

Verfahren zur Absatzvorausschätzung

Peter Naeve

1. Vorbemerkung

Aus der Sicht eines Statistikers ist man im ersten Augenblick geneigt, dem Zusatz „Absatz“ keine Bedeutung zuzubilligen. Ein Durchmustern der zahlreichen Literatur über Prognosemethoden, die sich an den Absatzspezialisten wendet, scheint die These zu stützen, daß es sich dabei im Grunde lediglich um *Probleme der Zeitreihenanalyse* handelt. Nimmt man für einen Moment diesen Standpunkt ein, dann ist eine Gliederung nach *univariaten* und *multivariaten* Prognoseverfahren nahelegend. Innerhalb jeder dieser Gruppen wäre dann noch eine Unterteilung nach der Anzahl der zu prognostizierenden Zeitreihen angebracht. Es leuchtet ein, daß die Absatzprognose für ein Sortiment von einigen tausend Artikeln nur mit Hilfe von EDV-Anlagen durchgeführt werden kann und daß dabei aus Kosten- und Zeitgründen nur einfache und daher statistisch nicht völlig befriedigende Verfahren eingesetzt werden können. Anders kann der Fall liegen, wenn es sich um hoch aggregierte Zeitreihen handelt. Da diese in der Regel nur jährlich oder halbjährlich prognostiziert werden, kann man aufwendigere Verfahren heranziehen.

Schon der Übergang von den univariaten Verfahren zu den multivariaten stellt einen Versuch dar, die Prognoseverfahren dem inhaltlichen Prognoseproblem anzupassen. Die Wahl der außer der Reihe der Absatzzahlen herangezogenen Zeitreihen sollte nach sachlichen und nicht nach statistischen Gesichtspunkten – z. B. Erzielung eines möglichst großen R^2 bei einer Regression – erfolgen. Prognoseverfahren verdienen aber erst dann das Prädikat „zur Absatzvorhersage“, wenn sie sich den speziellen Gegebenheiten des Absatzbereichs anpassen. Sie müssen in der Lage sein, die nicht als Zeitreihendaten anfallenden Aktivitäten im Absatzbereich für die Prognose heranzuziehen. Bei der Multiplen Regression versucht man dieses Problem mit der Einführung von sogenannten Dummy-Variablen höchst künstlich zu lösen.

Als Dummy-Variable bezeichnet man dabei eine Größe, die nur die Werte 0 und 1 annehmen kann. Ihren Gebrauch verdeutlicht am besten ein Beispiel. Nehmen wir an, in den Perioden $s, s + 1, \dots, s + r$

habe ein Streik in einem Wirtschaftssektor stattgefunden. Sicherlich beeinflußt ein Streik die Absatzmöglichkeiten vieler Artikel. Um diesen Einfluß in der Regression berücksichtigen zu können, wählt man die Dummy-Variable Streik. Diese Variable – d. h. die Zeitreihe ihrer „Beobachtungen“ – ist in allen Perioden 0 bis auf die Perioden s , $s + 1, \dots, s + r$; dort hat sie den Wert 1.

Zumindest im Falle von Prognoseproblemen, die nicht wie oben erwähnt mit großen Mengen von zu prognostizierenden Zeitreihen verbunden sind, scheint das Verlangen nach Konstruktion geeigneter „Absatz“-Prognosemodelle gerechtfertigt.

Gemäß dieser Vorbemerkung sollen in diesem Papier die folgenden Punkte behandelt werden:

- Exponential Smoothing
- Multiple Regression
- Spektralanalyse
- Prognosemodelle.

Dabei wird nicht eine geschlossene Darstellung der jeweiligen Methoden angestrebt, sondern es soll jeweils die grundlegende Idee herausgestellt und auf mit der Anwendung zusammenhängende Probleme eingegangen werden. Auf einen gewissen mathematischen Formalismus kann man dabei aber schon deswegen nicht verzichten, weil sonst zu den schon vorhandenen Problemen noch semantische Schwierigkeiten kämen.

