned}$$

reduziert werden. In diesem Fall braucht kein formales Entscheidungsverfahren eingeführt zu werden, denn die Rekursivität des MA-Modells ist augenscheinlich.

Zur Beurteilung der Rekursivität größerer dynamischer Modelle dient die sogenannte Strukturmatrix. Jede Zeile einer Strukturmatrix korrespondiert mit einer bestimmten Gleichung des betreffenden dynamischen Modells, jede Spalte dagegen mit einer entsprechenden unverzögerten Variablen. Die Elemente der Strukturmatrix bestimmen sich nach der Vorschrift: Trage in die Spalte, welche zu der betrachteten Gleichung gehört, den Wert Eins ein, wenn die dieser Spalte zugeordnete endogene unverzögerte Variable in der Gleichung auftritt. Alle freien Plätze erhalten eine Null. Entsprechend dieser Vorschrift erhält

man für das MA-Modell die Strukturmatrix

	Gleichungsnr.	$Y(t)$	$C(t)$	$I_i(t)$
$Y(t) = F_1[C(t), I_i(t)]$	1		[1 1 1]	
$C(t) = F_2[\text{const}]$	2		[0 1 0]	
$I_i(t) = F_3[C(t)]$	3		[0 1 1]	

In einer Strukturmatrix kann man die Zeilen und Spalten beliebig miteinander vertauschen, ohne den Informationsgehalt des Ansatzes zu verändern. Denn eine Vertauschung der Reihenfolge entspricht einer Umstellung der Reihenfolge der Gleichungen und eine Vertauschung der Spalten bewirkt keine Änderung der Verknüpfungsbehauptung.

Es liegt nun nahe, diesen 'Freiraum' der Spalten- und Zeilenanordnungen auf bestimmte Standardformen einzuschränken, die zugleich Auskunft über den Modelltyp liefern.

Zur Charakterisierung dieser Standardformen sind drei Strukturtypen von Matrizen einzuführen: Dreiecksmatrizen, unzerlegbare Matrizen und blocktrianguläre Matrizen.

Eine dreieckige Strukturmatrix ist eine quadratische Matrix der Form

$$\begin{bmatrix} * & 0 & 0 & \dots & 0 \\ * & * & 0 & \dots & 0 \\ * & * & * & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ * & * & * & \dots & * \end{bmatrix} \quad (24.15)$$

wobei mit dem Sternsymbol Elemente gekennzeichnet sind, die entweder eine 1 oder 0 repräsentieren können.

Eine quadratische Matrix wird als zerlegbar bezeichnet, wenn es gelingt, sie durch einen Austausch von Zeilen und Spalten in die Form

$$\left[\begin{array}{c|c} M_1 & 0 \\ \hline * & M_2 \end{array} \right] \quad (24.16)$$

zu überführen. M_1 ist eine quadratische ($r \times r$)-Matrix und M_2 wird damit eine $(n-r) \times (n-r)$ -Matrix.

Mit dem Symbol * ist eine $(n-r) \times (r)$ -Matrix gekennzeichnet, die beliebt

bige Werte annehmen darf, während in der oberen rechten Ecke eine Nullmatrix der Ordnung $r \times (n-r)$ enthalten ist.

Gelingt es nicht, eine Matrix durch Zeilen- und Spaltenaustausche in die Form (24.16) zu überführen, dann wird von einer vollkommen unzerlegbaren Matrix gesprochen.

Zur Verdeutlichung betrachten wir die Strukturmatrix,

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

welche wir durch den Austausch der ersten mit der dritten Zeile und Spalte in die Form

$$\begin{array}{|c|c|c|c|} \hline 1 & 0 & 0 & 0 \\ \hline 1 & 1 & 0 & 0 \\ \hline 1 & 1 & 1 & 1 \\ \hline 1 & 1 & 1 & 1 \\ \hline \end{array} \quad (24.17)$$

überführen können. Es handelt sich also um eine zerlegbare Matrix.

Eine unzerlegbare blocktriangulare Matrix ist eine quadratische Matrix der Form:

$$\begin{bmatrix} M_1 & 0 & 0 & \dots & 0 \\ * & M_2 & 0 & \dots & 0 \\ * & * & M_3 & \dots & 0 \\ * & * & * & \dots & 0 \\ * & * & * & \dots & M_r \end{bmatrix}$$

Die Untermatrizen 0 über der Hauptdiagonalen sind Nullmatrizen. Die Elemente der Hauptdiagonalen M_1, M_2, \dots, M_r bilden unzerlegbare Matrizen, während die mit * gekennzeichneten Blöcke beliebige Matrizen repräsentieren.

Die im letzten Beispiel gefundene Matrix (24.17) ist eine unzerlegbare blocktriangulare Matrix, weil sämtliche drei auf der Hauptdiagonalen liegenden Blöcke unzerlegbare Matrizen darstellen.

Wir haben uns nunmehr das Rüstzeug geschaffen, um ein generelles Be-

urteilungskriterium für rekursive und interdependente Modelle aufzustellen.

Es gilt der Satz: Ein dynamisches Modell ist stets nur dann rekursiv, wenn seine Strukturmatrix durch Zeilen- und Spaltenaustausche in eine Dreiecksmatrix überführt werden kann.

Die Strukturmatrix des Multiplikatormodells kann beispielsweise in die Dreiecksform

Gleichungsnr.	$C(t)$	$I_i(t)$	$Y(t)$
2		1 0	0
3		1 1	0
1		1 1	1

überführt werden. Die Dreiecksform der Strukturmatrix eines rekursiven Modells dient nicht nur seiner Identifizierung, sondern liefert wichtige Einsichten über die Systemstruktur. Im Falle der Strukturmatrix des MA-Modells erkennt man, daß bei einer periodischen Regressionslösung in einem ersten Schritt die endogene Variable C ermittelt werden muß, danach I_i und schließlich Y . Man kann das Berechnungsverfahren der Variablenwerte im Rahmen einer periodischen Regressionslösung auch durch folgendes Diagramm darstellen.

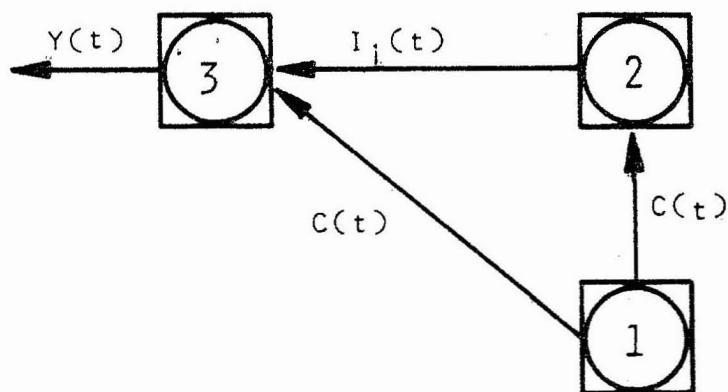


Abb. 24.5 Rekursive Anordnung der endogenen Variablen eines MA-Modells zur Ermittlung einer periodischen Regressionslösung

Für die Simulation eines rekursiven dynamischen Modells mit Hilfe eines Computers ist es notwendig, die vorgegebenen Gleichungen durch

Instruktionen darzustellen, die in Übereinstimmung mit der Numerierung einer Dreiecksstrukturmatrix abzuarbeiten sind. Dies bedeutet, daß bei größeren Modellen praktisch immer die Anordnungsfolge einer Dreiecksstrukturmatrix ermittelt werden muß.

Die Überlegenheit bestimmter Simulationsprachen wie zum Beispiel DYNAMO oder CSMP gegenüber FORTRAN wird durch die Tatsache deutlich, daß in diesen für rekursive Modelle geschaffenen Programmiersprachen die prozedurale Anordnung der Gleichungen von dem Programmsystem selbst vorgenommen wird. Geht man beispielsweise davon aus, daß in der Jumbo-Version der DYNAMO-Programmiersprache Modelle mit mehr als tausend endogenen Variablen simuliert werden können, so müßte vor der Programmierung derartiger Modelle in FORTRAN die rekursive Anordnung von tausend Variablen ermittelt werden.

B. Strukturmatrizen interdependenten Modelle

Eine zerlegbare Strukturmatrix, die sich nicht in eine Dreiecksmatrix überführen läßt, kann zumindest in eine blocktriangulare unzerlegbare Strukturmatrix umgewandelt werden.

Eine Matrix wie zum Beispiel

$$\begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix}$$

die nicht in mehrere unzerlegbare Diagonalblöcke M_1, M_2, \dots überführt werden kann, sondern sozusagen nur aus einem unzerlegbaren Diagonalsblock M_1 besteht, ist als ein Spezialfall einer unzerlegbaren blocktriangulären Matrix aufzufassen und wurde bereits als vollkommen unzerlegbare Matrix bezeichnet.

Es gilt der Satz: Ein interdependentes dynamisches Gleichungssystem besitzt eine unzerlegbare Strukturmatrix seiner endogenen unverzögerten Variablen.

Das erörterte interdependente MA-Modell

$$Y(t) = C(t) + I_i(t) + I_a(t)$$

$$C(t) = \alpha Y(t)$$

$$I_i(t) = \beta [C(t) - C(t-1)]$$

besitzt beispielsweise die Strukturmatrix

$$\begin{array}{ccc} Y & C & I_i \\ \left[\begin{array}{ccc} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{array} \right] \end{array}$$

die, wie man leicht erkennt, unzerlegbar ist.

In vielen Fällen ist es, wie gesagt, nicht ohne weiteres möglich, zu entscheiden, ob das vorliegende System interdependent oder rekursiv ist. Es ist darüberhinaus auch möglich, daß ein dynamisches Modell aus mehreren Subsystemen besteht, die in sich je ein interdependentes Modell bilden. In einem Modell ist aber nicht unmittelbar zu erkennen, welche endogenen Variablen derartige interdependente Subsysteme miteinander bilden.

Betrachten wir beispielsweise das Gleichungssystem

$$\begin{aligned} X_1 &= F[X_3, X_7] \\ X_2 &= F[X_4] \\ X_3 &= F[X_3] \\ X_4 &= F[X_5] \\ X_5 &= F[X_3, X_4] \\ X_6 &= F[X_1] \\ X_7 &= F[X_6, X_9] \\ X_8 &= F[X_4] \\ X_9 &= F[X_1, X_2] \end{aligned} \tag{24.18}$$

so ist nicht ohne weiteres klar, welche Variablen zu bestimmten interdependenten Teilmodellen gehören. Dieses Gleichungssystem besitzt die Strukturmatrix

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
x_1	1		1			1			
x_2		1		1					
x_3			1						
x_4				1	1				
x_5			1	1	1				
x_6	1					1			
x_7						1	1		1
x_8				1				1	
x_9	1	1							1

Ein entsprechende Vertauschung der Zeilen und Spalten dieser Struktur-matrix führt zu der blocktriangulären Matrix,

	x_3	x_4	x_5	x_2	x_1	x_6	x_7	x_9	x_8
x_3	1								
x_4		1	1						
x_5	1	1	1						
x_2		1		1					
x_1	1				1		1		
x_6					1	1			
x_7						1	1	1	
x_9				1	1			1	
x_8		1							1

aus der man die interdependenten Subsysteme und ihre Verknüpfung mit-einander deutlich erkennt. Das Verfahren zur Ermittlung derartiger blocktriangularer Strukturmatrizen wird im folgenden Abschnitt be-schrieben.

C. Ermittlung standardisierter Strukturmatrizen von rekursiven und interdependenten Modellen

Angesichts der Tatsache, daß es Computerprogramme gibt, die die beschwerliche Ermittlung einer standardisierten Strukturmatrix übernehmen, könnte man meinen, daß es nicht mehr notwendig sei, sich mit dem Problem zu beschäftigen, auf welche Weise die geschilderten Standardformen von Strukturmatrizen zu ermitteln sind. Dennoch soll im folgenden ein Verfahren beschrieben werden, weil seine Kenntnis das Verständnis für die Verknüpfungsweise dynamischer Systeme erhöht.

Neben der auf Seite 326 angeführten Strukturmatrix des rekursiven MA-Modells

$$\begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}$$

kann man von demselben Modell auch eine Boolsche Matrix aufstellen.

Im Gegensatz zur Strukturmatrix korrespondiert jede Spalte nur mit einer der in den Gleichungen als unabhängige Variable fungierenden endogenen unverzögerten Variablen, weshalb die Boolese Matrix auf der Hauptdiagonalen nur Nullen aufweist, während die Strukturmatrix dort nur mit Einsen besetzt ist.

Die Boolese Matrix eines MA-Modells weist daher folgende Form auf

$$\begin{array}{ccc} Y & C & I_i \\ \begin{matrix} Y \\ C \\ I_i \end{matrix} & \begin{bmatrix} 0 & 1 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} & \begin{aligned} Y(t) &= C(t) + I_i(t) + I_a(t) \\ C(t) &= \alpha Y(t-1) \\ I_i(t) &= \beta [C(t) - C(t-1)] \end{aligned} \end{array} \quad (24.19)$$

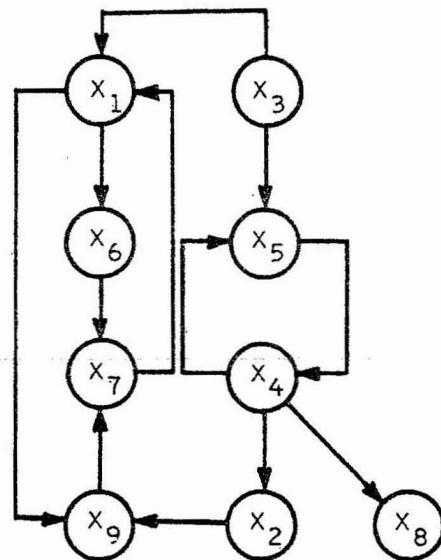
Die beschriebene Aufstellung einer Boolese Matrix ist unproblematisch, wenn jede Gleichung des dynamischen Modells als unabhängige Variable genau eine endogene unverzögerte Variable aufweist. In diesem Fall, der für die meisten dynamischen Modelle zutrifft, kann die jeweilige Gleichung als ein schwarzer Kasten angesehen werden, dessen Inputvariablen I_j die Outputvariable O_i bestimmen. Im Falle der

Boolschen Matrix des MA-Modells (24.19) wird die Outputvariable Y beispielsweise durch die Inputvariablen C und I; bestimmt.

Das Verfahren zur Entwicklung einer blocktriangularen, nicht weiter zerlegbaren Strukturmatrix soll am Beispiel des bereits zitierten Gleichungssystems (24.18) demonstriert werden. Die Boole'sche Matrix des Systems erhalten wir, wenn die Elemente der Hauptdiagonalen durch Nullen ersetzt werden. Denn diese Einsen auf der Hauptdiagonalen repräsentieren ja die Outputvariablen in jeder Gleichung, welche in der nur die Inputvariablen charakterisierenden Boole'schen Matrix nicht erscheinen dürfen.

Für unser Beispiel führt die Streichung zu der Boole'schen Matrix

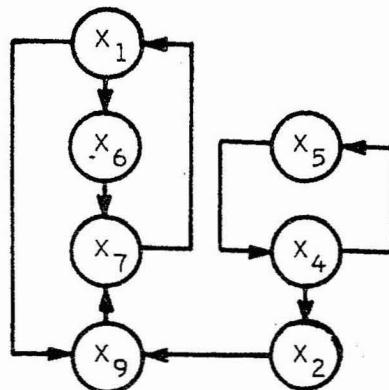
	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
x_1			1				1		
x_2				1					
x_3									
x_4					1				
x_5			1	1					
x_6	1								
x_7					1				1
x_8				1					
x_9	1	1							



Diejenigen Zeilen, welche leere Felder, d.h. nur unterdrückte Nullen enthalten, repräsentieren Variablen, die nicht von den anderen Variablen beeinflußt werden. Weiterhin wird deutlich, daß Spalten, welche nur unterdrückte Nullen aufweisen, solche Variablen beschreiben, die selbst keine anderen Variablen beeinflussen. In unserem Beispiel wird daher die Variable x_3 als erste berechenbare Variable in die erste Zeile einer Variablenliste eingetragen, während die Variable x_8 , welche für keine weitere Berechnung mehr gebraucht wird, am Ende derselben Liste aufgeführt wird, d.h.

$$\begin{matrix} x_3 \\ \vdots \\ x_8 \end{matrix}$$

Nach diesem ersten Schritt gewinnen wir eine reduzierte Matrix M durch Streichung der Zeilen und Spalten von x_3 und x_8 . Diese Matrix und der entsprechende Graph besitzen die Gestalt

$$M = \begin{array}{|c||c|c|c|c|c|c|c|} \hline & x_1 & x_2 & x_4 & x_5 & x_6 & x_7 & x_9 \\ \hline x_1 & & & & & & 1 & \\ \hline x_2 & & & 1 & & & & \\ \hline x_4 & & & & 1 & & & \\ \hline x_5 & & & -1 & & & & \\ \hline x_6 & 1 & & & & & & \\ \hline x_7 & & & & 1 & & 1 & \\ \hline x_9 & 1 & 1 & & & & & \\ \hline \end{array}$$


Von der Matrix M ist das Boolsche Matrizenprodukt M^2 zu ermitteln.

Die Boolsche Multiplikation einer Matrix sei kurz dargestellt:
Eine Boolese Multiplikation geht zum einen von der zusätzlichen Voraussetzung aus, daß A und B eine 0-1-Matrix bilden. Das übliche Matrizenprodukt M aus den Matrizen A und B ergibt sich wie bereits eingehend beschrieben, durch die Vorschrift $M_{ij} = A_i B_j$, d.h. das Skalarprodukt aus dem Zeilenvektor A_i und dem Spaltenvektor B_j liefert das Element M_{ij} der Produktmatrix M. Wendet man nunmehr zur Ermittlung einer Matrix C die zusätzliche Vorschrift

$$c_{ij} = \begin{cases} 0 & \text{wenn } M_{ij} = 0 \\ 1 & \text{wenn } M_{ij} \neq 0 \end{cases}$$

an, dann erhält man mit C das Boolese Matrizenprodukt von A und B.
Das Element einer quadratischen Potenzmatrix $A_4 B_3$ würde nach der allgemeinen Vorschrift zur Ermittlung von Vektorprodukten zu dem Ergebnis

$$A_4 B_3 = \begin{bmatrix} 1 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 0 \\ 1 \end{bmatrix} = 1 + 0 + 0 + 1 = 2$$

Führen. Da $2 \neq 0$, ist bei einem Booleschen Vektorprodukt statt 2 als Element eine 1 zu wählen.

Nach dieser Berechnungsvorschrift ergibt sich die Quadratmatrix M^2 .

	x_1	x_2	x_4	x_5	x_6	x_7	x_9
x_1					1		1
x_2				1			
x_4			1				
x_5				1			
x_6						1	
x_7	1	1					
x_9			1			1	

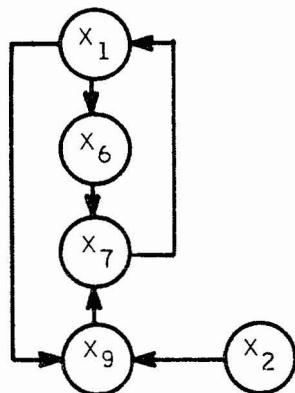
Mit Hilfe dieser Quadratmatrix lässt sich zeigen, welche Variablen eine andere Variable über eine weitere Variable beeinflussen. Die mit x_1 korrespondierende Zeile enthält beispielsweise in der mit x_6 und x_9 korrespondierenden Spalte eine 1. Dies bedeutet, daß x_6 und x_9 die Variable x_1 über eine weitere Variable beeinflussen. Wie wir uns anhand des Graphen überzeugen, ist dies der Fall, denn man findet ohne Schwierigkeiten die Beeinflussungsketten $x_6-x_7-x_1$ und $x_9-x_7-x_1$. Verallgemeinernd kann durch die Potenzmatrix M^n die über $n-1$ Variablen laufende Verknüpfung zwischen den mit einer Zeile und Spalte korrespondierenden Variablen gefunden werden.

Im Falle der Matrix M^2 enthalten die den Variablen x_4 und x_5 zugeordneten Zeilen nur auf der Hauptdiagonalen eine 1. Dies bedeutet, daß sich x_4 und x_5 über eine weitere Variable selbst beeinflussen und damit ein simultanes Gleichungssystem bilden. Die beiden Variablen bestimmen auch eine maximale Schleife, d.h. sie umfassen ein abgeschlossenes simultanes Gleichungssystem, weil sie nicht von den verbleibenden Variablen beeinflußt werden. Wir erhalten damit die erweiterte

Liste

x_3
 x_5, x_4 simultan
 \vdots
 x_8

Streicht man die entsprechenden Zeilen und Spalten von x_5 und x_4 in der Matrix M^2 , so ergibt sich die reduzierte Matrix M_2^*

$$M_2^* = \begin{array}{c|ccccc|} & x_1 & x_2 & x_6 & x_7 & x_9 \\ \hline x_1 & & & 1 & & 1 \\ x_2 & & & & & \\ x_6 & & & & 1 & \\ x_7 & 1 & 1 & & & \\ x_9 & & & & & 1 \end{array}$$


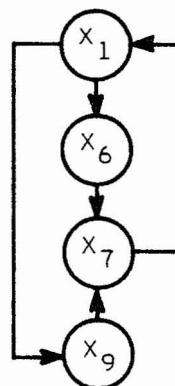
Anhand dieser Matrix erkennen wir, daß die mit x_2 korrespondierende Zeile nur Nullen besitzt, d.h. von keiner der noch vorhandenen Variablen beeinflußt wird. Die Zeile und Spalte von x_2 ist daher zu streichen, und wir erhalten die Matrix M_2^{**} .

$$M_2^{**} = \begin{array}{c|ccccc|} & x_1 & x_6 & x_7 & x_9 \\ \hline x_1 & & 1 & & 1 \\ x_6 & & & 1 & \\ x_7 & 1 & & & \\ x_9 & & & 1 & \end{array}$$

Durch Aufnahme von x_2 in die Variablenliste erhält man:

x_3
 x_5, x_4 simultan
 x_2
 \vdots
 x_8

Indem man in der ursprünglichen Matrix M ebenfalls die Zeilen und Spalten von x_5 , x_4 und x_2 streicht, ergibt sich M^* .

$$M^* = \begin{array}{c|c|c|c|c} & x_1 & x_6 & x_7 & x_9 \\ \hline x_1 & & & 1 & \\ \hline x_6 & 1 & & & \\ \hline x_7 & & 1 & & 1 \\ \hline x_9 & 1 & & & \end{array}$$


Die Matrix M_3 berechnet sich dann mit:

$$M_3 = M_2^{**} \cdot M^* =$$

	x ₁	x ₆	x ₇	x ₉
x ₁	1			
x ₆		1		
x ₇			1	
x ₉		1		1

Wir erkennen, daß die Variablen x_1 , x_6 und x_7 über zwei Variablen nur von sich selber abhängen. x_9 dagegen wird über zwei Variablen sowohl von sich selbst als auch von x_6 beeinflußt. Die vier Variablen bilden eine maximale Schleife, d.h. ein simultanes Gleichungssystem, weil jede der vier Schleifen über zwei andere Variablen läuft und daher alle Variablen voneinander abhängig sind.