2. Exponential Smoothing

Ehe wir das Verfahren vorstellen, sind einige Vereinbarungen notwendig. Den Wert der (Absatz-)Zeitreihe im Zeitpunkt t wollen wir mit x_t bezeichnen. Eine Prognose für den Zeitreihenwert im Zeitpunkt $t + 1$, die zum Zeitpunkt t erstellt wird, soll mit $\hat{x}_t(t + 1)$ bezeichnet werden. Es sei unterstellt, daß die Zeitreihenwerte und die Prognosewerte für äquidistante Zeitpunkte vorliegen bzw. interessieren.

Das auf Brown¹ zurückgehende Verfahren des Exponential Smoothing ist ein *univariates Verfahren*, d. h. außer den bekannten Werten der zu prognostizierenden Zeitreihe gehen keine weiteren Informationen in die Prognose ein. Wenn man auch ziemlich viel Mathematik im Zusammenhang mit diesem Verfahren anbringen kann, so ist es doch ein sogenanntes Ad-hoc-Verfahren, das auf folgender heuristischen Idee beruht:

$$\hat{x}_t(t + 1) = \hat{x}_{t-1}(t) + \alpha(x_t - \hat{x}_{t-1}(t)) \quad 0 \leq \alpha \leq 1. \quad (1)$$

Das Prinzip ist sofort zu erkennen. Man möchte die zum Zeitpunkt $t - 1$ für den Zeitpunkt t erstellte Prognose auch als Prognose zum Zeitpunkt t für den Zeitpunkt $t + 1$ beibehalten. Inzwischen – wir befinden uns im Zeitpunkt t – kann man aber die Abweichung der Prognose $\hat{x}_{t-1}(t)$ vom realisierten Zeitreihenwert x_t berechnen. Hat man den Wert unterschätzt, d. h. ist $x_t - \hat{x}_{t-1}(t) > 0$, so korrigiert man die alte Prognose um den α -ten Teil der Abweichung nach oben, ist die Abweichung negativ, d. h. hat man den realisierten Wert überschätzt, wird die alte Prognose nach unten um den α -ten Teil der Abweichung korrigiert.

Wie man sieht, drückt die Gleichung (1) durchaus ein oft zu beobachtendes menschliches Verhalten aus. Betrachten wir die beiden Grenzfälle $\alpha = 0$ und $\alpha = 1$. Im ersten Fall handelt es sich um ein durch nichts zu erschütterndes Verhalten; die einmal gefundene Prognose wird auch bei größten Abweichungen beibehalten. Der Fall $\alpha = 1$ stellt das schwankende Rohr im Winde dar. Immer der zuletzt realisierte Wert wird als neue Prognose genommen. Erfahrungen werden nicht gemacht. Durch Festlegung eines α erhält man aus dem heuristischen Ansatz (1) eine Prognoseformel. In der Literatur wird gewöhnlich mit α -Werten in der Gegend von 0.1 gearbeitet. Diese Wahl ist aber in der Regel methodisch und nicht sachlich motiviert.

Die Bezeichnung Exponential Smoothing für das vorgestellte Prognoseverfahren wird klar, wenn man die folgende Umformung von Gleichung (1) betrachtet:

$$\hat{x}_t(t+1) = \sum_{j=0}^{\infty} \alpha (1-\alpha)^j x_{t-j}. \quad (2)$$

Man gelangt zu dieser Darstellung, indem man die Formel (1) umformt zu

$$\hat{x}_t(t+1) = \alpha x_t + (1-\alpha) \hat{x}_{t-1}(t). \quad (3)$$

Verwendet man diese Form auch für die Prognose im Zeitpunkt $t - 1$, d. h.

$$\hat{x}_{t-1}(t) = \alpha x_{t-1} + (1-\alpha) \hat{x}_{t-2}(t-1), \quad (4)$$

und setzt diesen Prognosewert in (3) ein, so ergibt sich

$$\hat{x}_t(t+1) = \alpha x_t + \alpha(1-\alpha)x_{t-1} + (1-\alpha)^2 \hat{x}_{t-2}(t-1). \quad (5)$$

Ersetzt man mit Hilfe von (3) nacheinander $\hat{x}_{t-2}(t-1)$, $\hat{x}_{t-3}(t-2)$ usw., so gelangt man zur Darstellung der Gleichung (2).

Die Vergangenheitswerte x_{t-j} gehen jeweils mit dem Gewicht $\alpha(1-\alpha)^j$ in die Prognose ein. Die Gewichte fallen exponentiell ab, daher der

Name. In der Literatur ist es üblich, die Summenbildungsvorschrift in Gleichung (3) als Operator aufzufassen. Man setzt

$$S_t(x) = \sum_{j=0}^{\infty} \alpha(1 - \alpha)^j x_{t-j} . \quad (6)$$

In der Schreibweise $S_t(x)$ gibt der Index t an, in welchem Zeitpunkt der Operator startet; die Größe in der Klammer repräsentiert die Zeitreihe, auf die der Operator einwirkt.

Da die überwiegende Zahl der in der Anwendung befindlichen Prognose-systeme in der einen oder anderen Form das heuristische Prinzip des Exponential Smoothing verwendet, sei hier die Prognoseformel (1) als Algorithmus dargestellt. Wir verwenden eine von Knuth⁴ vorgeschlagene Darstellungsweise von Algorithmen.

Algorithmus: Exponential Smoothing

Auf den Plätzen X , S , A stehen der letzte beobachtete Zeitreihenwert, die Prognose (alte und neue) und der gewählte Wert für α .

A 1 (Initialisierung)	Bringe nach S einen Wert für eine Anfangsprognose und nach A den Wert des gewünschten α
A 2 (Neue Realisation)	Lese X ein
A 3 (Prognose)	$S \leftarrow S + A * (X - S)$
A 4 (Ausgabe)	Schreibe S
A 5 (Weitere Prognosen?)	Wenn weitere Prognosen gewünscht \rightarrow A 2
A 6	Ende

An diesem Algorithmus wird sofort deutlich, warum sich das Prinzip des Exponential Smoothing so großer Beliebtheit erfreut. Obwohl in die Prognose alle Vergangenheitswerte eingehen (siehe Gleichung (2)), benötigt man zur Berechnung einer neuen Prognose nur die 3 Werte X , S und A . Damit ist dieses Verfahren sehr gut geeignet für die in den Vorbemerkungen angeführten Fälle der Absatzprognose für eine Vielzahl von Zeitreihen (z. B. realistische Lagerhaltungsprobleme), denn für jede Zeitreihe braucht man höchstens 3 Werte und nicht die ganze Vergangenheit zu speichern. In der Regel arbeiten entsprechende Systeme mit einem α -Wert für eine größere Menge von Zeitreihen, was den Speicherbedarf noch weiter reduziert.

Das Prinzip des Exponential Smoothing wird vielleicht an dem folgenden Beispiel noch anschaulicher. Es sollte dabei die Zeitreihe

$$x_1 = 4.1, x_2 = 5.2, x_3 = 5.9, x_4 = 6.3 \quad \text{usw.}$$

sukzessive prognostiziert werden. Der „Schritt 3“ liefert die Prognose für x_4 . Man hatte $\alpha = 0,333$ gewählt. Die Prognose $\hat{x}_1(2) = 4.7$ ergab sich dadurch, daß die Anfangsprognose $\hat{x}_0(1) = 5$ gesetzt wurde. In der Graphik erkennt man sehr deutlich die exponentiell fallenden Gewichte, mit denen die Vergangenheitswerte belegt werden (Abb. 1).