Damit ergibt sich die endgültige Variablenliste

x_3
 x_5, x_4 simultan
 x_2
 x_1, x_6, x_7, x_9 simultan
 x_8

In dieser Reihenfolge können wir nunmehr die Zeilen der ursprünglichen Strukturmatrix anordnen und erhalten die uns bereits bekannte blocktrianguläre unzerlegbare Strukturmatrix

	x_3	x_4	x_5	x_2	x_1	x_6	x_7	x_9	x_8
x_3	(1)								
x_4		(1)	1						
x_5	1	1	(1)						
x_2		1		(1)					
x_1	1				(1)		1		
x_6					1	(1)			
x_7						1	(1)	1	
x_9				1	1				(1)
x_8		1							(1)

Anhand der ermittelten Strukturmatrix lassen sich die in Abbildung 24.6 dargestellten Teilsysteme voneinander abgrenzen.

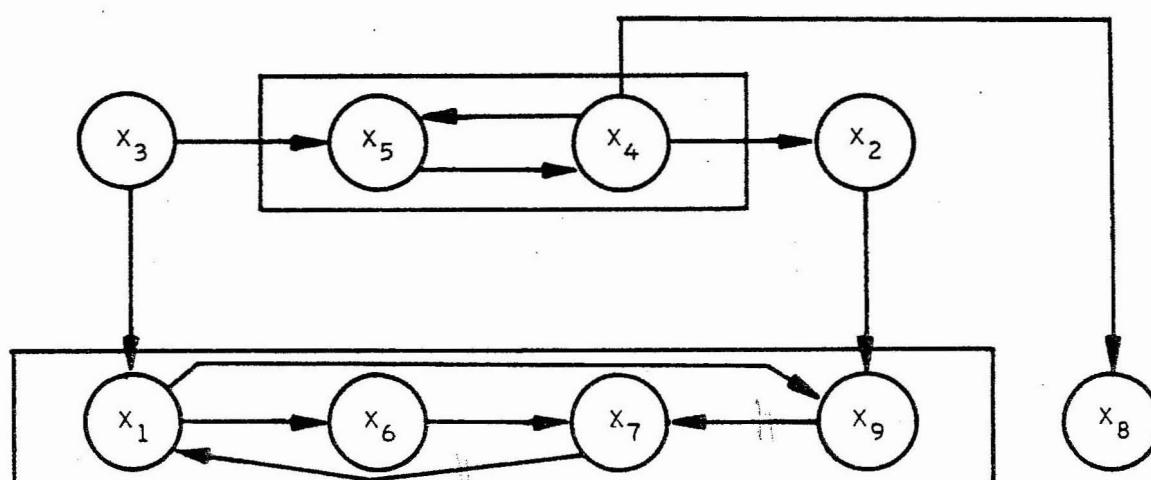


Abb. 24.6 Schema der sukzessiven Lösung eines Gleichungssystems mit interdependenten Teilsystemen

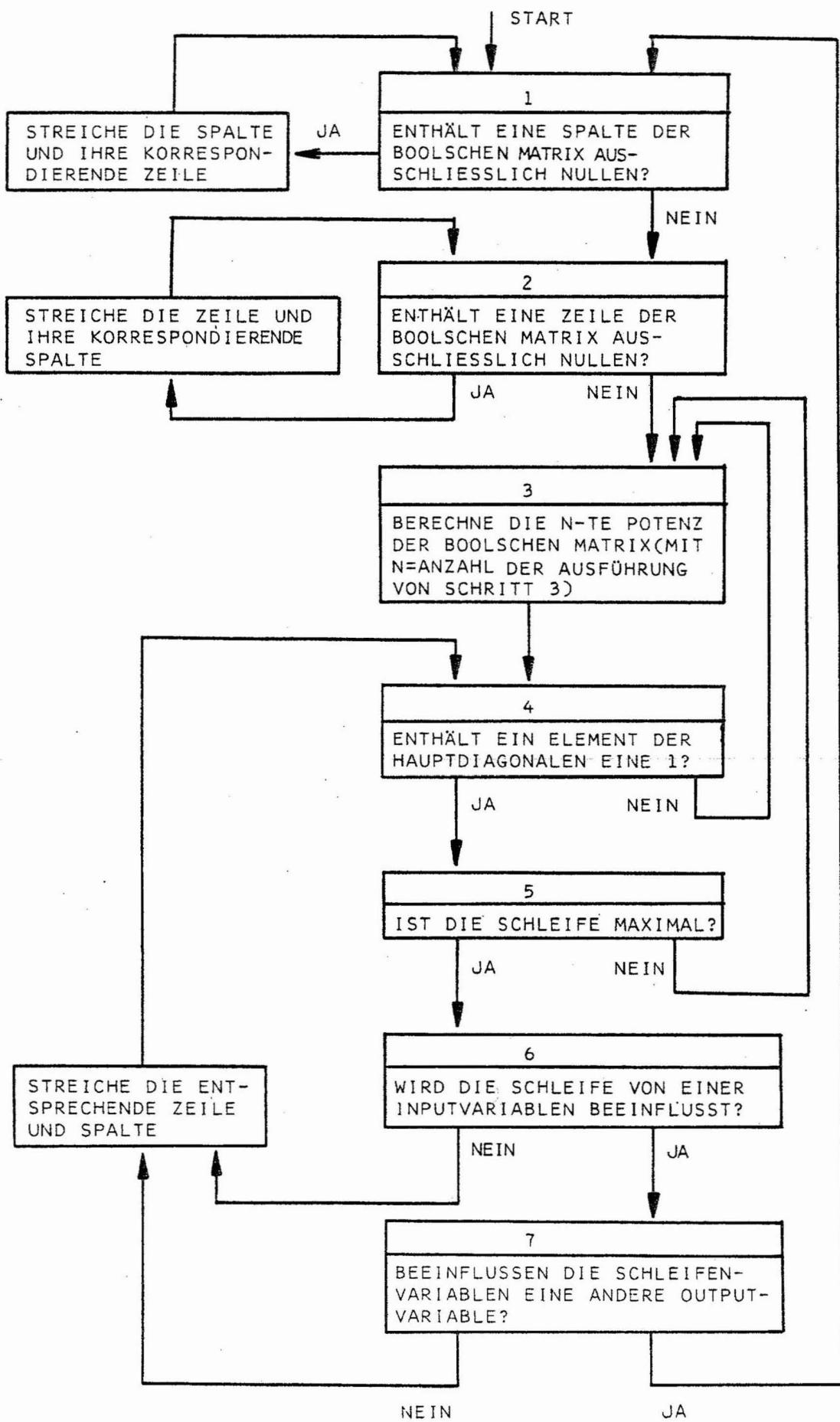


Abb. 24.7 Schema zur Zerlegung eines dynamischen Modells in minimale Subsysteme von interdependenten Gleichungen

Das beschriebene Verfahren läßt sich auf beliebige Modelle erweitern. Seine Durchführung ist anhand von Abbildung 24.7 noch einmal in allgemeiner Form zusammengefaßt.⁴

Die Simulationssysteme TROLL und SIMULATE, die für interdependent dynamische Modelle geeignet sind, führen die Ermittlung unzerlegbarer blocktriangularer Strukturmatrizen anhand der eingegebenen Gleichungen selbst durch und entlasten den Entwickler dynamischer Modelle von einer zeitaufwendigen Aufgabe. Die Programmiersprache DYNAMO hat eine Fehlerroutine, welche die vorhandenen maximalen interdependenten Schleifen eines Modells als unzulässig aufzeigt, da DYNAMO keine interdependenten Gleichungen zuläßt. Im Falle des anhand von Abbildung 24.6 dargestellten Systems würde in DYNAMO ein Ausdruck erfolgen, daß zum einen die Variablen X_5 und X_4 sowie die Variablen X_1 , X_6 , X_7 und X_1 , X_9 , X_7 jeweils ein interdependentes System bilden.⁵

D. Zeitpfadbestimmung in interdependenten Modellen

a) Zeitpfadbestimmung in linearen interdependenten Modellen

Interdependent Modelle repräsentieren eine Formalstruktur, die in der Mathematik als simultanes Gleichungssystem bezeichnet wird. Wir wollen daher im folgenden synonym von interdependenten oder simultanen Modellen sprechen.

Lineare interdependent Modelle mit konstanten Koeffizienten lassen sich durch die folgende Matrizen-Endgleichung darstellen:

$$M_0 Y(t) + M_1 Y(t-1) + \dots + M_n Y(t-n) = KE(t) \quad (24.20)$$

wobei M_0 eine blocktrianguläre unzerlegbare Matrix ist. Es ist nicht möglich, auf der Grundlage dieser Form mit Hilfe periodischer Regressionslösungen den Zeitverlauf der endogenen Variablen $Y(t) = [Y_1(t), Y_2(t), \dots, Y_n(t)]$ zu berechnen. Das Modell muß vielmehr in seine reduzierte Form überführt werden. Dieses Verfahren wird im folgenden kurz anhand eines Beispiels beschrieben.

Ist die Determinante $|M_0| \neq 0$, so besitzt M_0 eine Inverse M_0^{-1} , und das
⁴ entnommen mit geringfügigen Änderungen [84, S. 250]
⁵ siehe Seite 524

Gleichungssystem wird durch Multiplikation mit M_0^{-1} in seine reduzierte Form überführt, d.h.

$$Y(t) = M_0^{-1}[-M_1 Y(t-1) - \dots - M_n Y(t-n) + KE(t)]$$

Als Beispiel sei ein von KLEIN entwickeltes aus sechs Gleichungen bestehendes simultanes Modell der amerikanischen Wirtschaft angeführt:
[111, S.68]

$$K0(t) = 16,78 + 0,02NE(t) + 0,23NE(t-1) + 0,8[LP(t)+LS(t)]$$

$$NI(t) = 17,79 + 0,23NE(t) + 0,55NE(t-1) - 0,15KS(t-1)$$

$$LP(t) = 1,6 + 0,42 [VE(t)+SS(t)-LS(t)] + 0,16 [VE(t-1)+SS(t-1)-LS(t-1)] + 0,13 (T-1931)$$

$$VE(t) = K0(t) + NI(t) + SA(t) - SS(t)$$

$$NE(t) = VE(t) - LP(t) - LS(t)$$

$$KS(t) = KS(t-1) + NI(t)$$

Die Symbole besitzen folgende Bedeutung:

K0 = Konsum

NI = Nettoinvestition

LP = Lohnsumme der Privat-Beschäftigten

LS = Lohnsumme der vom Staat Beschäftigten (EXOGENE VAR.)

VE = Volkseinkommen

NE = Nichtarbeitnehmereinkommen

KS = Kapitalstock

SS = Steuersumme (EXOGENE VAR.)

SA = Staatsausgaben (EXOGENE VAR.)

T = Zeitindex (Jahr) (EXOGENE VAR.)

Bringt man sämtliche endogenen unverzögerten Variablen auf die linke Seite der Gleichungen, so ergibt sich:

$$K0(t) - 0,8LP(t) - 0,02NE(t) = 16,78 + 0,23NE(t-1) + 0,8LS(t)$$

$$NI(t) - 0,23NE(t) = 17,79 + 0,55NE(t-1) - 0,15KS(t-1)$$

$$LP(t) - 0,42VE(t) = 1,6 + 0,42[SS(t)-LS(t)] + 0,16[VE(t-1)+SS(t-1)-LS(t-1)] + 0,13(T-1931) \quad (24.21)$$

$$VE(t) - K0(t) - NI(t) = SA(t) - SS(t)$$

$$VE(t) - LP(t) - NE(t) = LS(t)$$

$$KS(t) - NI(t) = KS(t-1)$$

Die Koeffizientenmatrix der auf der linken Seite des Gleichungssystems stehenden unverzögerten endogenen Variablen ist

$$M = \begin{bmatrix} KO & NI & LP & VE & NE & KS \\ 1 & 0 & -0,8 & 0 & -0,02 & 0 \\ 0 & 1 & 0 & 0 & -0,23 & 0 \\ 0 & 0 & 1 & -0,42 & 0 & 0 \\ -1 & -1 & 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 1 & -1 & 0 \\ 0 & -1 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Die inverse Matrix M^{-1} ergibt sich mit

$$M^{-1} = \begin{bmatrix} 1,67 & 0,67 & 1,148 & 0,67 & -0,187 & 0 \\ 0,257 & 1,257 & -0,089 & 0,257 & -0,294 & 0 \\ 0,809 & 0,809 & 1,445 & 0,809 & -0,202 & 0 \\ 1,927 & 1,927 & 1,06 & 1,927 & -0,482 & 0 \\ 1,118 & 1,118 & -0,385 & 1,118 & -1,28 & 0 \\ 0,257 & 1,257 & -0,089 & 0,257 & -0,294 & 1 \end{bmatrix}$$

Mit den Definitionen

$$B(t) = \begin{bmatrix} B_1(t) \\ B_2(t) \\ B_3(t) \\ B_4(t) \\ B_5(t) \\ B_6(t) \end{bmatrix} = \begin{bmatrix} 16,78 + 0,23 NE(t-1) + 0,8 LS(t) \\ 17,79 + 0,55 NE(t-1) - 0,15 KS(t-1) \\ 1,6 + 0,42 [SS(t) - LS(t)] + 0,16 [VE(t-1) + SS(t-1) - LS(t-1)] + 0,13 (T-1931) \\ SA(t) - SS(t) \\ LS(t) \\ KS(t-1) \end{bmatrix} \quad (24.22)$$

und

$$x(t) = \begin{bmatrix} KO(t) \\ NI(t) \\ LP(t) \\ VE(t) \\ NE(t) \\ KS(t) \end{bmatrix}$$

läßt sich das Gleichungssystem (24.21) durch

$$MX(t) = B(t)$$

darstellen. Nach $X(t)$ aufgelöst ergibt sich die reduzierte Gleichungsform

$$X(t) = M^{-1}B(t)$$

und damit das voll ausgeschriebene Gleichungssystem der reduzierten Form

$$\begin{aligned} C(t) &= 1,67B_1(t) + 0,67B_2(t) + 1,148B_3(t) + 0,67B_4(t) - 0,187B_5(t) \\ I(t) &= 0,257B_1(t) + 1,257B_2(t) - 0,089B_3(t) + 0,257B_4(t) - 0,294B_5(t) \\ P(t) &= 0,809B_1(t) + 0,809B_2(t) + 1,445B_3(t) + 0,809B_4(t) - 0,202B_5(t) \\ V(t) &= 1,927B_1(t) + 1,927B_2(t) + 1,06B_3(t) + 1,927B_4(t) - 0,482B_5(t) \\ N(t) &= 1,118B_1(t) + 1,118B_2(t) - 0,385B_3(t) + 1,118B_4(t) - 1,28B_5(t) \\ K(t) &= 0,257B_1(t) + 1,257B_2(t) - 0,089B_3(t) + 0,257B_4(t) - \\ &\quad - 0,294B_5(t) + B_6(t) \end{aligned} \quad (24.23)$$

Dieses Gleichungssystem stellt mit (24.22) ein rekursives Modell dar. Damit ist es gelungen, das ursprünglich interdependenten Modell in ein rekursives zu überführen.

Auf der Grundlage dieses rekursiven Ansatzes ist es möglich, den Zeitpfad der Modellvariablen durch periodische Regressionslösungen zu ermitteln. Die Technik dieses Verfahrens wird anhand des entwickelten Modells von KLEIN im Abschnitt über die Simulation dynamischer Modelle behandelt.⁶

b) Zeitpfadbestimmung in nichtlinearen interdependenten Modellen

Zur Lösung nichtlinearer simultaner Differenzengleichungssysteme wird heute vorwiegend das Gauß-Seidel-Verfahren verwendet. Die Verläufe der endogenen Variablen nahezu aller größeren nichtlinearen interdependenten Modelle werden heute unter Anwendung dieses Verfahrens ermittelt. Auch bei der Retrodiktion nichtlinearer dynamischer Modelle, die eine Lösung simultaner Gleichungen erfordert, ist das Verfahren anwendbar.⁷

⁶ Siehe Seite 533

⁷ Vgl. Seite 471f.

Das Gauß-Seidel-Verfahren soll am Beispiel eines simultanen Systems mit zwei endogenen Variablen demonstriert werden. Als Ausgangspunkt dient das Gleichungssystem:

$$\begin{aligned} Y_1(t) &= F_1[Y_1(t), Y_2(t), Y_1(t-1), Y_2(t-1), E_1(t)] \\ Y_2(t) &= F_2[Y_1(t), Y_2(t), Y_1(t-1), Y_2(t-1), E_2(t)] \end{aligned} \quad (24.24)$$

$F_1[\dots]$ und $F_2[\dots]$ bilden nichtlineare Funktionen, welche es nicht erlauben, durch mathematische Umformungen die reduzierten Gleichungen von $Y_1(t)$ und $Y_2(t)$ abzuleiten. Aus diesem Grunde ist eine periodische Regressionslösung nicht anwendbar.

Die Lösung des Gleichungssystems ist daher nur mit Hilfe eines Näherungsverfahrens möglich. Das hier verwendete Gauß-Seidel-Verfahren erfordert die Bestimmung von zwei Anfangswerten $Y_1^0(t)$ und $Y_2^0(t)$, die wir in die rechte Seite von Gleichung (24.24) für $Y_1(t)$ und $Y_2(t)$ einsetzen. Wir erhalten damit

$$\begin{aligned} Y_1^1(t) &= F_1[Y_1^0(t), Y_2^0(t), Y_1(t-1), Y_2(t-1), E_1(t)] \\ Y_2^1(t) &= F_1[Y_1^0(t), Y_2^0(t), Y_1(t-1), Y_2(t-1), E_2(t)] \end{aligned}$$

Die zweite Iteration liefert

$$\begin{aligned} Y_1^2(t) &= F_1[Y_1^1(t), Y_2^1(t), Y_1(t-1), Y_2(t-1), E_1(t)] \\ Y_2^2(t) &= F_1[Y_1^1(t), Y_2^1(t), Y_1(t-1), Y_2(t-1), E_2(t)] \end{aligned}$$

Nach v Iterationen erhalten wir

$$\begin{aligned} Y_1^v(t) &= F_1[Y_1^{v-1}(t), Y_2^{v-1}(t), Y_1(t-1), Y_2(t-1), E_1(t)] \\ Y_2^v(t) &= F_1[Y_1^{v-1}(t), Y_2^{v-1}(t), Y_1(t-1), Y_2(t-1), E_2(t)] \end{aligned}$$

Die vorherbestimmten Variablen, d.h. $Y_1(t-1), Y_2(t-1), E_1(t)$ und $E_2(t)$ ändern ihre Werte während des Iterationsprozesses nicht.

Falls der Prozeß konvergiert, d.h. die Differenzen $Y_1^v(t) - Y_1^{v-1}(t)$ und $Y_2^v(t) - Y_2^{v-1}(t)$ mit wachsendem v abnehmen, muß eine Abbruchvorschrift definiert werden. Als Abbruchkriterium kann man beispielsweise

$$\left| \frac{Y_1^v(t) - Y_1^{v-1}(t)}{Y_1^{v-1}(t)} \right| < 10^{-3} \quad \text{und} \quad \left| \frac{Y_2^v(t) - Y_2^{v-1}(t)}{Y_2^{v-1}(t)} \right| < 10^{-3} \quad Y_1^{v-1}, Y_2^{v-1} \neq 0$$

wählen.

Die Geschwindigkeit der Konvergenz des Iterationsprozesses kann erhöht werden, wenn man den aus einer Gleichung gewonnenen neuen Wert der endogenen unverzögerten Variablen sofort als Eingangsgröße für die Berechnung der nachfolgenden endogenen Variablen verwendet.

Im Falle unseres Beispiels also nach dem Schema

$$Y_1^V(t) = F_1[Y_1^{V-1}(t), Y_2^{V-1}(t), Y_1(t-1), Y_2(t-1), E_1(t)]$$

$$Y_2^V(t) = F_2[Y_1^V(t), Y_2^{V-1}(t), Y_1(t-1), Y_2(t-1), E_2(t)]$$

Die Konvergenz des Verfahrens ist nicht zwingend gewährleistet. Sie hängt u.a. auch davon ab, welche der endogenen Variablen in einer Gleichung als unabhängige Variablen gewählt werden.

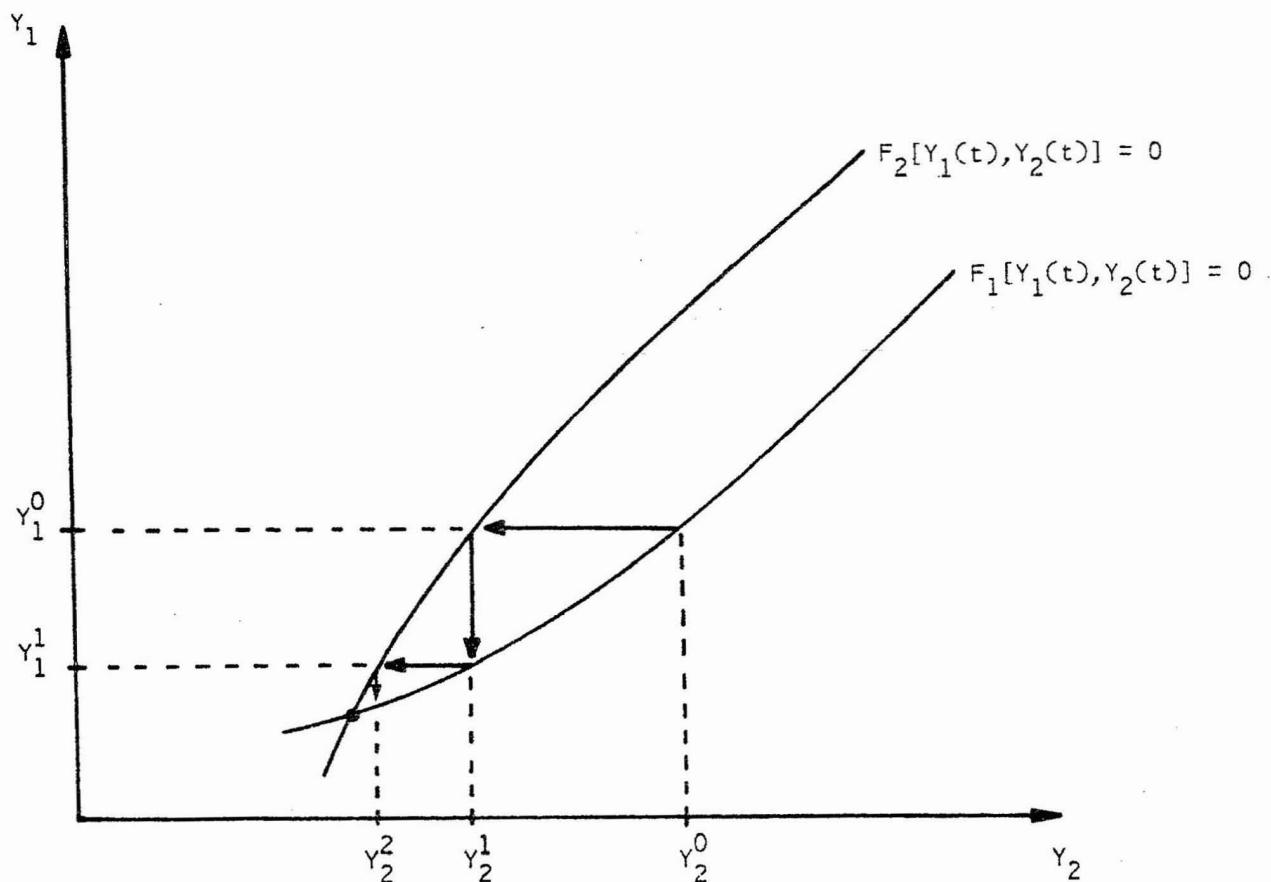


Abb. 24.8 Graphische Darstellung eines konvergierenden Gauß-Seidel-Verfahrens zur Lösung eines nichtlinearen Systems von zwei Gleichungen

Betrachten wir beispielsweise das simultane nichtlineare Gleichungssystem

$$F_1[Y_1(t), Y_2(t)] = 0 \quad (24.25)$$

$$F_2[Y_1(t), Y_2(t)] = 0$$

Die Kurvenzüge in Abbildung 24.8 beschreiben die Koordinatenwerte von F_1 und F_2 . Gesucht sind die Koordinatenwerte des Punktes G, die allein (24.25) befriedigen. Wir unterstellen, es sei möglich (24.25) in die Form

$$Y_1(t) = G_1[Y_2(t)] \quad (24.26)$$

$$Y_2(t) = G_2[Y_1(t)] \quad (24.27)$$

zu überführen. Wählt man als Anfangswert für Y_2 in (24.26) den Wert Y_2^0 , so erhält man, wie aus Abbildung 24.8 zu erkennen ist, für Y_1 den Wert Y_1^0 . Dieser in (24.27) eingesetzt ergibt Y_2^1 usw.; man erkennt, daß der Prozeß gegen G strebt.

Grundsätzlich wäre es aber auch möglich gewesen, das Gleichungssystem nach jeweils einer anderen Variablen aufzulösen, d.h.

$$Y_2(t) = H_1[Y_1(t)]$$

$$Y_1(t) = H_2[Y_2(t)]$$

Verfolgen wir den Iterationsprozeß anhand von Abbildung 24.9, dann erkennen wir, daß auch bei der Wahl desselben Anfangswertes Y_2^0 keine Konvergenz zustande kommt.

Die vorangegangenen Ausführungen bezogen sich nur auf ein simultanes Gleichungssystem mit zwei Variablen. Sie lassen sich jedoch auf ein System mit n Variablen erweitern.

Für den Fall ohne sukzessive Einsetzung während eines Iterations schrittes erhalten wir das Schema

$$Y_j^v(t) = F_j[Y_1^{v-1}, Y_2^{v-1}, \dots, Y_i^{v-1}, \dots, Y_n^{v-1}] \quad j=1, 2, \dots, n \quad (24.28)$$

Die Konvergenz kann auch dadurch erreicht werden, daß statt der Variablen Y_i^{v-1} der Ausdruck

$$wY_i^{v-1} + (1-w)Y_i^{v-2}$$

gewählt wird, in welchem w in der Regel zwischen Null und Eins liegt, in manchen Fällen aber auch, um eine Konvergenz herbeizuführen, größer als Eins gewählt werden muß. Wie erwähnt, kann die Konvergenz beschleunigt werden, wenn man bei der Berechnung von Y_j im v -ten Iterationsschritt die Variablenwerte Y_1^{v-1} bis Y_{i-1}^{v-1} in (24.27) durch die Variablenwerte Y_1^v bis Y_{i-1}^v ($i=j$) ersetzt, welche bereits berechnet wurden.

Nach einer entsprechenden Zahl von Ansätzen gelingt es oft, einen konvergierenden Prozeß zu erhalten.

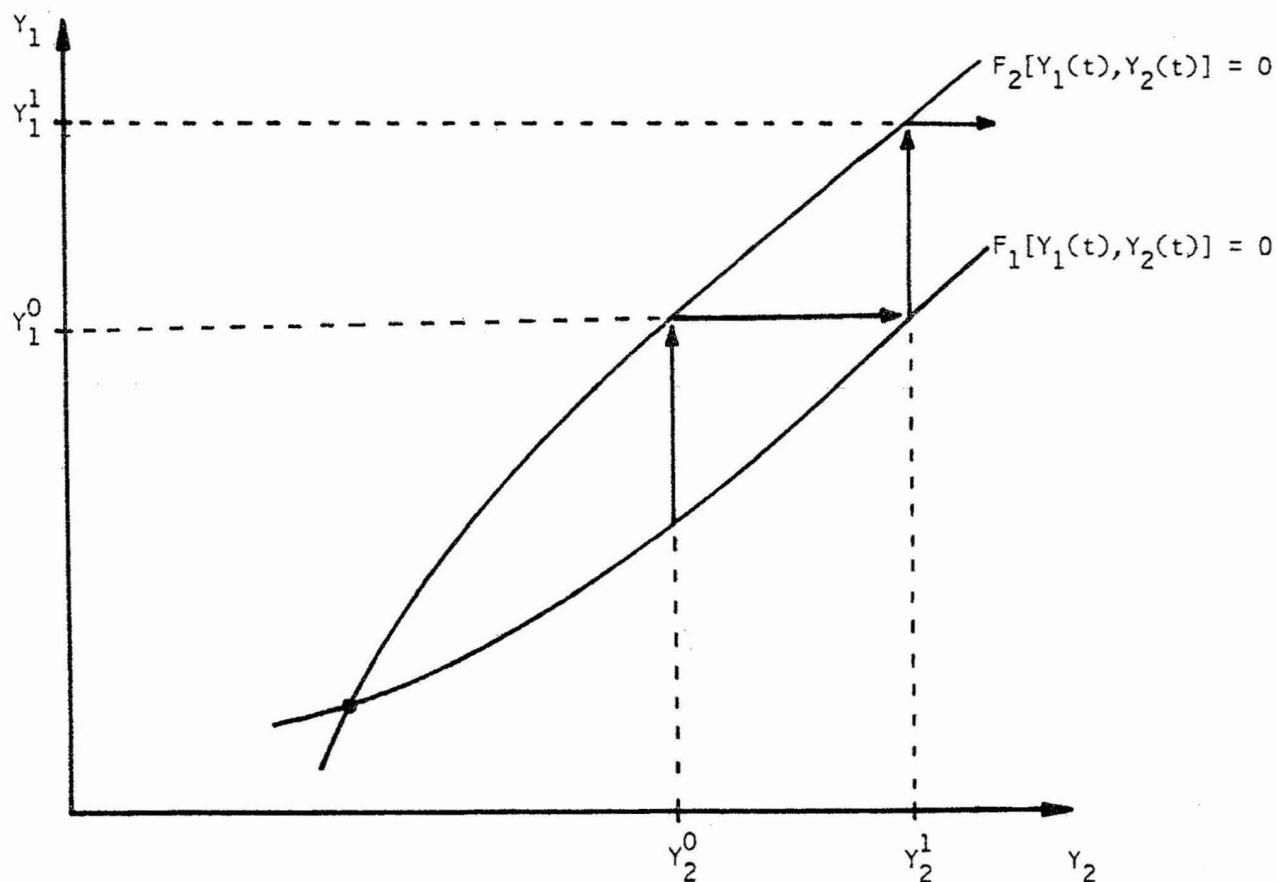


Abb. 24.9 Graphische Darstellung eines nicht konvergierenden Gauß-Seidel-Verfahrens zur Lösung eines nichtlinearen Systems von zwei Gleichungen

2.5. Zerlegbare, unzerlegbare und annähernd zerlegbare Modellformen

Bisher haben wir uns vorwiegend mit dem Verhalten einzelner endogener Modellvariablen befaßt. Für die Beurteilung eines Modells ist es jedoch auch von Interesse, die Verknüpfung zwischen seinen endogenen Variablen zu studieren.

So kann man der Frage nachgehen, ob sich in einem dynamischen Modell eine Gruppe von Variablen finden läßt, die das restliche System beeinflußt, ohne wiederum von diesem Restsystem beeinflußt zu werden. Im Gegensatz zu diesem Fall ist es auch instruktiv zu wissen, ob etwa alle endogenen Variablen direkt oder indirekt miteinander verknüpft sind. Im folgenden werden bestimmte Modelltypen entwickelt, denen sich die oben beschriebenen Verknüpfungsweisen eindeutig zuordnen lassen. Daran anschließend wird der Versuch unternommen, ein Maß für die Verknüpfungsintensität oder auch Komplexität eines Modells zu entwickeln. Schließlich wird die Frage diskutiert, ob und bis zu welchem Grade es zulässig ist, ein nur 'lose' mit dem Hauptsystem verbundenes Subsystem isoliert zu untersuchen.

2.5.1. Begriffliche Klärung und empirische Interpretation

Für die Verknüpfung einer endogenen Variablen $Y_1(t)$ mit einer anderen endogenen Variablen $Y_2(t)$ ist es unmaßgeblich, ob $Y_1(t)$ verzögert oder unverzögert auf $Y_2(t)$ einwirkt. Die Unterscheidung zwischen verzögerten und unverzögerten endogenen Variablen ist daher für die Darstellung der Verknüpfungsstruktur eines Systems bedeutungslos.

Zur Analyse der Modellverknüpfung wird ein parametrisch-generelles Modell wie

$$\begin{aligned} Y(t) &= C(t) + I_i(t) + I_a(t) \\ C(t) &= \alpha Y(t-1) \\ I_i(t) &= \beta [C(t) - C(t-1)] \end{aligned}$$

auf seine nichtparametrische Form unter Vernachlässigung der exogenen Variablen reduziert. In unserem Beispiel erhält man

$$\begin{aligned} Y &= F[C, I_i] \\ C &= F[Y] \\ I_i &= F[C] \end{aligned} \tag{25.1}$$

Die exogenen Variablen bleiben bei dieser Darstellung unberücksichtigt, und auch die spezielle Art der Verknüpfung zwischen den endogenen Variablen ist nicht mehr von Interesse. Die unabhängigen Variablen in den Gleichungen repräsentieren jeweils den Einfluß der entsprechenden verzögerten oder auch unverzögerten Variablen.

In Analogie zur Ermittlung der Strukturmatrix der endogenen unverzögerten Variablen können wir eine Strukturmatrix der endogenen Variablen entwickeln, die als Verknüpfungsmatrix bezeichnet werden soll.

Um die Verknüpfungsweisen eines dynamischen Modells besser beurteilen zu können, ist es erstrebenswert, seine Verknüpfungsmatrix durch entsprechenden Zeilen- und Spaltenaustausch (Permutationen) auf bestimmte standardisierte Grundformen zu überführen.

Gelingt es, die Verknüpfungsmatrix eines Modells in eine vollkommen zerlegbare Matrix zu überführen, dann zeigt sich, daß der aufgestellte Modellansatz aus zumindest zwei voneinander unabhängigen Modellen besteht.

Eine vollkommen zerlegbare Matrix ist eine quadratische Matrix der Form

$$\begin{bmatrix} M_1 & 0 \\ M_2 & \ddots \\ 0 & \ddots & M_r \end{bmatrix}$$

Sie zeichnet sich dadurch aus, daß außer den Untermatrizen auf der Hauptdiagonalen alle übrigen Elemente Nullmatrizen sind.

Als Beispiel sei das dynamische Modell

$$Y_1(t) = 0,5Y_1(t-1) + Y_3(t)$$

$$Y_2(t) = Y_4(t) - 0,2Y_2(t-1)$$

$$Y_3(t) = 0,1Y_1(t)$$

$$Y_4(t) = Y_4(t-1) + Y_2(t)$$

angeführt, dessen Verknüpfungsmatrix

$$\begin{matrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \end{matrix} \left[\begin{array}{cccc} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{array} \right]$$

auf die Grundform

$$\begin{matrix} Y_1 \\ Y_3 \\ Y_4 \\ Y_2 \end{matrix} \left[\begin{array}{cccc} 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{array} \right]$$

überführt werden kann. Man erkennt, daß Y_1 und Y_3 auf der einen und Y_4 und Y_2 auf der anderen Seite zwei voneinander völlig unabhängige Modelle bilden. Dieser Fall eines vollkommen zerlegbaren Modells dürfte relativ selten auftreten.

Wichtiger sind dagegen die hierarchisch zerlegbaren Modelle. Ein solcher Typ liegt immer dann vor, wenn es gelingt, die Verknüpfungsmatrix eines Modells in eine unzerlegbare blocktriangulare Matrix zu überführen, d.h. eine Matrix der Form¹

$$\left[\begin{array}{ccccc} M_1 & 0 & 0 & \dots & 0 & 0 \\ * & M_2 & 0 & \dots & 0 & 0 \\ * & * & M_3 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ * & * & * & \dots & M_{r-1} & 0 \\ * & * & * & \dots & * & M_r \end{array} \right]$$

Dabei wird unterstellt, daß zumindest eine der durch * gekennzeichneten Matrizen in jeder Spalte keine Nullmatrix ist.

¹ Vgl. Seite 327

Anhand einer solchen Matrix können wir uns verdeutlichen, warum man von einem hierarchisch zerlegbaren Modell sprechen kann. Bezeichnen wir mit dem Vektor $[Y_1]$ die der Matrix M_1 zugeordneten Variablen und definieren entsprechend die Vektoren $[Y_2]$ bis $[Y_r]$, so können wir folgende hierarchische Gliederung erkennen: $[Y_1]$ beeinflußt zumindest eine der endogenen Variablen in $[Y_2]$ bis $[Y_r]$, wird aber von keiner dieser Variablen beeinflußt.

$[Y_1]$ und $[Y_2]$ beeinflussen zwar zumindest eine der Variablen aus $[Y_3]$ bis $[Y_r]$, werden aber von diesen nicht beeinflußt. Verallgemeinernd erkennt man, daß die Variablengruppen $[Y_1], \dots, [Y_i]$ nur einseitig einen Einfluß auf die Gruppen $[Y_{i+1}], \dots, [Y_r]$ ausüben.

Man erhält daher ein hierarchisches Beeinflussungsverhältnis der Variablengruppen, an dessen Basis die Variablengruppe $[Y_1]$ anzurufen ist, während die Gruppe $[Y_r]$ die Spitze der Hierarchie einnimmt.

Die Kenntnis der hierarchischen Struktur eines dynamischen Modells zeigt die Möglichkeiten der 'Weitergabe' bestimmter Systemverhaltensweisen unter den Variablengruppen. Auch wird deutlich, welche Variablen man im Rahmen der Gruppe $[Y_1]$ von dem Gesamtmodell 'abkoppeln', d.h. isoliert untersuchen kann. Zugleich zeigt sich, welche der zusammengefaßten hierarchisch aufsteigenden Gruppen $[Y_1]$ und $[Y_2]$ oder $[Y_1], [Y_2]$ und $[Y_3]$ usw. losgelöst von den übrigen Modellteilen untersucht werden kann.

Im folgenden soll die Grundform der Verknüpfungsmatrix eines einfachen linearen Modells bestimmt werden. Wir gehen aus von dem Ansatz

$$\begin{aligned} Y_1(t) &= 0,5Y_3(t-1) + E_1(t) \\ Y_2(t) &= 1,5Y_5(t) + 3,0Y_4(t) \\ Y_3(t) &= 0,25Y_5(t-1) + E_2(t) \\ Y_4(t) &= 0,05Y_3(t-1) + Y_2(t-1) \\ Y_5(t) &= 0,1Y_1(t-1) + Y_2(t-1) \end{aligned} \tag{25.2}$$

Die Verknüpfungen zwischen den (verzögerten und unverzögerten) endogenen Systemvariablen können auf das folgende Gleichungssystem reduziert werden.

$$\begin{aligned}
 Y_1 &= F[Y_3] \\
 Y_2 &= F[Y_5, Y_4] \\
 Y_3 &= F[Y_5] \\
 Y_4 &= F[Y_3, Y_2] \\
 Y_5 &= F[Y_1]
 \end{aligned}$$

Die Boolesche Relationenmatrix ergibt:

	Y_1	Y_2	Y_3	Y_4	Y_5
Y_1			1		
Y_2				1	1
$M = Y_3$					1
Y_4		1	1		
Y_5	1				

Die Quadratmatrix von M , d.h.

	Y_1	Y_2	Y_3	Y_4	Y_5
Y_1					1
Y_2	1	1	1		
$M^2 = Y_3$	1				
Y_4				1	1
Y_5			1		

zeigt, daß die Variablen Y_2 und Y_4 eine maximale Schleife bilden, weil sie keine andere Variable des Systems beeinflussen. Durch Streichen der Spalten und Zeilen von Y_2 und Y_4 erhält man die Matrix M_2^*

	Y_1	Y_3	Y_5
Y_1			1
Y_3	1		
Y_5		1	

Durch Streichen der Spalten und Zeilen von Y_2 und Y_4 in der ursprünglichen Matrix M erhält man M^*

	Y_1	Y_3	Y_5
Y_1		1	
Y_3			1
Y_5	1		

Die Matrix $M_3 = M_2^* \cdot M^*$

	Y_1	Y_3	Y_5
Y_1	1		
Y_3		1	
Y_5			1

zeigt, daß die Variablen Y_1 , Y_3 und Y_5 eine maximale Schleife bilden.
Die entsprechende Umordnung der Booleschen Relationenmatrix ergibt:

	Y_1	Y_5	Y_3	Y_4	Y_2
Y_1			1		
Y_5	1				
Y_3		1			
Y_4			1		1
Y_2		1		1	

und damit erhält man die Verknüpfungsmatrix

	Y_1	Y_5	Y_3	Y_4	Y_2
Y_1	1		1		
Y_5	1	1			
Y_3		1	1		
Y_4			1	1	1
Y_2		1		1	1

(25.3)

Der vorliegende Ansatz erweist sich als ein zweistufig hierarchisch zerlegbares Modell. Man erkennt, daß die Gruppe der Variablen Y_1 , Y_5 und Y_3 die Variablengruppe Y_4 und Y_2 beeinflußt, ein rückwirkender Einfluß jedoch nicht gegeben ist. Interessieren den Modellentwickler nur die Implikationen der Variablen Y_1 , Y_5 und Y_3 , dann kann er das Gleichungssystem, welches diese Variablen erklärt, isoliert untersuchen.

Gelingt es nicht, die Verknüpfungsmatrix eines dynamischen Modells in eine blocktriangulare Form zu überführen, dann handelt es sich um ein unzerlegbares Modell.

Als Beispiel kann das MA-Modell dienen, dessen Gleichungssystem der endogenen Verknüpfung (25.1) zu der Booleschen Matrix

	C	I_i	Y
C			1
$M = I_i$	1		
Y	1	1	

führt.