SCHRITT: 3

4.1	5.2	5.9 [*]	6.3	7.1	6.8	7.0	5.8	5.1
4.7	4.9	5.2						
	*	=						

EXP. GEGLAETTETER WERT DER VORPERIODE = 4.9
 ZU GLAETTENDER WERT IN PERIODE 3 = 5.9

BERECHNUNG:
 $0.333 * 5.9 + 0.667 * 4.9 = 5.2$

NR	WERT	GEWICHTUNG (PROZENT)
1	4.1	XXXXXXXXXXXXXXXXXXXX 14.81
2	5.2	XXXXXXXXXXXXXXXXXXXX 22.21
3	5.9	XXXXXXXXXXXXXXXXXXXX 33.30
4	6.3	
5	7.1	
6	6.8	
7	7.0	
8	5.8	
9	5.1	

SUMME DER GEWICHTUNGSFAKTOREN VORHERIGER, HIER
 NICHT AUFGEFUEHRTER DATEN IN PROZENT = 29.67

Abb. 1. Prinzip des Exponential Smoothing

Untersucht man das hier vorgestellte Verfahren unter statistischen Gesichtspunkten, so stellt sich heraus, daß nur Zeitreihen des Typs $x_t = a_0 + u_t$, wobei a_0 eine Konstante und u_t eine zufällige Größe mit dem Mittelwert 0 sind, im Mittel richtig prognostiziert werden. Dies Ergebnis ist aufgrund der heuristischen Überlegungen im Zusammenhang mit der Gleichung (1) nicht sonderlich überraschend. Nun zeichnen sich ökonomische Zeitreihen aber gerade dadurch aus, daß sie nicht von dem simplen Typ $x_t = a_0 + u_t$ sind, sondern in der Regel Trend und Saisonkomponenten enthalten. Es läßt sich zeigen, daß man auch für Zeitreihen eines solchen Typs die

Idee des Exponential Smoothing sinnvoll – d. h. so, daß auch statistischen Kriterien Genüge getan wird – anwenden kann. Dies soll hier nicht ausgeführt werden, da die verständliche Herleitung und Begründung der benötigten Formeln und Algorithmen den hier zur Verfügung stehenden Platz sprengen würde. Im Prinzip laufen aber alle diese Verfahren darauf hinaus, das hier vorgestellte einfache oder – wie man auch sagt – exponentielle Glätten 1. Ordnung auf aus den Zeitreihen geeignet konstruierte Größen anzuwenden, es mehrfach hintereinander auszuführen – d. h. man nimmt als Prognose $\hat{x}_t(t+1) = S_t(S(x))$, genannt exponentielles Glätten 2. Ordnung usw. –, oder Kombinationen dieser Vorgehensweisen zu benutzen.

Die folgende Graphik stellt das Prognoseverhalten des Exponential Smoothing 1. Ordnung (\circ) dem der 2. Ordnung (\triangle) gegenüber. Man erkennt deutlich den systematischen Fehler des ersten Verfahrens bei einer Zeitreihe mit linearem Trend (Abb. 2):