Die Boole'sche Quadratmatrix von M ergibt

$$M^2 = \begin{array}{c|ccc} & C & I_1 & Y \\ \hline C & 1 & 1 & \\ & & & \\ & & & 1 \\ \hline I_1 & & & \\ & & & \\ & & & \\ \hline Y & 1 & & 1 \end{array}$$

Es zeigt sich, daß das System keine maximale Schleife besitzt, die nur über zwei Variablen führt, denn es läßt sich keine Spalte finden, die nur auf der Hauptdiagonalen eine 1 besitzt. Eine weitere Matrixmultiplikation liefert

$$M^3 = \begin{array}{c|ccc} & C & I_1 & Y \\ \hline C & 1 & & 1 \\ & & & \\ & & & \\ \hline I_1 & 1 & 1 & \\ & & & \\ & & & \\ \hline Y & 1 & 1 & 1 \end{array}$$

Da die gesamte Hauptdiagonale mit Einsen besetzt ist, befinden sich alle drei endogenen Variablen in einer geschlossenen Beeinflussungskette, d.h. einer Schleife. Eine graphische Darstellung der Verknüpfungsstruktur ergibt

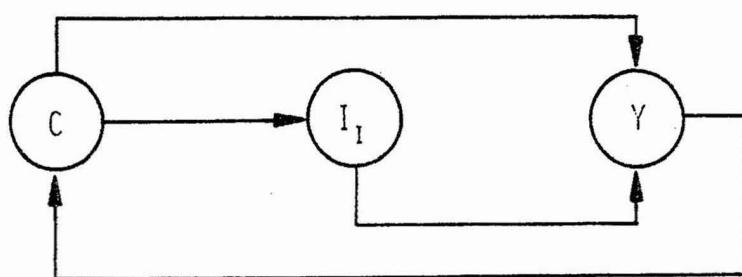


Abb. 25.1 Graphische Darstellung der Verknüpfungsstruktur eines MA-Systems

Es kann gezeigt werden, daß es in einem unzerlegbaren Modell stets eine geschlossene Beeinflussungskette gibt, die über alle endogenen Variablen führt.² Eine derartige geschlossene Beeinflussungskette soll als Totalschleife bezeichnet werden. Erweist sich ein Modell als unzerlegbar, so wird deutlich, daß man keine Submodelle isoliert untersuchen kann. Auch spricht in diesem Falle vieles für die Vermutung, daß sich spezielle Systemverhaltensweisen eines Teilbereiches wie ein fluktuerendes Verhalten über die vorhandene Totalschleife auf das gesamte System ausbreiten werden.

Die vorangegangenen Erörterungen erlauben nunmehr, das bereits im Abschnitt 2.3.1. angekündigte Klassifizierungskriterium von zyklischen Modellen zu entwickeln.³ In einem rekursiven Modell ist der hier verwendete Begriff einer Schleife identisch mit einem Feedbackkreis. Wenn interdependente Modelle nur Schleifen zwischen ihren endogenen unverzögerten Variablen aufweisen, ist es dagegen wohl kaum möglich, von einem Feedbackkreis zu sprechen, weil gemeinhin von der Auffassung ausgegangen wird, daß ein Feedbackkreis verzögerte Rückwirkungen beschreibt.

Die Frage, ob ein vorliegendes Modell zyklisch ist, läßt sich wie folgt beantworten. Enthalten die primären Hypothesengleichungen autoregressive Variablen, dann liegt stets ein zyklisches Modell vor. Sind solche autoregressiven Beziehungen nicht vorhanden, so ist ein Modell nur dann zyklisch, wenn es zumindest eine Schleife enthält, die im Falle eines interdependenten Modells eine verzögernde Beziehung enthalten muß. Besitzt das Modell eine Totalschleife, so besitzt das Modell einen Feedbackkreis, der alle endogenen Variablen umfaßt.

2 Siehe Seite 361

3 Siehe Seite 286

2.5.2. Verknüpfungs- und Komplexitätsmaße dynamischer Modelle

Bisher haben wir ein System durch Eigenschaften gekennzeichnet, die anhand des entsprechenden Systemmodells aufzeigbar sind. In diesem Sinne wurde von einem nichtlinearen oder einem offenen System gesprochen. Diesem Sprachgebrauch entsprechend müßte ein komplexes System durch ein komplexes Modell darstellbar sein. Beim Begriff der Komplexität ist jedoch eine solche Begriffsübertragung nicht möglich. Oder anders ausgedrückt: es ist nicht generell möglich zu sagen, daß ein hoch (bzw. wenig) komplexes Modell auch ein hoch (bzw. wenig) komplexes System repräsentiere.

Als Maß für die Komplexität eines Modells könnte man eine bestimmte Gewichtung aus den Größen 'Stärke der Nichtlinearität', 'notwendige Variablenzahl' und 'Verknüpfungsintensität der endogenen Variablen' ansehen.

Die Größe oder Stärke der Nichtlinearität eines Modells ist schwer präzise zu fassen.⁴ Im folgenden soll der Begriff der Modellkomplexität allein anhand der Bestimmungsgrößen 'notwendige Variablenzahl' und 'Verknüpfungsintensität' präzisiert werden.

Unter der 'notwendigen Variablenzahl' soll die Zahl der Variablen verstanden werden, die zur adäquaten Erfassung eines Systems erforderlich ist. Von den notwendigen Variablen sind die interessierenden Variablen zu unterscheiden, d.h. die Variablen, welche der Modellentwickler für explikative, prognostische oder normative Zwecke benötigt. Die Zahl der notwendigen Variablen ist daher größer oder gleich der Zahl der interessierenden Variablen.

Die Bestimmung der Zahl der notwendigen Variablen ist nicht unproblematisch. Betrachten wir beispielsweise ein MA-Modell, d.h. einen Ansatz der Form

$$\begin{aligned} Y(t) &= C(t) + I_i(t) + I_a(t) \\ C(t) &= \alpha Y(t-1) \\ I_i(t) &= \beta [C(t) - C(t-1)] \end{aligned}$$

⁴ Vgl. Seite 272

und unterstellen, den Modellentwickler interessiere allein die Variable $Y(t)$. Dann kann man zu der Auffassung kommen, die Zahl der notwendigen Variablen sei drei. Da es jedoch nicht grundsätzlich ausgeschlossen ist, daß ein Modellentwickler direkt die Endgleichungshypothese (12.9), d.h.

$$Y(t) = (\alpha + \alpha\beta)Y(t-1) - \alpha\beta Y(t-2) + I_a(t)$$

formuliert und diese Hypothese ausreicht, um die interessierende Variable $Y(t)$ zu beschreiben, ist die Zahl der notwendigen Variablen in diesem Fall 1.

Im Falle nichtlinearer Systeme ist jedoch nicht immer ein eindeutiges Urteil möglich, ob sich die Zahl der notwendigen Variablen auf die Zahl der interessierenden Variablen reduzieren läßt. Angesichts dieser Sachlage kann letztlich nur die Zahl der tatsächlichen Variablen eines Modells als eine klar definierte Bestimmungsgröße der Komplexität herangezogen werden. Dies hat zur Folge, daß ein bestimmtes System durch Modelle beschrieben werden kann, die zwar miteinander logisch konsistent sind, sich aber in der Variablenzahl unterscheiden. Es ist nicht von der Hand zu weisen, daß viele sozioökonomische Modelle dadurch 'komplexer' gemacht werden, daß ihre Entwickler eine Fülle von Zwischenvariablen einführen, mit der Folge, daß die Anzahl der Gleichungen enorme Dimensionen annimmt.

Neben der leicht bestimmbaren tatsächlichen Anzahl der endogenen Variablen eines Modells ist eine Maßeinheit für die Intensität der Verknüpfung zwischen den endogenen Variablen zu entwickeln.

Als Maß der Komplexität eines Modells soll das Produkt aus der Anzahl der Schleifen seiner Verknüpfungsmatrix mit der Zahl der vorhandenen endogenen Modellvariablen dienen, d.h.

$$\text{Modellkomplexität} = \left[\begin{matrix} \text{Anzahl der existierenden Schleifen} & \times & \text{Anzahl der endogenen Modellvariablen} \end{matrix} \right]$$

Das in Abbildung 25.1 dargestellte MA-Modell besitzt beispielsweise die beiden Schleifen C, I_1, Y, C und C, Y, C . Seine Modellkomplexität beträgt daher $M=2 \cdot 3=6$.

Der Begriff der Modellkomplexität ist allein modell- und nicht system-

spezifisch aufzufassen. Gelingt es beispielsweise, von einem MA-Modell die Endgleichungen von Y , C und I_i aufzustellen, dann besitzt dieses Modell eine Komplexität von Null.

Der Wissenschaftstheoretiker HEMPEL verwendet, wie erwähnt, im Zusammenhang mit der Explikation des Begriffes einer wissenschaftlichen Erklärung den Terminus 'Minimalgesetz'. Unter der Zielsetzung, daß die Redundanz eines Systems von Aussagen zur Beschreibung der Realität so niedrig wie möglich sein soll, ist es stets erstrebenswert, solche Minimalgesetze zu ermitteln. Die Endgleichung einer endogenen Variablen kann als ein derartiges Minimalgesetz aufgefaßt werden.⁵

Der höchste Grad der Redundanzverminderung eines Modells, der durch die Darstellung der Endgleichungen zum Ausdruck gebracht wird, führt also zur Modellkomplexität Null. Der hier entwickelte Begriff der Modellkomplexität erweist sich daher als ein Maß für die 'Formulierungsredundanz' des Modells.

Betrachten wir unter diesem Gesichtspunkt noch einmal die verschiedenen Versionen des MA-Modells, so besitzt die Endgleichungsversion

$$\begin{aligned} Y(t) &= (\alpha + \alpha\beta)Y(t-1) - \alpha\beta Y(t-2) + I_a(t) \\ I_i(t) &= (\alpha + \alpha\beta)I_i(t-1) - \alpha\beta I_i(t-2) + \alpha\beta I_a(t-1) - \alpha\beta I_a(t-2) \\ C(t) &= (\alpha + \alpha\beta)C(t-1) - \alpha\beta C(t-2) + \alpha I_a(t-1) \end{aligned}$$

eine Modellkomplexität von

$$M = 0 \cdot 3 = 0$$

Der Ansatz

$$\begin{aligned} Y(t) &= C(t) + I_i(t) + I_a(t) \\ I_i(t) &= \beta[C(t) - C(t-1)] \\ C(t) &= \alpha Y(t-1) \end{aligned}$$

besitzt eine Komplexität von

$$M = 2 \cdot 3 = 6$$

während die Zustandsraumdarstellung des MA-Modells

⁵ Vgl. Seite 46

$$\begin{bmatrix} Y(t) \\ I_i(t) \\ C(t) \end{bmatrix} = \begin{bmatrix} \alpha + \beta\alpha & 0 & -\beta \\ \alpha\beta & 0 & -\beta \\ \alpha & 0 & 0 \end{bmatrix} \begin{bmatrix} Y(t-1) \\ I_i(t-1) \\ C(t-1) \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} E(t)$$

das Komplexitätsmaß von

$$M = 1 \cdot 3 = 3$$

aufweist.

Die Berechnung der Anzahl der existierenden Schleifen eines Modells soll anhand eines Beispieles demonstriert werden.

Betrachten wir die durch die folgenden Funktionen zum Ausdruck kommenden Verknüpfung der endogenen Variablen eines dynamischen Modells

$$Y_1 = F[Y_2, Y_3]$$

$$Y_2 = F[Y_3]$$

$$Y_3 = F[Y_1, Y_2]$$

dann lassen sich die hier zu Tage tretenden Verknüpfungen auch durch das Schema

Gleichungsnr.	Y_1	Y_2	Y_3
1	0	•12	•13
2		0	•23
3	•31	•32	0

zum Ausdruck bringen, indem in jeder Zeile die abhängigen Variablen durch einen leeren, die unabhängigen Variablen durch einen geschwärzten Kreis gekennzeichnet sind.

In unserem Beispiel existiert eine Zweierschleife, weil Y_3 die Gleichung für Y_2 beeinflußt und Y_2 wiederum die Gleichung von Y_3 . Man erhält, wie in dem nachfolgenden Schema graphisch demonstriert, eine geschlossene Kette

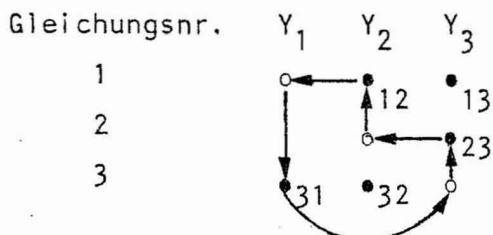
Gleichungsnr.	Y_1	Y_2	Y_3
1	0	•12	•13
2			•23
3	•31	•32	0

Bezeichnen wir nunmehr die Koeffizienten einer beliebigen Verknüpfungsmatrix mit a_{ij} , so wird anhand der graphischen Aufweisung einer Zweierschleife die allgemeine Behauptung evident: Ein Modell besitzt so viele Zweierschleifen wie sich Eins werdende Koeffizientenprodukte

$$a_{\alpha\beta} * a_{\beta\alpha} = 1 \quad \text{mit } \alpha, \beta \in \{1, 2, \dots, n\} \text{ und } \alpha \neq \beta$$

seiner Verknüpfungsmatrix finden lassen.

Für das Auffinden einer Dreierschleife gilt eine analoge Betrachtung. Eine Dreierschleife wird in unserem Beispiel durch das folgenden Schema



beschrieben und durch die Koeffizientenfolge a_{12}, a_{23}, a_{31} charakterisiert. Analog zum Fall einer Dreierschleife können wir feststellen: ein Modell besitzt so viele Dreierschleifen wie sich Eins werdende Koeffizientenprodukte

$$a_{\alpha\beta} * a_{\beta\gamma} * a_{\gamma\alpha} = 1 \quad \text{mit } \alpha, \beta, \gamma \in \{1, 2, \dots, n\} \text{ und } \alpha \neq \beta \neq \gamma$$

seiner Verknüpfungsmatrix finden lassen.

Aus den Beispielen einer Zweier- und Dreierschleife läßt sich die Verallgemeinerung finden, daß es so viele über s Variablen führende Schleifen gibt wie Koeffizientenprodukte

$$a_{i_1 j_1} * a_{i_2 j_2} * \dots * a_{i_v j_v} * \dots * a_{i_s j_s} = 1 \quad \text{mit } i_v \neq j_v \quad \text{und } i_{v+1} = j_v \quad (25.4)$$

existieren, in welchen i_1, i_2, \dots, i_s und j_1, j_2, \dots, j_s Permutationen der Elemente der Menge T bilden, welche eine Teilmenge von $\{1, 2, \dots, n\}$ ist. Eine $n \times n$ Verknüpfungsmatrix ist dann und nur dann unzerlegbar, wenn sich für den Fall $s=n$ zumindest ein Koeffizientenprodukt finden läßt, welches die in (25.4) geforderten Bedingungen erfüllt.⁶

⁶ Zum Beweis [152, S. 109]

Da in diesem Fall sämtliche endogenen Variablen von zumindest einer geschlossenen Beeinflussungskette berührt werden, enthält das Modell auf jeden Fall eine Totalschleife.

Damit wird aber deutlich, daß ein Modell stets nur dann eine Totalschleife besitzt, wenn seine Verknüpfungsmatrix unzerlegbar ist.

Besitzt ein Modell keine Totalschleife, d.h. kann es in eine blocktriangulare Verknüpfungsmatrix mit unzerlegbaren Diagonalblöcken umgewandelt werden, dann umfaßt jeder Diagonalblock zumindest eine Schleife von der Dimension dieser unzerlegbaren Untermatrix, und auch alle weiteren Schleifen niedrigen Grades hängen allein von der Konstellation der Koeffizienten in diesen Untermatrizen ab.

Dies soll am Beispiel der Verknüpfungsmatrix (25.3) des dynamischen Modells (25.2) demonstriert werden. Ersetzt man die Einsen dieser Matrix durch die Elemente a_{ij} , so erhält man die Matrix

a_{11}		a_{13}		
a_{21}	a_{22}			
	a_{32}	a_{33}		
		a_{43}	a_{44}	a_{45}
	a_{52}		a_{54}	a_{55}

Nach dem Gesagten sind nur die Koeffizienten in den beiden Diagonalblöcken für die Schleifenbestimmung von Bedeutung. Es zeigt sich:
Im oberen Diagonalblock gibt es nur eine Dreierschleife

$$a_{21} \quad a_{13} \quad a_{32}$$

und im unteren die Zweierschleife

$$a_{54} \quad a_{45}$$

Die Modellkomplexität beläuft sich daher auf $M=2 \cdot 5=10$.

Zur Bestimmung der Schleifen in einer Verknüpfungsmatrix stehen bestimmte Algorithmen zur Verfügung. [49]

Im Rahmen des Simulationssystems IFICUS, welches zur Simulation sozioökonomischer dynamischer Modelle entwickelt wurde und große Ähnlichkeiten mit dem Programmiersystem DYNAMO besitzt, ist ein Dienstprogramm KISS zur Erkennung von Kreisstrukturen entwickelt worden.

KISS (Kreise im Simulationssystem) erkennt und dokumentiert alle verzögerten Kreisbeziehungen zwischen den endogenen Variablen eines Modells.⁷ KISS liefert folgende Informationen:⁸

- (1) Eine Statistik über die Anzahl der Kreise.

ERKANNT WURDEN 259 KREISE

- (2) Eine Aufstellung der in einem Kreislauf berührten endogenen Variablen:

1)	AR	KV	EWTL	QTWL	WTL
	TLR	TLA			
2)	AR	KV	EWTL	QTWL	WTL
	TLR	TLA	AK	QLB	LB
	PER	PR	PMABA	PINQU	DUMS
	DAR				
3)	DAR				
4)	PMABA	PINQU	REINP		

Die unter 3) angegebene Variable kennzeichnet einen Fall, in dem eine endogene Variable allein von ihrer eigenen Verzögerung beeinflußt wird. Im Sinne der von uns durchgeföhrten Definition darf man in diesem Fall nicht von einer Schleife reden.

- (3) Eine Statistik über die Häufigkeit, mit der eine endogene Variable sowie bestimmte Kettenglieder mit dieser endogenen Variablen in einem Kreis enthalten sind.

⁷ Das System ist wie DYNAMO nur auf rekursive dynamische Modelle ausgelegt. Unverzögerte Kreisbeziehungen, d.h. simultane Gleichungen wären daher Modellierungsfehler.

⁸ Beispiel entnommen [194,S.287]

AR IST IN 140 KREIS(EN) ENTHALTEN
 AR KV IN 140 KREIS(EN)
 ASL IST IN 48 KREIS(EN) ENTHALTEN
 ASL VR IN 48 KREIS(EN)
 KV IST IN 140 KREIS(EN) ENTHALTEN
 KV ETWL IN 79 KREIS(EN)
 KV EWMA IN 61 KREIS(EN)
 AR KV ETWL QTLW WTL
 TLR TLA

Die Stärke der Verknüpfung eines Modells kann auch durch eine weitere Maßzahl beschrieben werden, die in Analogie zu dem im Rahmen der Input-Output-Analyse verwendeten Begriff eines Zirkularitätsgrades dargestellt und als Rückführungsgrad bezeichnet werden soll. Zur Entwicklung dieses Begriffes gehen wir davon aus, daß es gelungen ist, die blocktriangulare Strukturmatrix eines dynamischen Modells zu entwickeln. Als Beispiel nehmen wir die auf Seite 333 dargestellte Matrix. Denken wir uns in dieser Matrix alle Einsen über der Hauptdiagonalen gleich Null gesetzt, dann erhält man die folgende Matrix.

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
x_1									
x_2									
x_3	.								
x_4									
x_5			1	1					
x_6	1								
x_7						1			
x_8				1					
x_9	1	1							

Diese Matrix beschreibt ein System, in dem sich die endogenen Variablen in rekursiver Weise beeinflussen.⁹ Das Pfeildiagramm a zeigt
 9 Vgl. zum rekursiven Aufbau eines Systems von Variablen Seite 316ff

den sich ergebenden Zusammenhang. Im Gegensatz zum Pfeildiagramm b des ursprünglichen Modells enthält es keinen Feedbackkreis, weil die nach oben führenden Pfeile, im folgenden Rückführungen genannt, abgeschnitten sind.

Es liegt nahe davon auszugehen, daß in einem Modell mit wachsender Anzahl der Rückführungen auch die Modellkomplexität zunimmt, weil

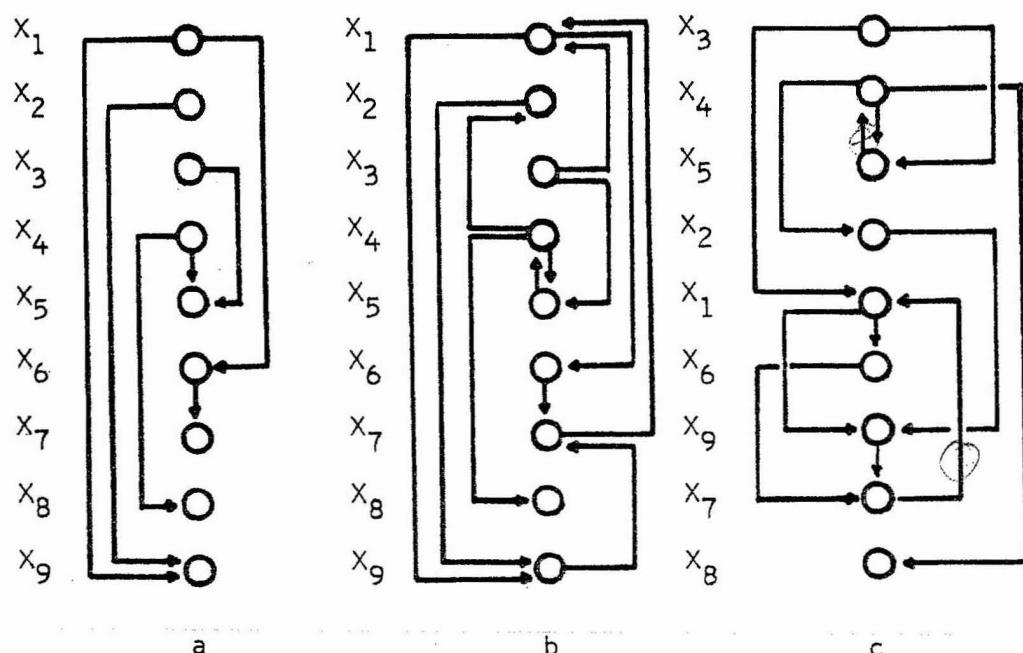


Abb. 25.2 Pfeildiagramme eines multivariablen Modells

damit (tendenziell) auch die Zahl der vorhandenen Feedbackkreise wächst. Die Zahl der Rückführungen ist jedoch kein präziser Begriff.

Zur Verdeutlichung betrachten wir die Pfeildiagramme b und c in Abbildung 25.2. Beide beschreiben dasselbe Modell bei unterschiedlichen Anordnungen der endogenen Variablen. Legt man nun mehr fest, daß die nach oben führenden Richtungspfeile als Anzahl der Rückführungen anzusehen sind, so besitzt das Modell im Fall b fünf, im Falle c dagegen nur zwei Rückführungen. Ihre Anzahl hängt daher von der Anordnung der Variablen ab. Wegen dieser Mehrdeutigkeiten wollen wir einen präziseren Rückführungs begriff einführen, der als die Minimalrückführung eines Modells bezeichnet wird. Er resultiert aus folgender Überlegung:

in dem Pfeildiagramm eines Modells wird eine Richtung als 'rekursive Richtung' festgelegt. Auf der Grundlage dieser Festlegung wird die Forderung erhoben, die Anordnung der Variablen zu finden, bei der die Anzahl der in die nichtrekursive Richtung weisenden Pfeile minimiert wird. Die unter dieser Bedingung aufweisbaren Rückführungen sollen als die Minimalrückführungen eines Modells bezeichnet werden. Unter diesen Umständen bildet auch jede Rückführung das Element zumindest eines Feedbackkreises, so daß bei einer Reduzierung der Rückführungen stets auch die Zahl der wirkenden Feedbackkreise abnimmt. Versucht man das beschriebene Extremierungsverfahren auf formaler Ebene anhand der Verknüpfungsmatrix eines Modells vorzunehmen, so lautet die entsprechende Vorschrift: Finde durch Zeilen- und Spaltenaustausch (Permutationen) die Verknüpfungsmatrix des Modells mit der Eigenschaft, daß die Summe der Elemente über der Hauptdiagonalen minimiert wird.

Das Verfahren zur Ermittlung einer derartigen Matrix wird von HELMSTÄDTER als Triangulation bezeichnet [80]. Diese Namensgebung ist irreführend, da man gewöhnlicherweise in der Mathematik darunter die Gewinnung einer Dreiecksmatrix versteht. Im folgenden soll daher speziell im Hinblick auf die hier zur Diskussion stehende Frage von der Ermittlung einer rückführungsminimalen Verknüpfungsmatrix gesprochen werden. Es stehen verschiedene Algorithmen zur Realisierung rückführungsminimaler Verknüpfungsmatrizen zur Verfügung. [113] Man kann sich fragen, welche Einsichten aus einer rückführungsminimalen Verknüpfungsmatrix gezogen werden können. Definiert man unter Zugrundelegung einer rückführungsminimalen Verknüpfungsmatrix das Verhältnis der rückführenden zur Gesamtzahl der Pfeile als Rückführungsgrad, dann sinken mit wachsendem Rückführungsgrad die Chancen, durch 'Kappen' der Rückführungen ein in der Komplexität reduziertes, aber dennoch empirisch akzeptables Modell zu erhalten. Die rückführungsminimale Verknüpfungsmatrix gibt zudem darüber Auskunft, welche Beziehungen für eine derartige Komplexitätsreduzierung überhaupt in Frage kommen.

Präzise ergibt sich bei Vorliegen einer rückführungsminimalen Verknüpfungsmatrix der Rückführungsgrad ξ aus¹⁰

$$\xi = \sum_{i>j}^n x_{ij} / \sum_{i,j=1}^n x_{ij}$$

x_{ij} beschreiben hierbei die Elemente der rückführungsminimalen Matrix. Es stellt sich die Frage nach dem Rückführungsgrad des durch die Pfeildiagramme b. und c in Abbildung 25.2 beschriebenen Modells. Die Ermittlung der rückführungsminimalen Verknüpfungsmatrix soll ohne Anführung eines streng formalen Verfahrens anhand dieses Modells demonstriert werden.

Wir erinnern uns, daß bereits ein Verfahren beschrieben wurde, aufgrund dessen die auf Seite 333 beschriebene Verknüpfungsmatrix des Modells in ihre blocktriangulare Form überführt wurde. Zur Ermittlung der rückführungsminimalen Matrix genügt es, isoliert im Rahmen der einzelnen Diagonalblöcke, die Zahl der über der Hauptdiagonalen liegenden nicht Null werdenden Elemente zu minimieren. Zu diesem Zweck betrachten wir die auf Seite 338 angeführte blocktriangulare Verknüpfungsmatrix des Modells.

In unserem Beispiel sind nur zwei Blöcke mit Einser-Elementen besetzt. Ein Austausch der Spalten und Zeilen von X_4 und X_5 im ersten Block zeigt, daß sich die Zahl der Einser-Elemente über der Hauptdiagonalen nicht verändert. Im Falle des zweiten Blocks sind $4!=24$ Permutationen möglich, unter denen die Anordnung X_1, X_6, X_9, X_7 rückführungsminimal ist.

Das Pfeildiagramm c in Abbildung 25.2 repräsentiert daher den Fall einer Minimalrückführung. Man erkennt, daß sich unter insgesamt zwölf

10 Der Begriff des Rückführungsgrades ist nicht identisch mit dem von HELMSTÄDTER eingeführten Begriff des Zirkularitätsgrades, weil im Rahmen der statischen Input-Outputrechnung die x_{ij} -Wertgrößen repräsentieren, so daß der Zirkularitätsgrad zugleich den quantitativen Einfluß der Variableneinflüsse berücksichtigt. Im vorliegenden Fall werden die Einflüsse gleichgewichtig behandelt, weil bei nichtlinearen dynamischen Beziehungen eine Erfassung des Beeinflussungsausmaßes durch eine Maßzahl kaum möglich sein dürfte.

Einflußpfeilen zwei rückführende Pfeile befinden, d.h. der Rückführungsgrad des Systems beträgt 0,166.

Zusammenfassend zeigt Abbildung 25.3 die ursprüngliche Verknüpfungsmatrix des betrachteten Modells sowie die aus ihr abgeleitete rückführungsminimale Matrix.

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
x_1			1				1		
x_2				1					
x_3									
x_4					1				
x_5			1	1					
x_6	1								
x_7						1			1
x_8				1					
x_9	1	1							

a

	x_3	x_4	x_5	x_2	x_1	x_6	x_9	x_7	x_8
x_3									
x_4				1					
x_5	1	1							
x_2				1					
x_1	1							1	
x_6					1				
x_9					1	1			
x_7						1	1		
x_8		1							

b

Abb. 25.3 Verknüpfungsmatrix eines primären Modellansatzes (a) sowie die rückführungsminimale Verknüpfungsmatrix desselben Modells (b)

2.5.3. Subsystemabspaltung in dynamischen Modellen

Ein dynamisches Modell beschreibt stets nur einen Teilbereich der Realität, d.h. ein bestimmtes Subsystem. Dieses Subsystem ist, wie die Erfahrung zeigt, fast nie von dem restlichen System völlig isoliert, sondern wird von ihm beeinflußt. Beeinflussungen des Restsystems kommen in offenen Modellen durch die Wahl bestimmter Verläufe der exogenen Variablen zum Ausdruck, während sie in geschlossenen Modellen einfach vernachlässigt werden. In diesem Falle tut man so, als ob es keine Außeneinflüsse gäbe.

Dieses Vorgehen kann, wie eine Untersuchung von SIMON und ANDO zeigt, im Falle linearer geschlossener Systeme unter bestimmten Umständen gerechtfertigt sein. [179]

Die Frage nach der Zulässigkeit einer solchen Subsystemanalyse soll am Beispiel eines aus vier Variablen bestehenden Differenzengleichungssystems

$$Y(t) = A \cdot Y(t-1) \quad (25.5)$$

erörtert werden, welches durch

$$Y(t) = \begin{bmatrix} Y_1(t) \\ Y_2(t) \\ Y_3(t) \\ Y_4(t) \end{bmatrix} \quad A = \left[\begin{array}{cc|cc} 0,20 & 0,50 & 0,05 & 0,00 \\ 0,30 & 0,40 & 0,10 & 0,11 \\ \hline 0,04 & 0,10 & 0,40 & 0,45 \\ 0,13 & 0,08 & 0,50 & 0,30 \end{array} \right] \quad \text{und} \quad Y(0) = \begin{bmatrix} Y_1(0) \\ Y_2(0) \\ Y_3(0) \\ Y_4(0) \end{bmatrix}$$

beschrieben wird.

Betrachtet man die Koeffizienten von A in den Nebendiagonalblöcken, so kann man feststellen, daß sie im Vergleich zu den Koeffizienten in den Hauptdiagonalblöcken relativ kleine numerische Werte besitzen mit der Folge, daß offenbar zwischen den Variablengruppen Y_1 und Y_2 sowie Y_3 und Y_4 nur geringfügige Einwirkungen zum Tragen kommen.

Es liegt die Frage nahe, ob die Subsysteme Y_1 und Y_2 sowie Y_3 und Y_4 nicht isoliert voneinander untersucht werden können. Isoliert heißt in diesem Fall, daß man die in den Nebendiagonalblöcken von A befindlichen Koeffizienten Null setzt, was durch die Beziehung

$$Y(t) = A^* Y(t-1) \quad (25.6)$$

mit $A^* = \begin{bmatrix} 0,2 & 0,5 & 0 & 0 \\ 0,3 & 0,4 & 0 & 0 \\ 0 & 0 & 0,4 & 0,45 \\ 0 & 0 & 0,5 & 0,3 \end{bmatrix}$

zum Ausdruck kommt.

Die Beziehungen zwischen A^* und A können durch

$$A = A^* + \epsilon C \quad (25.7)$$

beschrieben werden, wobei C als Verbindungsmaatrix bezeichnet wird, weil sie die Verbindung zwischen den beiden Subsystemen beschreibt.

Die Konstante ϵ , die hier für eine später erfolgende Überlegung eingeführt wird, soll den Wert 1 besitzen.

Sind die numerischen Werte der Elemente der Verbindungsmaatrix (wie im vorliegenden Fall) relativ niedrig im Vergleich zu den numerischen Werten der Elemente in den Hauptdiagonalblöcken von A , so kann man die Matrix A als eine annähernd zerlegbare Matrix bezeichnen und analog dazu das eine solche Matrix enthaltende Modell als ein annähernd zerlegbares Modell.

Ersetzt man in (25.5) die annähernd zerlegbare Matrix A durch die zerlegbare Matrix A^* und definiert

$$A^* = \begin{bmatrix} A_1^* & 0 \\ 0 & A_2^* \end{bmatrix}$$

so können beide Subsysteme A_1^* und A_2^* isoliert untersucht werden, und zwar anhand der Beziehungen

$$\begin{bmatrix} Y_1(t) \\ Y_2(t) \end{bmatrix} = A_1^* \begin{bmatrix} Y_1(t-1) \\ Y_2(t-1) \end{bmatrix}$$

und

$$\begin{bmatrix} Y_3(t) \\ Y_4(t) \end{bmatrix} = A_2^* \begin{bmatrix} Y_3(t-1) \\ Y_4(t-1) \end{bmatrix} \quad (25.8)$$

Im folgenden wollen wir der Frage nachgehen, in welchem Umfang es berechtigt ist, eine derartige isolierte Subsystemanalyse durchzuführen. Wir betrachten dazu nur das Beispiel für A_2^* .

Definiert man die Nebendiagonalblöcke der Matrix A in allgemeiner Form, d.h. durch

$$A = \begin{bmatrix} 0,2 & 0,5 & a_{13} & a_{14} \\ 0,3 & 0,4 & a_{23} & a_{24} \\ \hline a_{31} & a_{32} & 0,4 & 0,45 \\ a_{41} & a_{42} & 0,5 & 0,3 \end{bmatrix}$$

so läßt sich aus dem Ansatz (25.5) auf analytischem Wege die Beziehung

$$\begin{bmatrix} Y_3(t) \\ Y_4(t) \end{bmatrix} = A_2^{(1)} \begin{bmatrix} Y_3(t-1) \\ Y_4(t-1) \end{bmatrix} + A_2^{(2)} \begin{bmatrix} Y_3(t-2) \\ Y_4(t-2) \end{bmatrix} + A_2^{(3)} \begin{bmatrix} Y_3(t-3) \\ Y_4(t-3) \end{bmatrix} \quad (25.9)$$

ableiten. In dieser Darstellungsform sind die Variablen Y_3 und Y_4 des Subsystems allein in ihrer Abhängigkeit von den verzögerten Variablen des eigenen Subsystems beschrieben. Dennoch erfaßt diese Darstellung die Y_3 und Y_4 berührenden Interdependenzen zwischen den beiden Subsystemen. Die Beziehung (25.9) beschreibt daher den tatsächlichen Verlauf der Variablen Y_3 und Y_4 in Abhängigkeit von ihren verzögerten Ausprägungen, während (25.8) die durch die Annahme einer isolierten Subsystemanalyse vereinfachte Form darstellt. Die Frage nach der Zulässigkeit einer isolierten Subsystemanalyse kann daher auf die Frage reduziert werden, ob im Hinblick auf den Zeitverlauf von Y_3 und Y_4 (25.8) als eine hinreichende Approximation der Beziehung (25.9) akzeptiert werden kann.

Um dieser Frage nachzugehen, soll (25.9) stärker konkretisiert werden. Durch geeignete Umformungen erhält man:

$$Y_3(t) = [p+k]Y_3(t-1) + [1-kp]Y_3(t-2) + [(a_{23}a_{41} - 0,5 \cdot 0,3) \frac{j}{a_{41}} - 1 \cdot p]Y_3(t-3) + mY_4(t-1) + [\frac{0,3j}{a_{41}} + n - mp]Y_4(t-2) + [(a_{41}a_{24} - 0,3 \cdot 0,3) \frac{j}{a_{41}} - np]Y_4(t-3)$$

und

$$\begin{aligned} Y_4(t) &= bY_3(t-1) + \left[\frac{0,3a}{a_{31}} + c - fb \right] Y_3(t-2) + \left[(a_{31}a_{32} - 0,3 \cdot 0,4) \frac{a}{a_{31}} - fc \right] Y_3(t-3) + \\ &\quad + [d+f] Y_4(t-1) + [e-fd] Y_4(t-2) + \left[(a_{24}a_{31} - 0,3 \cdot 0,45) \frac{a}{a_{31}} - fe \right] Y_4(t-3) \end{aligned}$$

mit den folgenden Abkürzungen:

$$\begin{aligned} a &= 0,5a_{41} + 0,4a_{42} - a_{32}g & b &= g + 0,5 \\ c &= a_{13}a_{31} + a_{42}a_{23} - 0,4g & d &= 0,3 \\ e &= a_{14}a_{41} + a_{42}a_{24} - 0,45g & f &= \frac{0,4a_{31} - 0,3a_{32}}{a_{31}} \\ g &= \frac{0,2a_{41} + 0,3a_{42}}{a_{31}} & i &= \frac{0,2a_{31} + 0,3a_{32}}{a_{41}} \\ j &= 0,5a_{31} + 0,4a_{32} - ia_{42} & k &= 0,4 \\ l &= a_{31}a_{13} - a_{32}a_{23} - 0,5i & m &= i + 0,45 \\ n &= a_{14}a_{31} + a_{32}a_{24} - 0,3i & p &= \frac{0,4a_{41} - 0,3a_{42}}{a_{41}} \end{aligned}$$

Für die Matrizen $A_2^{(1)}$, $A_2^{(2)}$ und $A_2^{(3)}$ ergibt sich:

$$A_2^{(1)} = \begin{bmatrix} (p+k) & m \\ b & (d+f) \end{bmatrix}$$

$$A_2^{(2)} = \begin{bmatrix} (1-kp) & \left(\frac{0,3j}{a_{41}} + \psi - mp \right) \\ \left(\frac{0,3a}{a_{31}} + c - fb \right) & (e-fd) \end{bmatrix}$$

$$A_2^{(3)} = \begin{bmatrix} \left[(a_{23}a_{41} - 0,15) \frac{j}{a_{41}} - lp \right] & \left[(a_{41}a_{24} - 0,09) \frac{j}{a_{41}} - np \right] \\ \left[(a_{31}a_{32} - 0,12) \frac{a}{a_{31}} - fc \right] & \left[(a_{24}a_{31} - 0,135) \frac{a}{a_{31}} - fe \right] \end{bmatrix}$$

Aus (25.6) und (25.7) erkennt man, daß zwischen den Elementen der Nebendiagonalblöcke von A und C die Beziehung

$$a_{ij} = \varepsilon c_{ij}$$

gilt.

Ersetzt man nun in den Matrizen $A_2^{(1)}$, $A_2^{(2)}$, $A_2^{(3)}$ die a_{ij} durch ϵc_{ij} und führt den Grenzübergang $\epsilon \rightarrow 0$ mit $A \rightarrow A^*$ durch, so folgt $\epsilon c_{ij} \rightarrow 0$.

Daraus folgt:

$$b \rightarrow 0,5, \quad d \rightarrow 0,3, \quad k \rightarrow 0,4, \quad m \rightarrow 0,45$$

Die übrigen Ausdrücke a, c, e, f, i, l, j, n, p, g verschwinden. Für die Matrizen $A_2^{(1)}$, $A_2^{(2)}$ und $A_2^{(3)}$ ergeben sich die Übergänge

$$\begin{aligned} A_2^{(2)} &\rightarrow \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} & A_2^{(3)} &\rightarrow \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \\ A_2^{(1)} &\rightarrow \begin{bmatrix} k & m \\ b & d \end{bmatrix} = \begin{bmatrix} 0,4 & 0,45 \\ 0,5 & 0,3 \end{bmatrix} = A_2^* \end{aligned}$$

d.h. es liegt das folgende isolierte Submodell vor:

$$\begin{bmatrix} Y_3(t) \\ Y_4(t) \end{bmatrix} = \begin{bmatrix} 0,4 & 0,45 \\ 0,5 & 0,3 \end{bmatrix} \begin{bmatrix} Y_3(t-1) \\ Y_4(t-1) \end{bmatrix}$$

Man kann daher feststellen, daß die Beziehung (25.9) in (25.8) übergeht, wenn die Elemente der Verbindungsmatrix gegen Null streben.

In welchem Maß jedoch diese Vereinfachung (25.8) eine akzeptable Approximation des 'wahren' Modells (25.9) darstellt, hängt von den numerischen Werten der Elemente der Verbindungsmatrix C ab.

SIMON und ANDO haben das hier beispielhaft dargestellte Problem in systematischer Weise untersucht [179]. Unter Verallgemeinerung der bisher gebrauchten Matrzenschreibweise soll A^* eine blockdiagonale Ma-

$$A^* = \begin{bmatrix} A_1^* & & & \\ & A_2^* & & 0 \\ & & \ddots & \\ 0 & & & A_i^* \\ & & & \ddots \\ & & & A_n^* \end{bmatrix}$$

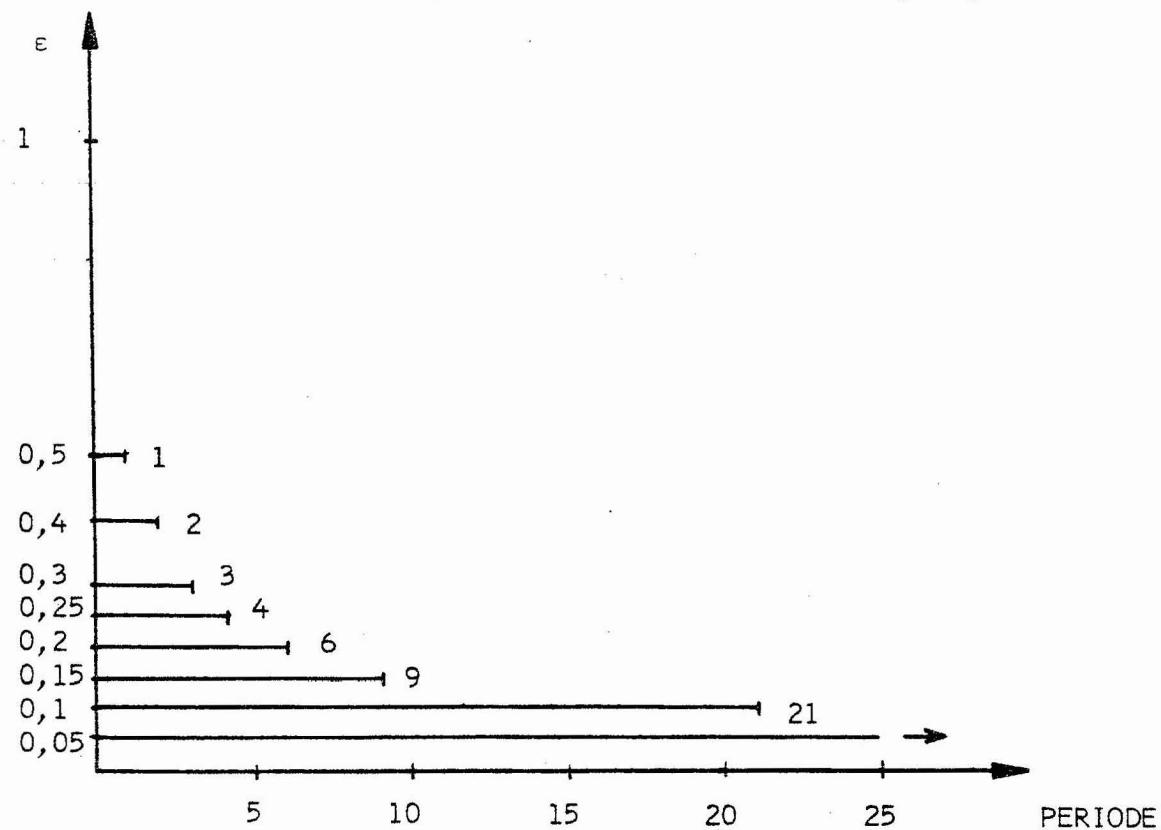
trix sein, die ein vereinfachendes A^* -Modell beschreibt, welches aus

einem A-Modell mit einer annähernd zerlegbaren Matrix A gewonnen wurde. Die Verknüpfung zwischen beiden Matrizen kann in Übereinstimmung mit dem angeführten Beispiel durch (25.7) beschrieben werden.

SIMON und ANDO haben nunmehr nachgewiesen, daß sich unter Zugrundelegung der Beziehung (25.7) bei hinreichend kleinem ϵ stets ein Zeitraum von der Periode 0 bis zur Periode T angeben läßt, in dem der Zeitverlauf der Variablen des vereinfachten A^* -Modells annähernd mit dem Zeitverlauf übereinstimmt, den die entsprechenden Variablen des A-Modells beschreiben. Diese Feststellung läßt sich anhand des erörterten Beispiels illustrieren.

Als tolerierbare Abweichung des Unterschiedes zwischen den Variablenwerten des A_2 und A_2^* -Modells wurde zehn Prozent angenommen. Für den Spaltenvektor der Anfangswerte $Y(0)=[Y_1(0), Y_2(0), Y_3(0), Y_4(0)]$ wurden die Werte $Y(0)=[100, 200, 300, 400]$ gewählt.

Abbildung 25.4 zeigt den Zusammenhang zwischen verschiedenen Werten von ϵ und dem Zeitbereich, in welchem die als zulässig angesehene Ab-



25.4 Abhängigkeit zwischen dem Koeffizientenwert ϵ der Verbindungs-matrix ϵC und dem Zeitraum, in welchem die Verlaufsabweichung zwi-schen den Variablen des A_2 - und A_2^* -Modells unter zehn Prozent liegt.

weichung von nicht mehr als zehn Prozent vorlag.

Das Theorem von SIMON und ANDO erschüttert die gelegentlich vorgetragene Aufassung, isolierte Subsystemanalysen seien im Sinne einer ganzheitlichen Systemerfassung generell zu verwerfen. Seine praktische Anwendbarkeit ist jedoch begrenzt, denn es kann nicht als Kriterium fungieren, ob in einem gegebenen Fall eine isolierte Subsystemanalyse angemessen ist oder nicht. Die Entscheidung über den Erfolg einer derartigen Untersuchung hängt von der Struktur der Verbindungsmaatrix ϵC ab, die im Einzelfall aber nicht bekannt ist. Selbst wenn ϵC bekannt wäre, würde das Theorem keinen Hinweis liefern, ob eine bestimmte Subsystemanalyse akzeptabel ist.

SIMON und ANDOs Theorem kann daher nur als Ermunterung zur Entwicklung von isolierten Submodellen gedeutet werden: es ist eine Chance gegeben, akzeptable Teilbereichsmodelle zu entwickeln; ob dies im Einzelfall auch tatsächlich gelungen ist, zeigt allerdings erst die empirische Überprüfung des entwickelten Ansatzes.

2.6. Deterministische und stochastische Modellformen

Der fundamentale Unterschied zwischen deterministischen und stochastischen Modellen erklärt sich aus den grundsätzlich unvereinbaren 'Weltsichten', die diesen Modellen zugrunde liegen. Als Folge davon besitzen auch die Implikationen beider Modellformen einen grundlegend andersartigen Charakter: den Implikationen deterministischer Modelle in Form sicherer Behauptungen stehen die Wahrscheinlichkeitsbehauptungen stochastischer Modellimplikationen gegenüber. Der Unterschied zwischen beiden Modellformen wirkt sich bis auf das anzuwendende mathematische Instrumentarium zur Aufdeckung dieser Implikationen aus: über die Beachtung der Vorschriften der Algebra hinausgehend, verlangt die Implikationenaufdeckung stochastischer Modelle die Einhaltung bestimmter statistischer Grundgesetze, die in den sogenannten Kolmogoroffschen Axiomen ihren Niederschlag gefunden haben.

Weil im Wissenschaftsbetrieb die Aufdeckung von Modellimplikationen

fast ausschließlich anhand von deterministischen Modellen erfolgt, wurden die bisherigen Betrachtungen stillschweigend auf die Ermittlung deterministischer Modellimplikationen eingeschränkt, obwohl im Rahmen der diskutierten Modelltypen auch stochastische Varianten und daher auch die Aufdeckung stochastischer Implikationen denkbar wäre.

In diesem Abschnitt sollen die deduktiven und pseudoinduktiven Verfahren der Implikationenaufdeckung stochastischer Modelle dargestellt werden.

Deterministische Modelle werden nur kurz und mehr aus systematischen Gründen angesprochen, denn die Behandlung ihrer Formen und die Verfahren ihrer Implikationenerschließung wurde in den übrigen Teilen dieses Kapitels ausführlich beschrieben.

Als erstes wird anhand von zwei Beispielen die Ermittlung des Zeitverlaufes des Erwartungswertes und der Standardabweichung der endogenen Variablen eines linearen stochastischen Störgrößenmodells demonstriert. Dem schließt sich eine Erörterung über die Bestimmung derselben Implikationen im Rahmen einer pseudoinduktiven Modellschließung an. Neben einer kurzen Bemerkung zum praktischen Wert sogenannter varianzreduzierender Verfahren im Rahmen einer pseudoinduktiven Varianzermittlung werden abschließend die Beziehungen zwischen subjektiven Entscheidermodellen und stochastischen Modellen dargestellt.

2.6.1. Deterministische Modellformen

Deterministische Modelle repräsentieren sichere Wenn-Dann-Aussagen über die Realität in Form empirisch interpretierter Differenzengleichungen und algebraischer Gleichungen. Sind in einem deterministischen Modell die vorherbestimmten Variablen numerisch spezifiziert, d.h. die Wenn-Komponenten durch singuläre Tatsachenbehauptungen belegt, dann liefern die Dann-Komponenten in Form der endogenen Variablenverläufe dem Modellanwender zumeist unbekannte 'neue' Tatsachenbehauptungen. Diese 'neuen' Tatsachenbehauptungen, die vom zeitlichen Stand-

punkt des Modellanwenders gesehen meistens Prognosen sind, stellen Implikationen des deterministischen Modells dar und sind mit Hilfe bestimmter deduktiver und pseudoinduktiver Methoden zu ermitteln. Diese Methoden wurden in den verschiedenen Abschnitten des zweiten Kapitels erörtert, so daß sie dem Leser bereits bekannt sein dürften.

2.6.2. Stochastische Modellformen

A. Begriffliche Klärung und empirische Interpretation

Stochastische dynamische MZÄ-Modelle werden durch stochastische Differenzengleichungen beschrieben, die in ihrer äußeren Darstellungsform große Ähnlichkeit mit deterministischen Differenzengleichungen besitzen.

Der stochastische Charakter der in die Gleichungen eingehenden stochastischen Variablen kommt in der formalen Darstellung allein dadurch zum Ausdruck, daß die Wahrscheinlichkeitsverteilungen der stochastischen exogenen Variablen zusätzlich anzugeben sind. Als Beispiel sei die stochastische Investitionshypothese angeführt¹

$$I_i(t) = \beta[C(t) - C(t-1)] + \varepsilon(t) \quad \varepsilon(t) \in NV(\mu=0, \sigma=\text{konst.})$$

$\varepsilon(t)$ ist hierbei eine stochastische Variable, die einer Normalverteilung mit dem Erwartungswert $\mu=0$ und der Standardabweichung $\sigma=\text{konstant}$ angehört. Man erkennt, daß die endogene Variable I_i ebenfalls eine stochastische Variable ist. Die endogenen Variablen eines stochastischen dynamischen Modells werden daher durch ihre Wahrscheinlichkeitsverteilung charakterisiert.

Wie bei den deterministischen Modellen die wichtigste Implikation die Ermittlung des Zeitverlaufes der endogenen Variablen ist, so ist bei stochastischen Modellen die Bestimmung des zeitlichen Verlaufes der Wahrscheinlichkeitsverteilung ihrer endogenen Variablen die zentrale Aufgabe der Implikationenaufdeckung.

In den meisten Fällen begnügt man sich mit der Bestimmung des zeitlichen Verlaufes des Erwartungswertes und der Standardabweichung der

¹ Vgl. Seite 74

Wahrscheinlichkeitsverteilung der endogenen Variablen, zumal diese Parameter einige Verteilungstypen erschöpfend charakterisieren. Der Erwartungswert spielt bei der Analyse stochastischer Modelle eine entscheidende Rolle. Betrachtungen über Sensitivitäten, Multiplikatoren und dynamische Verhaltensweisen werden oft unter Verwendung des Erwartungswertes als endogener Variable analog zu deterministischen Modellen vorgenommen.

Es läge nahe, von der Annahme auszugehen, daß die Werte der Verläufe der endogenen Variablen eines deterministischen Modells mit den Verläufen der Erwartungswerte des entsprechenden stochastischen Modells übereinstimmen. Träfe diese Annahme zu, so könnte man stets die Untersuchung der Erwartungswerte eines stochastischen Modells an einem wesentlich einfacher zu handhabenden deterministischen Äquivalent vornehmen. Wie HOWERY und KELEJIAN jedoch gezeigt haben, führt die Determinisierung nichtlinearer Modelle nicht zwingend zu Verläufen der endogenen Variablen, die dem Erwartungswert entsprechen [94]. Die generelle Unterstellung, die Verläufe der endogenen Variablen eines nichtlinearen determinisierten Modells könnten als die Erwartungswerte des stochastischen Modells angesehen werden, ist deshalb falsch.

Der Übergang zu einem determinisierten Modell stellt zudem auch eine Informationsbeschränkung dar, weil prinzipiell ermittelbare Informationen über andere Parameter der Wahrscheinlichkeitsverteilung wie zum Beispiel über den Verlauf der Standardabweichung auf jeden Fall verloren gehen.

B. Deduktive Analyse stochastischer Modelle

Ein deterministisches lineares Differenzengleichungssystem wird dadurch stochastisiert, daß man von der Unterstellung ausgeht, zumindest einer der Koeffizienten oder eine exogene Variable sei nur durch eine Wahrscheinlichkeitsverteilung beschreibbar. Im Rahmen dieses Modelltyps verwendet man, wie erwähnt, in den Wirtschafts- und Sozialwissenschaften fast ausschließlich sogenannte Störgrößenmodelle.²

Es wird dabei unterstellt, daß die sogenannte Störgröße oder Schock-

² Vgl. Seite 40

variable durch eine Normalverteilung mit dem Mittelwert Null und einer unveränderten Standardabweichung beschrieben werden kann.

Als Beispiel wurde die Hypothese der induzierten Investitionen I_i des MA-Modells

$$I_i(t) = \beta[C(t) - C(t-1)] + \varepsilon(t)$$

angeführt, in der $\varepsilon(t)$ die Störgröße repräsentiert. Man muß sich darüber klar sein, daß in Fällen, in denen (wie im MA-Modell) alle endogenen Variablen eine Totalschleife bilden, durch ε auch alle endogenen Variablen des Systems stochastisch verseucht werden, d.h. generell nur Informationen über ihre Wahrscheinlichkeitsverteilungen zur Verfügung stehen.

Nehmen wir vereinfachend an, daß die Konsumfunktion $C(t)$ deterministischer Natur ist, d.h.

$$C(t) = \alpha Y(t-1)$$

so gelangt man unter Berücksichtigung der Definitionsgleichung

$$Y(t) = C(t) + I_i(t) + I_a(t)$$

und unter analoger Anwendung der Endgleichungsbestimmung im deterministischen Fall³ zu der stochastischen Endgleichung von Y

$$Y(t) = (\alpha + \alpha\beta)Y(t-1) - \alpha\beta Y(t-2) + I_a(t) + \varepsilon(t) \quad (26.1)$$

Unterstellen wir jedoch eine ebenfalls stochastische Konsumfunktion

$$C(t) = \alpha Y(t-1) + \mu(t) \quad (26.2)$$

so erhalten wir, wie man sich leicht überzeugen kann, die stochastische Endgleichung

$$Y(t) = (\alpha + \alpha\beta)Y(t-1) - \alpha\beta Y(t-2) + I_a(t) + \eta(t) \quad (26.3)$$

mit $\eta(t) = (1+\beta)\mu(t) - \beta\mu(t-1) + \varepsilon(t)$

Generell läßt sich in Analogie zu den Endgleichungsbestimmungen deterministischer Systeme feststellen, daß es im Falle eines linearen Störgrößenmodells mit konstanten Koeffizienten immer dann möglich ist, eine stochastische Endgleichung zu ermitteln, wenn die Determinante der Operatormatrix des entsprechenden determinisierten Modells

³ Vgl. Seite 36f.

ungleich Null ist.⁴

In den stochastischen Endgleichungen eines linearen Systems treten dabei die Störgrößen $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ der Hypothesengleichungen als Linear kombinationen ihrer verzögerten und unverzögerten Werte auf.

Die Methoden zur Analyse stochastischer linearer Differenzengleichungen der beschriebenen Art gehen auf Arbeiten von YULE, WOLD und SLUTSKY zurück. [222], [224], [182]

Sie sollen im folgenden kurz erörtert werden. Gehen wir davon aus, daß es gelungen ist, durch entsprechende Umformungen eine stochastische Endgleichung der allgemeinen Form

$$Y(t) = \omega_1 Y(t-1) + \omega_2 Y(t-2) + \dots + \omega_n Y(t-n) + \varepsilon(t) + E(t) \quad (26.4)$$

aufzustellen. Nach dem Postulat der ungestörten Überlagerung setzt sich der Zeitverlauf von $Y(t)$ aus zwei Komponenten zusammen: der deterministischen Komponente $D(t)$, welche durch den Eingang $E(t)$ und die Anfangswerte bewirkt wird und der stochastischen Komponente $S(t)$, welche auf $\varepsilon(t)$ zurückzuführen ist, d.h. es gilt:

$$Y(t) = D(t) + S(t)$$

Die stochastische Komponente $S(t)$ kann in ihr sequentielles Äquivalent umgeformt werden, d.h. in die Form⁵

$$S(t) = \sum_{\tau=0}^{\infty} g(\tau) \varepsilon(t-\tau) \quad (26.5)$$

Hierbei stellt $g(\tau)$ die Gewichtsfunktion der Endgleichung dar.

Die Gewinnung einer Aussage über die Dichtefunktion von $Y(t)$ ist grundsätzlich immer dann möglich, wenn es anhand der statistischen Eigenschaften von ε gelingt, die Dichtefunktion von $S(t)$ zu ermitteln.

Um die Abhängigkeit zwischen der stochastischen Variablen ε und $S(t)$ in (26.5) aufzuweisen, müssen wir auf einige bekannte Sätze der Wahrscheinlichkeitstheorie zurückgreifen.⁶

S₁ Der Erwartungswert einer Summe von v unabhängigen Zufallsvariablen, deren Glieder mit einem konstanten Vielfachen multipliziert sind, berechnet sich nach:

⁴ Siehe im einzelnen Seite 245

⁵ Vgl. Seite 295f.

⁶ Zum Beweis dieser Sätze siehe [216, S. 74, 97 und 141]

$$E\{b_1 \varepsilon(1) + b_2 \varepsilon(2) + \dots + b_v \varepsilon(v)\} = b_1 E\{\varepsilon(1)\} + b_2 E\{\varepsilon(2)\} + \dots + b_v E\{\varepsilon(v)\}$$

S₂ Die Varianz einer Zufallsvariablen, die aus der Summe von v unabhängigen Zufallsvariablen gebildet wird, entspricht der Summe der Varianzen der sie bildenden Zufallsvariablen, d.h.

$$\text{var} \sum_{i=1}^v \varepsilon(i) = \sum_{i=1}^v \text{var } \varepsilon(i)$$

S₃ Die Varianz eines konstanten Vielfachen einer Zufallsvariablen entspricht der mit dem Quadrat dieser Konstanten multiplizierten Varianz der Zufallsvariablen, d.h.

$$\text{var}(a\varepsilon) = a^2 \text{var } \varepsilon$$

S₄ Wenn die unabhängigen Zufallsvariablen $\varepsilon(1), \varepsilon(2), \dots, \varepsilon(v)$ einer Normalverteilung entstammen, dann ist auch die lineare Kombination dieser Zufallsvariablen normalverteilt.

Nach Satz S₁ ist aufgrund der Beziehung (26.5) auch der Erwartungswert von S(t) Null, da voraussetzungsgemäß der Erwartungswert von ε Null sein soll.

Nach der Bestimmung des Erwartungswertes wenden wir uns der Ermittlung der Varianz von S(t) zu. Da die Störgrößen $\varepsilon(t-\tau)$ unabhängig voneinander sein sollen, kann man den Satz S₂ auf die Beziehung (26.5) anwenden, d.h.

$$\text{var } S(t) = \sum_{\tau=0}^{\infty} \text{var } g(\tau) \varepsilon(t-\tau)$$

Da die Varianz von $\varepsilon(t-\tau)$ für $\tau=0, 1, 2, \dots$ konstant sein soll, d.h.

$\text{var } \varepsilon(t-\tau) = \text{var } \varepsilon$, gilt gemäß Satz S₃

$$\text{var } S(t) = \text{var } \varepsilon \sum_{\tau=0}^{\infty} g^2(\tau) \quad (26.6)$$

Die Varianz von S(t) bestimmt sich damit aus dem Produkt der Varianz von ε und einer unendlichen Summe. Die Beurteilung, unter welchen Bedingungen $\text{var } S(t)$ einen endlichen Wert besitzt und damit eine stochastische Stabilität vorliegt, läuft damit auf die Frage hinaus, unter welchen Bedingungen die unendliche Summe

$$\sum_{\tau=0}^{\infty} g^2(\tau) \quad (26.7)$$

einen endlichen Grenzwert besitzt. Die Gewichtsfunktion $g(\tau)$ eines linearen Systems läßt sich, wie wir gesehen haben, durch die Funktionslösung⁷

$$g(\tau) = b_1 \lambda_1^\tau + b_2 \lambda_2^\tau + \dots + b_n \lambda_n^\tau$$

beschreiben. Der Ausdruck $g^2(\tau)$ wird durch

$$g^2(\tau) = \sum_{i=1}^n \sum_{j=1}^n b_i b_j (\lambda_i \lambda_j)^\tau \quad (26.8)$$

dargestellt. Falls sämtliche Wurzeln der Funktionslösung $|\lambda_i| < 1$, dann hat die unendliche Summe

$$\sum_{\tau=0}^{\infty} g^2(\tau) = \sum_{\tau=0}^{\infty} \sum_{i=1}^n \sum_{j=1}^n b_i b_j (\lambda_i \lambda_j)^\tau \quad (26.9)$$

einen endlichen Grenzwert. Denn für jedes Glied $b_i b_j (\lambda_i \lambda_j)^\tau$ in (26.9) gilt in diesem Fall $|\lambda_i \lambda_j| < 1$. Eine geometrische Reihe besitzt bekanntermaßen den Summenwert

$$a^0 + a^1 + a^2 + a^3 + \dots + a^\tau = \frac{1-a^{\tau+1}}{1-a} \quad (26.10)$$

Daher berechnet sich die Summe eines Gliedes $b_i b_j (\lambda_i \lambda_j)^\tau$ aus (26.9) mit $a = \lambda_i \lambda_j$

$$b_i b_j [(\lambda_i \lambda_j)^0 + (\lambda_i \lambda_j)^1 + \dots + (\lambda_i \lambda_j)^\tau] = \frac{1 - (\lambda_i \lambda_j)^{\tau+1}}{1 - \lambda_i \lambda_j} b_i b_j$$

für $\tau \rightarrow \infty$ führt dies für $|a| < 1$ zu dem Grenzwert

$$\sum_{\tau=0}^{\infty} b_i b_j (\lambda_i \lambda_j)^\tau = \frac{1}{1 - \lambda_i \lambda_j} b_i b_j$$

Da diese Beziehung für jeden Ausdruck in (26.9) gilt, hat auch (26.9) einen endlichen Grenzwert. Die Forderung, daß im Falle stochastischer Stabilität die Absolutwerte sämtlicher Wurzeln kleiner als Eins sein müssen, ist nicht nur hinreichend, sondern auch notwendig. Denn findet sich nur eine Wurzel λ_s mit $|\lambda_s| > 1$, dann besitzt der Ausdruck $b_s b_s (\lambda_s \lambda_s)^\tau$ keinen Grenzwert, womit auch die unendliche Summe (26.9) keinen Grenzwert besitzt. Die Folge ist, daß ein lineares Störgrößenmodell, dessen deterministisches Äquivalent ein stabiles Verhal-

⁷ Vgl. Seite 204

ten aufweist, stets dann auch stochastische Stabilität besitzt.

Um die stochastische Stabilität eines derartigen Modells zu beurteilen, kann man sich daher auf die Analyse seines deterministischen Äquivalentes beschränken. Will man allerdings neben der Beurteilung der stochastischen Stabilität auch die Höhe der Varianz ermitteln, so ist wegen (26.6) die Summe (26.7) zu berechnen. Ist ε normalverteilt, so ist $S(t)$ wegen des Satzes S₄ ebenfalls normalverteilt. Da eine Normalverteilung durch ihren Erwartungswert und ihre Standardabweichung vollständig beschrieben werden kann, ist die Wahrscheinlichkeitsverteilung von $S(t)$ damit eindeutig gekennzeichnet.

Dieser Fall soll anhand von zwei Beispielen demonstriert werden. Wir wenden uns dabei nicht direkt einer stochastisierten Version des uns schon bekannten MA-Modells zu, sondern werden uns zuvor mit dem noch einfacheren Fall eines Modells beschäftigen, welches durch eine stochastische Differenzengleichung ersten Grades beschrieben wird.

a) Analyse eines stochastischen Modells der Lager- und Bestellpolitik

Im folgenden Modell wird die Lager- und Bestellpolitik eines Betriebes beschrieben. Der stochastische Charakter des Modells resultiert aus der Tatsache, daß es eine exogene Variable, die Absatzmenge, enthält, von welcher allein die Dichtefunktion bekannt ist. Es werden folgende Größen verwendet:⁸

- $L(t)$ Lagerbestand am Ende der Periode t
- \bar{L} Angestrebter Lagerbestand
- $P(t)$ produzierte Menge in der Periode t
- \bar{P} Durchschnittliche produzierte Menge
- $R(t)$ Änderungsrate der produzierten Menge in der Periode t
- $A(t)$ Abweichung des wirklichen Lagerbestandes vom angestrebten Lagerbestand \bar{L}
- $N(t)$ Absatzmenge während der Periode t
- \bar{N} Durchschnittliche Absatzmenge
- $\varepsilon(t)$ Zufallsvariable, die die Abweichung der realisierten Nachfrage von der durchschnittlichen Nachfrage \bar{N} kennzeichnet. Annahme einer Normalverteilung mit dem Mittelwert 0 und der Varianz σ_N^2 .

⁸ Zur Entwicklung dieses Modells vgl. [45] und [138]

Die Produktionsmenge in der Periode t bestimmt sich aus der durchschnittlichen Produktionsmenge \bar{P} und einer Komponente $R(t)$, die die Abweichung zwischen dem angestrebten und dem realisierten Lagerbestand in der Vorperiode berücksichtigt.

$$P(t) = \bar{P} + R(t)$$

Diese Komponente wird als ein konstantes Vielfaches α der Abweichung zwischen angestrebtem und realisiertem Lagerbestand der Vorperiode $A(t-1)$ gewählt, d.h.

$$R(t) = \alpha A(t-1)$$

Die Abweichung $A(t)$ bestimmt sich aus der Differenz des realisierten Lagerbestandes $L(t)$ mit dem angestrebten Bestand \bar{L} , d.h.

$$A(t) = \bar{L} - L(t)$$

Unter Berücksichtigung der Definitionsgleichung des Lagerbestandes

$$L(t) = L(t-1) + P(t) - \varepsilon(t) - \bar{N}$$

folgt als Endgleichung für $P(t)$ und $A(t)$

$$P(t) = (1-\alpha)P(t-1) + \alpha\bar{N} + \alpha\varepsilon(t-1)$$

$$A(t) = (1-\alpha)A(t-1) - \bar{P} + \varepsilon(t) + \bar{N}$$

Aus den Endgleichungen für $P(t)$ und $A(t)$ erkennt man, daß $P(t)$ und $A(t)$ sich stets dann durch stochastische Stabilität auszeichnen, wenn $|1-\alpha| < 1$. Der Verlauf des Erwartungswertes der endogenen Variablen $P(t)$, d.h. $D(t)$, ergibt sich aus der Funktionslösung des entsprechenden deterministischen Äquivalentes und ist

$$D(t) = (1-\alpha)^t (P(0) - \bar{N}) + \bar{N}$$

Die Gewichtsfunktion $g(\tau)$ wird durch

$$g(\tau) = (1-\alpha)^\tau$$

beschrieben. Damit wird

$$g^2(\tau) = (1-\alpha)^{2\tau}$$

Mit der Summenformel (26.10) gilt

$$\sum_{\tau=0}^t g^2(\tau) = \frac{1-(1-\alpha)^2(t+1)}{1-(1-\alpha)^2}$$

und damit bestimmt sich nach (26.6) der zeitliche Verlauf der Varianz durch

$$\text{var } P(t) = \frac{1-(1-\alpha)^2(t+1)}{1-(1-\alpha)^2} \text{ var } (\alpha \varepsilon) \quad (26.11)$$

Der die stochastische Stabilität kennzeichnende Grenzwert der Varianz P_g ergibt sich, wenn man t in (26.11) gegen unendlich laufen lässt und wird

$$\text{var } P_g = \frac{1}{2\alpha-\alpha^2} \text{ var } (\alpha \varepsilon) = \alpha^2 \frac{1}{2\alpha-\alpha^2} \text{ var } \varepsilon = \frac{\alpha}{2-\alpha} \text{ var } \varepsilon$$

Da α ein deterministischer Parameter ist, der zudem eine Kontrollvariable darstellt, kann es von Interesse sein, welche Sensitivität die Varianzen der endogenen Variablen bezüglich der Änderung von α besitzen.

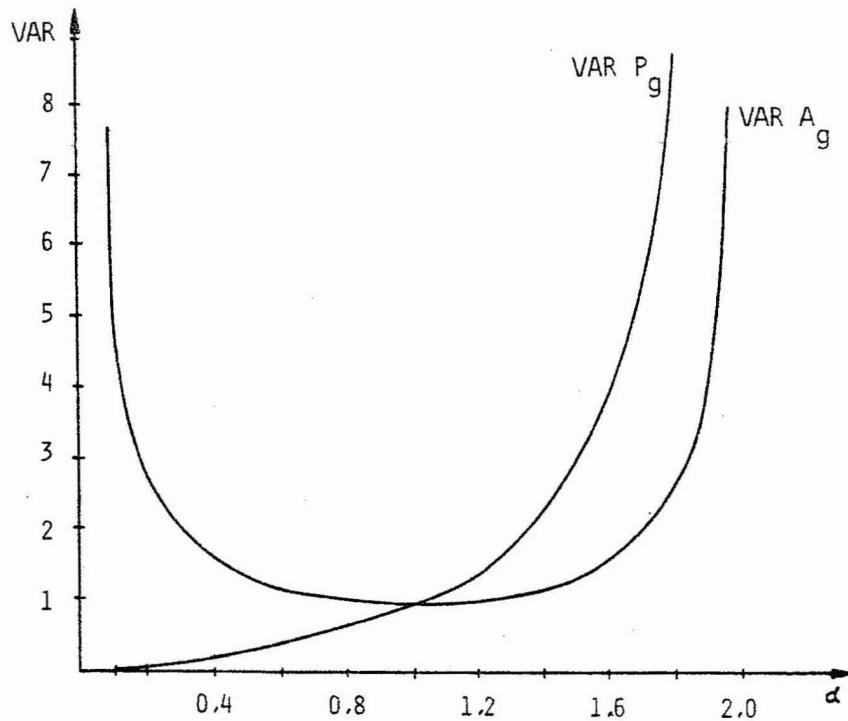


Abb. 26.1 Die Varianz der produzierten Menge $P(t)$ und der Lagerbestandsabweichung $A(t)$ in Abhängigkeit von dem Faktor α unter der Annahme $\text{var } \varepsilon=1$ und $t \rightarrow \infty$

Im Falle stochastischer Modelle erweitert sich damit der Bereich der Untersuchungsmöglichkeit, indem man die Empfindlichkeit der Varianzen bestimmter endogener Variablen im Hinblick auf bestimmte kritische Parameter untersuchen kann.

b) Analyse eines stochastischen MA-Modells

Zur Analyse eines linearen stochastischen MA-Modells gehen wir von der uns bereits bekannten stochastischen Endgleichung

$$Y(t) = bY(t-1) - cY(t-2) + I_a(t) + \varepsilon(t) \quad (26.12)$$

$$\text{mit } b = \alpha + \alpha\beta \quad (26.13)$$

$$\text{und } c = \alpha\beta$$

aus. Diese besitzt die Gewichtsfunktion

$$g(\tau) = b_1 \lambda_1^\tau + b_2 \lambda_2^\tau \quad (26.14.)$$

womit sich $g^2(\tau)$ durch

$$g^2(\tau) = b_1^2 \lambda_1^{2\tau} + b_2^2 \lambda_2^{2\tau} + 2b_1 b_2 (\lambda_1 \lambda_2)^\tau$$

bestimmt. Die Summen dieser drei Komponenten bestimmen sich für

$|\lambda_1, \lambda_2| < 1$ nach der Summenformel (26.10), d.h.

$$b_1^2 \sum_{\tau=0}^t \lambda_1^{2\tau} = \frac{(1-\lambda_1^{2(t+1)})b_1^2}{1-\lambda_1^2}$$

$$b_2^2 \sum_{\tau=0}^t \lambda_2^{2\tau} = \frac{(1-\lambda_2^{2(t+1)})b_2^2}{1-\lambda_2^2}$$

$$2b_1 b_2 \sum_{\tau=0}^t (\lambda_1 \lambda_2)^\tau = \frac{[1-(\lambda_1 \lambda_2)^{t+1}]2b_1 b_2}{1-\lambda_1 \lambda_2}$$

Mit (26.6) ermittelt sich der Zeitpfad der Varianz $\text{var } Y(t)$ aus

$$\text{var } Y(t) = \text{var } \varepsilon \left[\frac{(1-\lambda_1^{2(t+1)})b_1^2}{1-\lambda_1^2} + \frac{(1-\lambda_2^{2(t+1)})b_2^2}{1-\lambda_2^2} + \frac{(1-(\lambda_1 \lambda_2)^{t+1})2b_1 b_2}{1-\lambda_1 \lambda_2} \right] \quad (26.15)$$

Gemäß (21.75) sind die Koeffizienten der Gewichtsfunktion (26.14) bei ungleichen Wurzeln

$$b_1 = -\lambda_1 / (\lambda_2 - \lambda_1) \quad b_2 = \lambda_2 / (\lambda_2 - \lambda_1) \quad (26.16)$$

Mit

$$\lambda_1 = \frac{b}{2} + \sqrt{\frac{b^2}{4} - c} \quad \text{und} \quad \lambda_2 = \frac{b}{2} - \sqrt{\frac{b^2}{4} - c}$$

folgt

$$b_1 = \frac{\left(\frac{b}{2} - \sqrt{\frac{b^2}{4} - c}\right)}{2\sqrt{\frac{b^2}{4} - c}}$$

und

(26.17)

$$b_2 = \frac{-\frac{b}{2} + \sqrt{\frac{b^2}{4} - c}}{2\sqrt{\frac{b^2}{4} - c}}$$

Setzt man die Parameter b_1 und b_2 in Gleichung (26.15) ein, so erhält man

$$\begin{aligned} \text{var } Y(t) &= \text{var } \varepsilon \left[\frac{\left[1 - \left(\frac{b}{2} + \sqrt{\frac{b^2}{4} - c}\right)^2 (t+1)\right] \left[b - \left(\frac{b}{2} - \sqrt{\frac{b^2}{4} - c}\right)\right]^2}{4\left(1 - \frac{b^2}{2} + c - 2\frac{b}{2}\sqrt{\frac{b^2}{4} - c}\right) \left(\frac{b^2}{4} - c\right)} + \frac{2(1-c^{t+1})(-c)}{4(1-c)(\frac{b^2}{4} - c)} \right. \\ &\quad \left. + \frac{\left[1 - \left(\frac{b}{2} - \sqrt{\frac{b^2}{4} - c}\right)^2 (t+1)\right] \left[\left(\frac{b}{2} + \sqrt{\frac{b^2}{4} - c}\right) - b\right]^2}{4\left(1 - \frac{b^2}{2} + c + 2\frac{b}{2}\sqrt{\frac{b^2}{4} - c}\right) \left(\frac{b^2}{4} - c\right)} \right] \end{aligned} \quad (26.18)$$

Mit der Reduzierung der Koeffizienten b und c auf die strukturellen Parameter α und β des MA-Modells gemäß (26.13) ergibt sich schließlich der gewünschte Ausdruck, der die zeitliche Entwicklung der Varianz von $Y(t)$ in Abhängigkeit von den Ausprägungen des Multiplikators α und des Akzelerators β beschreibt, d.h.

$$\text{var } Y(t) = \text{var } \varepsilon \left[\frac{\left[1 - \left(\frac{\alpha + \alpha\beta}{2} + \sqrt{\frac{(\alpha + \alpha\beta)^2}{4} - \alpha\beta} \right)^2(t+1) \right] \left[\frac{\alpha + \alpha\beta}{2} + \sqrt{\frac{(\alpha + \alpha\beta)^2}{4} - \alpha\beta} \right]^2}{\left[1 - \left(\frac{\alpha + \alpha\beta}{2} + \sqrt{\frac{(\alpha + \alpha\beta)^2}{4} - \alpha\beta} \right)^2 \right] \left[(\alpha + \alpha\beta)^2 - 4\alpha\beta \right]} - \frac{2[1 - (\alpha\beta)^{t+1}] \alpha\beta}{[1 - \alpha\beta][(\alpha + \alpha\beta)^2 - 4\alpha\beta]} + \right. \\ \left. + \frac{\left[1 - \left(\frac{\alpha + \alpha\beta}{2} - \sqrt{\frac{(\alpha + \alpha\beta)^2}{4} - \alpha\beta} \right)^2(t+1) \right] \left[-\frac{\alpha + \alpha\beta}{2} + \sqrt{\frac{(\alpha + \alpha\beta)^2}{4} - \alpha\beta} \right]^2}{\left[1 - \left(\frac{\alpha + \alpha\beta}{2} - \sqrt{\frac{(\alpha + \alpha\beta)^2}{4} - \alpha\beta} \right)^2 \right] \left[(\alpha + \alpha\beta)^2 - 4\alpha\beta \right]} \right] \quad (26.19)$$

Dieser voluminöse Formelausdruck zeigt, welche engen Grenzen einer deduktiven Untersuchung stochastisch linearer Systeme gesetzt sind. Schon in etwas weniger einfachen linearen Systemen wird eine Analyse dieser Art äußerst aufwendig. In diesen Fällen wird man sich in der Regel mit stochastischen Simulationen begnügen.

Läßt man den Zeitindex t in Gleichung (26.19) gegen unendlich streben, so erhält man die Varianz von $Y(t)$ bei Vorliegen eines Gleichgewichtes. Diese wird durch

$$\text{var } \bar{Y} = \text{var } \varepsilon \frac{1 + \alpha\beta}{1 + \alpha\beta - \alpha^2 - 2\alpha^2\beta - 2\alpha^2\beta^2 + \alpha^3\beta + 2\alpha^3\beta^2} \quad (26.20)$$

beschrieben.

Gleichung (26.20) ermöglicht die Aufstellung eines Höhenliniendiagrammes der Varianzmultiplikatoren in Abhängigkeit von α und β . Als Varianzmultiplikator wird dabei der in (26.20) angegebene Faktor bezeichnet, dessen Multiplikation mit $\text{var } \varepsilon$ die Gleichgewichtsvarianz $\text{var } \bar{Y}$ ergibt. Man erkennt, daß mit wachsendem α die Varianz $\text{var } \bar{Y}$ stärker zunimmt als mit wachsendem β . Es zeigt sich, daß im Bereich der Grenzlinie zum instabilen Bereich $\alpha = 4\beta(1+\beta)^2$ die Varianzmultiplikatoren sehr empfindlich auf Parameteränderungen reagieren.

Die Ermittlung der einzelnen Koordinatenwerte für $\text{var } \bar{Y}$ in Abhängigkeit von α und β hätte im Falle einer pseudoinduktiven Analyse, d.h.

der Durchführung von stochastischen Simulationen, einen beträchtlichen Rechenaufwand zur Folge gehabt.

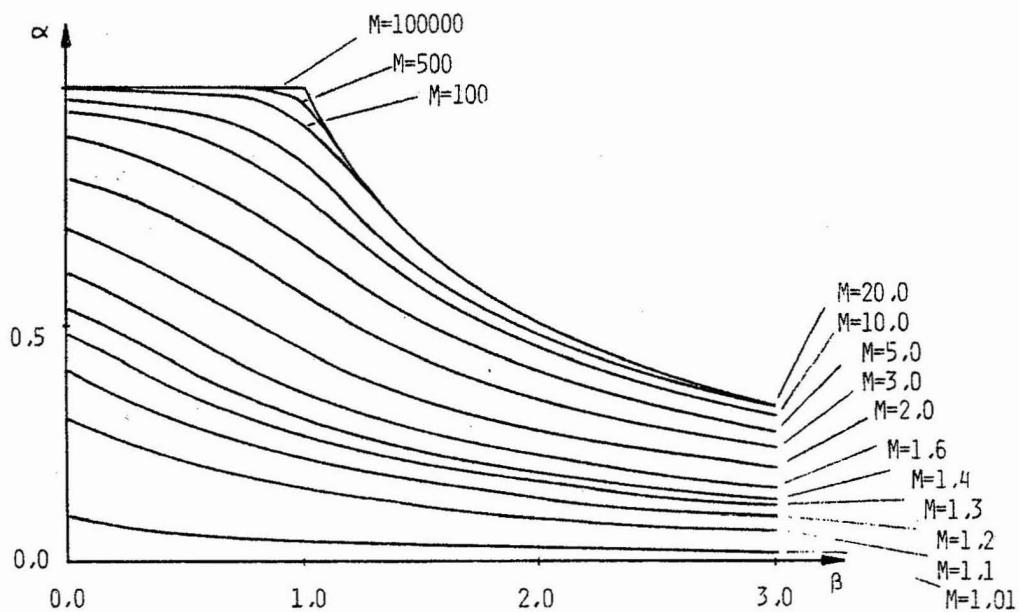


Abb. 26.2 Höhenlinienverlauf der Varianzmultiplikatoren der Variablen Y eines MA-Modells in Abhängigkeit von α und β

Neben dem erörterten Verfahren zur Analyse linearer Störgrößenmodelle gibt es eine Reihe von Einzelformen stochastischer Modelltypen, für welche eine Ermittlung des zeitlichen Verlaufs der Dichtefunktion der endogenen Variablen möglich ist. Eine generelle Theorie existiert jedoch nicht. In den letzten Jahren ist ein Verfahren entwickelt worden, das in beschränktem Umfang ein Urteil über die stochastische Stabilität nichtlinearer Modelle zuläßt. Dieses auf BUCY und KUSHNER zurückgehende Verfahren zur Beurteilung der stochastischen Stabilität nichtlinearer stochastischer Differenzengleichungen beruht auf einer Ausdehnung der deterministischen Stabilitätstheorie von LJAPUNOW auf stochastische Fälle. [27], [121]

Auch in diesem Fall ergeben sich aber dieselben Schwierigkeiten bei der Wahl einer geeigneten LJAPUNOW-Funktion, die schon im deterministischen Fall geschildert wurden.⁹ Bei der Analyse ökonomischer Zu-

⁹ Vgl. Seite 265

sammenhänge ist dieses Verfahren der stochastischen Stabilitätsanalyse von TURNOVSKY zur Untersuchung von Marktgleichgewichten verwendet worden. [213]¹⁰

C. Pseudoinduktive Analyse stochastischer Modelle

a) Grundlagen der Parameterschätzung stochastischer Modellimplikationen

Mangels geeigneter deduktiver mathematischer Verfahren zur Analyse stochastischer Modelle bedient man sich heute vorwiegend einer Untersuchungsmethode, die wir bereits im Einführungsteil als pseudoinduktive Modellerschließung bezeichnet haben.¹¹

Die Implikationen eines stochastischen Modells werden hierbei durch Simulationsexperimente am Modell ermittelt. Vorab stellt sich die Frage, was man unter der Simulation eines stochastischen dynamischen Modells zu verstehen hat.

Betrachten wir die stochastische Endgleichung

$$\begin{aligned} Y(t) &= 1,89Y(t-1) - 0,99Y(t-2) + 4 \cdot 10^3 + \varepsilon(t) \\ Y(0) &= 60 \cdot 10^9 \quad \text{und} \quad Y(1) = 65 \cdot 10^9 \end{aligned} \quad (26.21)$$

wobei $\varepsilon(t)$ einer Normalverteilung mit dem Mittelwert 0 und der Standardabweichung 10^9 angehören soll. Eine Simulation des Zeitpfades von $Y(t)$ kommt dadurch zustande, daß in jeder Periode t eine Stichprobe aus der ε zugrunde liegenden Normalverteilung gezogen wird, und der ermittelte numerische Wert von ε damit eine sukzessive numerische Berechnung von Y ermöglicht. Wenn vom Ziehen einer Stichprobe gesprochen wird, so ist dies nicht allzu wörtlich zu nehmen. Die Stichproben werden mit Hilfe von Computerprogrammen, den sogenannten Pseudozufallszahlengeneratoren, erzeugt.¹²

¹⁰ Eine ökonomische Analyse stochastischer Differentialgleichungssysteme mit Hilfe von stochastischen LJAPUNOW-Funktionen hat WEINBERG vorgenommen. [217]

¹¹ Vgl. Seite 99

¹² Zur Entwicklung von Programmen, welche in der Lage sind, Zufallszahlen bestimmter Dichtefunktionen zu erzeugen siehe [146]

Von Pseudozufallszahlen spricht man, weil es dem Wesen einer Zufallszahl an sich widerspricht, wenn sie von einer bestimmten Programmroutine erzeugt wird. Dennoch werden die so erzeugten Zahlen als Zufallszahlen akzeptiert, wenn sich zeigt, daß alle einschlägigen Tests sie als Zufallszahlen bestätigen.

Im Falle des obigen Beispiels kann man im Rahmen der Simulations-sprache DYNAMO die Endgleichung (26.21) durch die Instruktion¹³

$$Y.K=1.89*VE1(Y.K,60E9)-0.99*VE2(Y.K,60E9,65E9)+4E3+NORMRN(0,1E9)$$

darstellen. Der Ausdruck NORMRN(0,1E9) bildet hierbei die Anweisung, eine normalverteilte Zufallszahl mit den geforderten Parametern ihrer Ausgangsverteilung zu liefern.

Durch wiederholte Simulationen eines Modells erhält man s Zeitverläufe $Y^i(t)$ mit ($i=1,2,\dots,s$) der endogenen Variablen Y . Diese Zeitverläufe dienen als Ausgangsbasis der pseudoinduktiven Modelluntersuchung.

Auf der Grundlage dieser Zeitreihen ist es möglich, eine Information über die zeitliche Entwicklung des Erwartungswertes und der Varianz der endogenen Variablen zu erhalten.

Eine Schätzung des Erwartungswertes $\mu(t)$ erhält man durch

$$\tilde{\mu}(t) = \frac{1}{n} \sum_{i=1}^n Y^i(t) \quad (26.22)$$

während die Varianz $\sigma^2(t)$ anhand von

$$\tilde{\sigma}^2(t) = \frac{1}{n-1} \sum_{i=1}^n [Y^i(t) - \tilde{\mu}(t)]^2 \quad (26.23)$$

geschätzt werden kann.

Da für die Schätzung des Wertes der Varianz $\tilde{\sigma}^2(t)$ der tatsächliche Erwartungswert $\mu(t)$ erforderlich ist, dieser jedoch kaum bekannt sein dürfte, so behilft man sich oft damit, daß $\mu(t)$ in (26.23) durch $\tilde{\mu}(t)$ ersetzt wird.

Mit Hilfe sogenannter Vertrauensbereiche kann man die Genauigkeit der Schätzung des Erwartungswertes und der Varianz beurteilen.

Nach dem zentralen Grenzwertsatz sind die Schätzwerte von μ bei genügend großem n ($n>30$) annähernd normalverteilt und besitzen den Mittelwert μ sowie die Varianz σ^2/n . Dabei ist es unerheblich, durch welche Verteilung $Y(t)$ beschrieben wird.

¹³ Zur Erklärung siehe im einzelnen Seite 522ff.

Wenn μ normalverteilt ist, dann ist es auch die Verteilung

$$z = \frac{\tilde{\mu} - \mu}{\sigma} \sqrt{n}$$

welche den Mittelwert Null und die Varianz Eins besitzt. Für diesen Verteilungstyp läßt sich die Wahrscheinlichkeit γ bestimmen, mit der die Realisation von z in das Intervall $\pm c_\gamma$ fällt, d.h.

$$P(-c_\gamma \leq \frac{\tilde{\mu} - \mu}{\sigma} \sqrt{n} \leq c_\gamma) = \gamma$$

Hieraus folgt die Beziehung

$$P(\tilde{\mu} - c_\gamma \frac{\sigma}{\sqrt{n}} \leq \mu \leq \tilde{\mu} + c_\gamma \frac{\sigma}{\sqrt{n}}) = \gamma \quad (26.24)$$

welche besagt, daß der Bereich

$$\tilde{\mu} \pm c_\gamma \frac{\sigma}{\sqrt{n}} \quad (26.25)$$

den unbekannten Mittelwert μ mit der Wahrscheinlichkeit γ einschließt. Der Sicherheitsfaktor c_γ wird üblicherweise anhand einer der folgenden drei Fehler-Intervalle gewählt.

Wahrscheinlichkeit, daß μ in das Fehlerintervall fällt	Sicherheitsfaktor c_γ
95 %	1,96
99 %	2,58
99,9 %	3,29

Da die für (26.25) erforderliche tatsächliche Standardabweichung σ nicht bekannt ist, muß sie durch den aus der Stichprobe ermittelten Schätzwert $\tilde{\sigma}$ ersetzt werden.

Ist die Verteilung von Y eine Normalverteilung, so wird die Zufallsvariable

$$T = \frac{\tilde{\mu} - \mu}{\sigma^*} \sqrt{n}$$

durch eine t-Verteilung mit $n-1$ Freiheitsgraden beschrieben. σ^* ist

hierbei eine Zufallsvariable, von der $\tilde{\sigma}$ einen Beobachtungswert repräsentiert. Unter Verwendung einer t-Verteilungstafel kann man den Wert für c_{γ} mit dem Vertrauensbereich

$$\tilde{\mu} \pm c_{\gamma} \frac{\tilde{\sigma}}{\sqrt{n}} \quad (26.26)$$

ermitteln.

Bei einer Stichprobe von $n > 30$ ist die Beziehung auch auf beliebige Verteilungen von Y anwendbar. Die t-Verteilung nähert sich in diesem Fall einer Normalverteilung so, daß zur Bestimmung des Intervales in (26.26) die angeführten Werte von c_{γ} verwendet werden können.

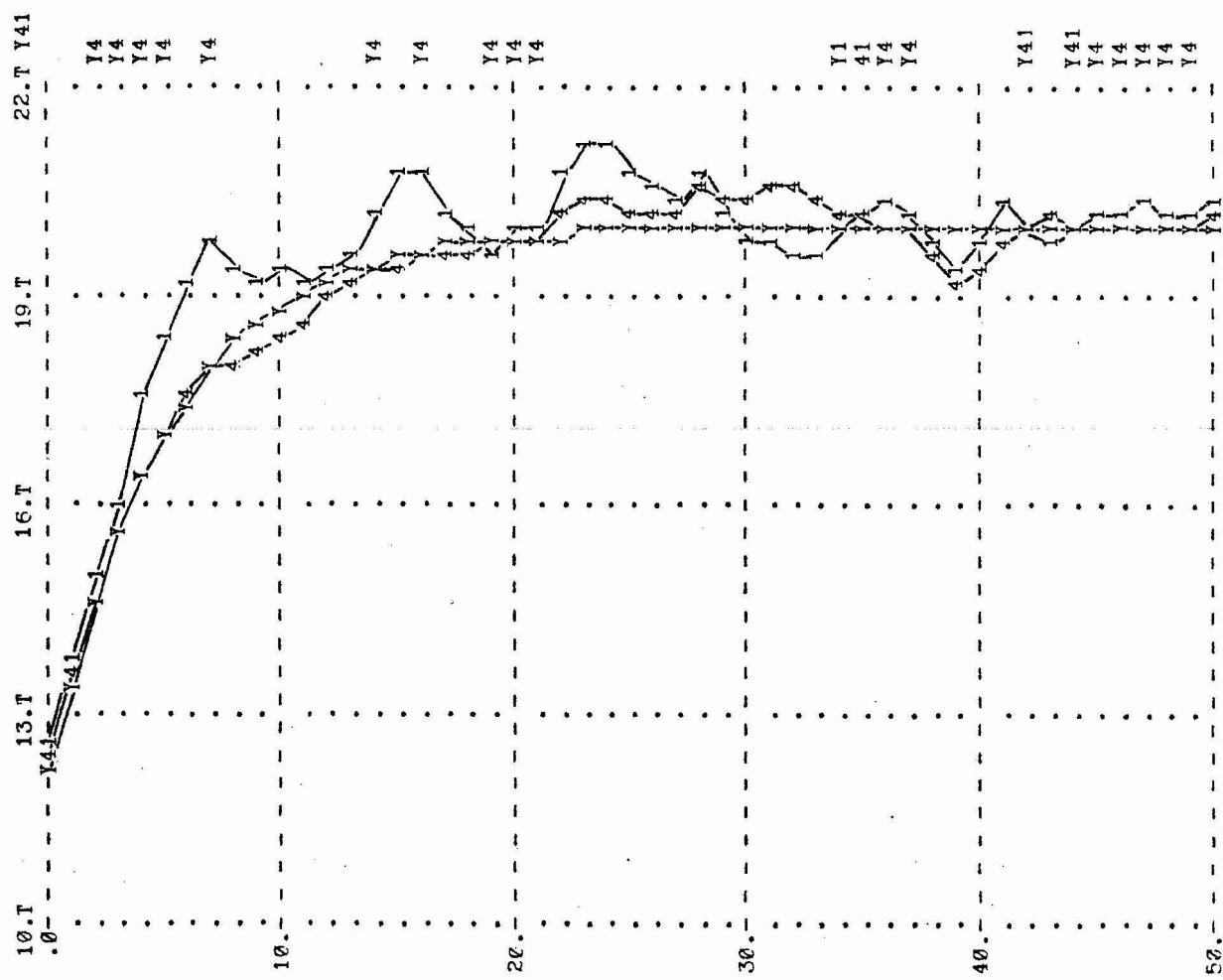


Abb. 26.3 Zeitlicher Verlauf des Erwartungswertes von $Y(t)$ [Symbol 'Y'] und seiner Schätzwerte $\tilde{Y}_{10}(t)$ [Symbol '1'] und $\tilde{Y}_{40}(t)$ [Symbol '4']. [Einheit T: Tausend]

Abbildung 26.3 zeigt für das auf den Seiten 74 und 97 beschriebene stochastische MA-Modell den analytisch berechneten Verlauf des Erwartungswertes $\bar{Y}(t)$ sowie die Verläufe der Schätzwerte für $\bar{Y}(t)$ anhand von zehn und vierzig Schätzungen $\tilde{Y}_{10}(t)$ und $\tilde{Y}_{40}(t)$. Eine analoge Darstellung liefert Abbildung 26.4 für die Standardabweichung.

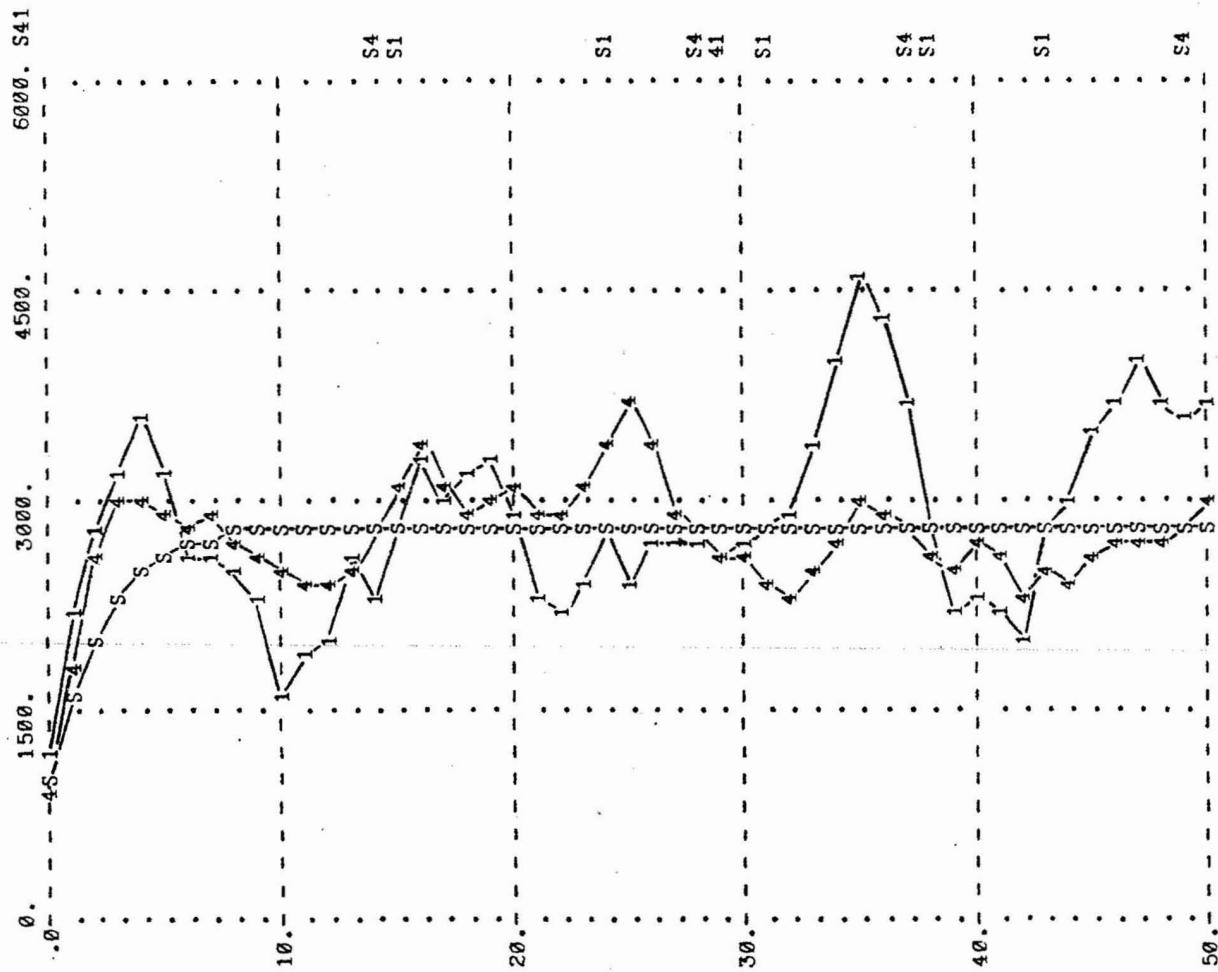


Abb. 26.4 Zeitlicher Verlauf der Standardabweichung σ der endogenen Variablen $Y(S)$ sowie ihrer Schätzwerte $\tilde{\sigma}_{10}(1)$ und $\tilde{\sigma}_{40}(4)$

In der Praxis werden derartige Ermittlungen stochastischer Implikationen allerdings selten vorgenommen. Zumeist werden stochastische Modelle determinisiert, und der Zeitverlauf des deterministischen Äquivalentes wird als Erwartungswert interpretiert.

b) Varianzreduzierende Verfahren im Rahmen der Parameterschätzung
stochastischer Modellimplikationen

Die Durchführung pseudoexperimenteller Schätzverfahren anhand stochastischer dynamischer Modelle konzentriert sich fast ausschließlich auf die Schätzung der Erwartungswerte und Standardabweichungen einzelner interessierender endogener Modellvariablen.

Hat man den Wunsch, den durch (26.25) beschriebenen Vertrauensbereich zu vermindern, so zeigt sich, daß dies mit einer überproportionalen Erhöhung der Stichprobenanzahl n erkauft werden muß. Eine Verminderung des Vertrauensbereiches um den Faktor 10 erfordert beispielsweise eine Erhöhung der Stichprobenanzahl um den Faktor 100. Dieser überproportionale Mehrbedarf an Simulationsläufen kann zu einer Erhöhung der Rechenzeit führen, die auch bei einer Verwendung der heute zur Verfügung stehenden Hochleistungsrechner ins Gewicht fällt. Zur Abschwächung des Konkurrenzverhältnisses zwischen Rechenzeitbedarfserhöhung und Verminderung des Vertrauensbereiches dient die Anwendung sogenannter varianzreduzierender Verfahren.

Ist für die Schätzung von μ beispielsweise ein bestimmter Vertrauensbereich vorgegeben, so ist es unter Anwendung varianzreduzierender Verfahren möglich, die Zahl der zur Einhaltung dieses Bereiches erforderlichen Simulationen n_e zu vermindern. Der Grund hierfür ist folgender: Es läßt sich zeigen, daß zwischen der zur Einhaltung eines bestimmten Vertrauensbereiches (unter Vorgabe einer bestimmten Sicherheitswahrscheinlichkeit) erforderlichen Stichprobenanzahl n_e und der Varianz σ^2 der betrachteten Zufallsvariablen die Beziehung

$$n_e = b\sigma^2(t) \quad b>0 \quad (26.27)$$

gilt. [109, S.106]

Gelingt es nunmehr, die stochastische Struktur eines Modells so zu verändern, daß die interessierende Variable zwar denselben Erwartungswert behält, aber ihre Varianz vermindert wird, dann kann wegen (26.27) die für den angestrebten Vertrauensbereich erforderliche Stichprobenanzahl vermindert werden. Wenn also die Varianz um die Hälfte redu-

ziert wird, kann n_e um ein Viertel gesenkt werden.

Die Existenz dieser Beziehung zwischen n_e und σ^2 bildet den Ausgangspunkt für die Entwicklung varianzreduzierender Verfahren.

Die theoretische Fundierung dieser Methoden ist in der Literatur ausführlich dargestellt [132], [109], [137]. Sie werden jedoch insbesondere bei MZÄ-Modellen kaum angewendet. Hierbei spielt sicher eine Rolle, daß das Verfahren einer stochastischen Simulation und die daran anknüpfende Parameterschätzung keine weitgehenden statistischen Kenntnisse voraussetzen, während dies bei der Anwendung varianzreduzierender Verfahren der Fall ist. Selbst mit statistischen Methoden weitgehend vertraute Anwender scheinen offenbar höhere Rechenzeiten der Entwicklung eines der Varianzreduzierung dienenden Computerprogrammes vorzuziehen.

Wegen der geringen praktischen Relevanz sollen daher bis auf diese kurzen Bemerkungen varianzreduzierende Verfahren nicht weiter erwähnt werden.

D. Subjektive Entscheidermodelle und stochastische Analyse

Wie erwähnt, werden in dynamischen Entscheidermodellen die unbeeinflußbaren Parameter durch subjektive Schätzungen der Modellanwender ermittelt.¹⁴ Das Endergebnis ist ein deterministisches Modell. Derartige Modelle können jedoch als determinisierte stochastische Modelle gedeutet werden, in welchen der stochastische Parameter durch den vom Modellanwender subjektiv geschätzten Erwartungswert oder den häufigsten Wert ersetzt worden ist. Geht man davon aus, daß die gesamte Information über die subjektiven Wahrscheinlichkeitsverteilungen der einzelnen Parameter in das Modell mit einfließen sollen, so gelangt man von dem ursprünglich deterministischen Modell zu einem stochastischen Modell. Dieses Modell gestattet Aussagen über die subjektiven Wahrscheinlichkeiten, mit denen bestimmte Verläufe der endogenen Variablen auftreten. Solche zusätzlichen Informationen sind für die Beurteilung der prognostischen Qualität der aufgestellten Modellthesen von erheblicher Bedeutung.

¹⁴ Vgl. Seite 150

Betrachten wir beispielsweise ein subjektives determinisiertes Entscheidermodell:

$$Y(t) = 0,97Y(t-1) + 120 \quad (26.29)$$

so zeigt sich, daß dieses Modell gegen den Gleichgewichtspfad 4000 konvergiert. Nehmen wir nunmehr an, daß der Parameter von 0,97 als häufigster Wert der folgenden subjektiven Wahrscheinlichkeitsverteilung gewählt wurde, dann ist die Wahrscheinlichkeit, daß ein Parameter realisiert wird, der größer als Eins ist, höher als der umgekehrte Fall. Damit ist die Wahrscheinlichkeit höher, daß das System

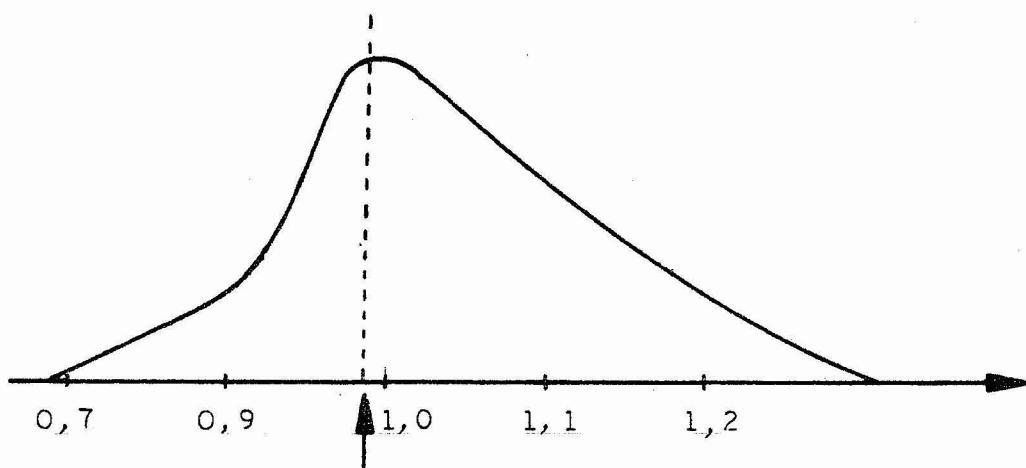


Abb. 26.5 Subjektive Wahrscheinlichkeitsverteilung eines Parameters

explodiert und daher nicht stabil ist. Liegt dieser Informationsstand vor, dann wäre es für einen Modellanwender nicht sinnvoll, von der Hypothese (26.29) auszugehen. Es zeigt sich somit, daß die Entscheidung, ob eine angenommene Hypothesengleichung sinnvollerweise zu akzeptieren ist, erst dann getroffen werden kann, wenn die subjektive Wahrscheinlichkeitsverteilung der Parameter mitberücksichtigt wird.

Diese Einsicht legt es nahe, nicht mehr mit deterministischen Entscheidermodellen zu arbeiten, sondern allein ihre stochastisierte Form zu verwenden. Als Ausweg bietet sich ein 'Validierungsverfahren' an, mit dem man ein deterministisches Modell nachträglich daraufhin

überprüfen kann, ob es sich weiterhin als Prognoseinstrument verwenden läßt. Dieses Verfahren einer Hypothesenüberprüfung durch Sensitivitätsanalysen läßt sich etwa folgendermaßen umreißen:

Zeigt sich, daß nur eine geringfügige Änderung eines oder mehrerer subjektiv geschätzter Parameter zu einem grundsätzlich anderen Systemverhalten führt, dann wird die empirische Gültigkeit der unterstellten Hypothesen erschüttert. Im anderen Fall bewährt sich die untersuchte Hypothese. Diesen Schlußfolgerungen kann auf der Basis, daß von einem entsprechenden subjektiven stochastischen Entscheidermodell ausgegangen wird, eine methodisch befriedigende Rechtfertigung gegeben werden. Einer Parameteränderung liegt bei dieser Betrachtungsweise die Annahme zugrunde, daß die geringfügige Änderung eines Parameters 'genauso wahrscheinlich' ist wie der Ausgangswert. Führt diese Änderung jedoch zu einem grundsätzlich anderen Systemverhalten, so kann man daraus schließen, daß das eine Systemverhalten genauso wahrscheinlich ist wie das andere. Als Folge davon ist eine für praktische Zwecke geeignete Prognose nicht möglich, weil das Modell stark voneinander abweichende Alternativen mit annähernd gleichen Wahrscheinlichkeiten auszeichnet.

Würde ein entsprechendes stochastisches Modell zur Verfügung stehen, so könnte man diese Argumentation allerdings durch die Angabe subjektiver Wahrscheinlichkeiten in exakterem Maße belegen als in der recht unsystematischen Form einer Sensitivitätsanalyse.