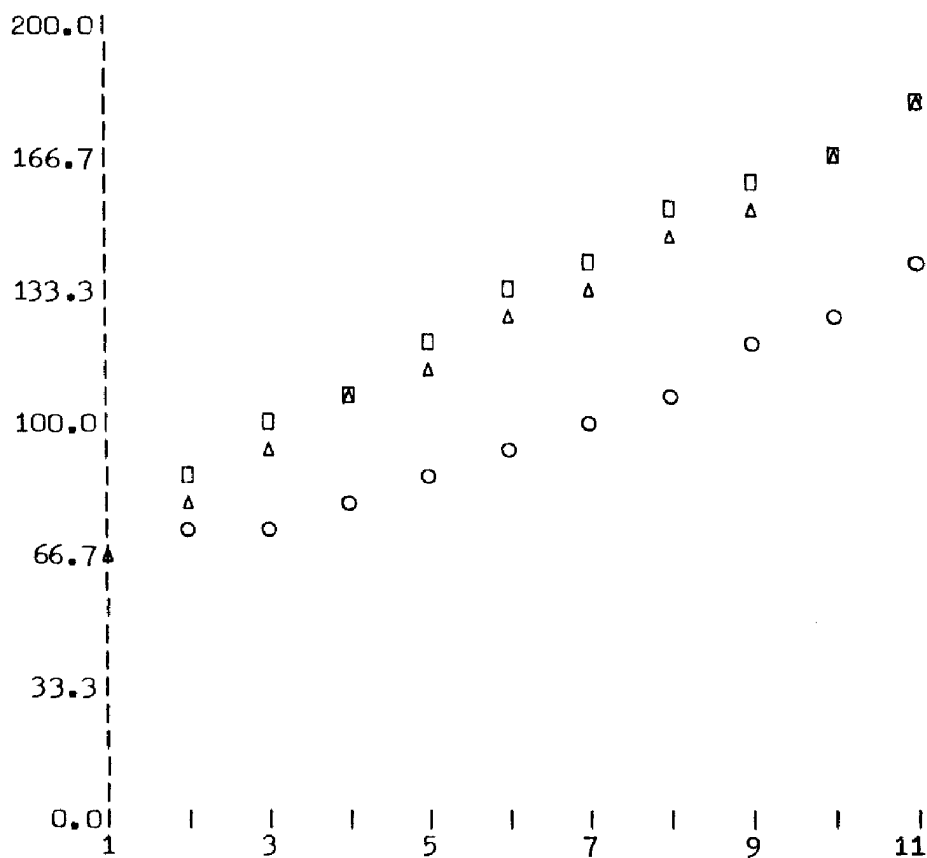


Abb. 2. Exponential Smoothing 1. u. 2. Ordnung im Vergleich

3. Multiple Regression

Prognosen mit Hilfe von Regressionsmodellen gehören zur Klasse der multivariaten Verfahren. Ausgangspunkt ist – oder besser: sollte sein – die Überlegung, daß die betrachtete Absatzzeitreihe in ihrer Ausprägung zum Zeitpunkt t nicht nur von den Vergangenheitswerten der gleichen Reihe abhängt, sondern auch von den Werten anderer Größen, wie z. B. der Kaufkraft, dem Absatz anderer Produkte, Preisen usw. Eine solche Überlegung legt es nahe, als Prognoseformel $\hat{x}_t(t+1)$ eine Funktion aller dieser aus sachlichen Überlegungen in Frage kommenden Größen zu wählen:

$$\hat{x}_t(t+1) = f(x_{1t}, x_{1t-1}, \dots, x_{nt}, x_{nt-1}, \dots) \quad (7)$$

Mit x_{it-j} ist der Wert der i -ten Variablen (auch Einflußgröße genannt) zum Zeitpunkt $t-j$ bezeichnet. Eine der Variablen x_i steht dabei für die Vergangenheitswerte der zu prognostizierenden Reihe. Sind von den übrigen Variablen schon die Werte im Zeitpunkt $t+1$ bekannt, dann können sie durchaus in (7) mit auftreten.

Das Problem mit einer Prognose gemäß (7) besteht nun in zwei Dingen. Einmal muß festgelegt werden, welche Größen von Einfluß auf die Prognose sind, und zum anderen gilt es noch, die funktionale Form der Gleichung (7) zu bestimmen, m. a. W. die Funktion f muß explizit angegeben werden.

Für eine Klasse von Funktionen bietet nun die Statistik ein Vorgehen an, das beide Probleme zu lösen in der Lage ist. Beschränkt man sich in (7) auf lineare Funktionen, d. h.

$$\hat{x}_t(t+1) = a_{00} + a_{10}x_{1t} + a_{11}x_{1t-1} + \dots + a_{n0}x_{nt} + \dots \quad (8)$$

dann lassen sich die in Statistik und Ökonometrie entwickelten Techniken der Regressionsanalyse anwenden. Man erhält nicht nur Verfahren zur Bestimmung der noch unbekanntenen Koeffizienten a_{ij} , sondern auch Entscheidungskriterien für die Auswahl der Einflußgrößen x_i .

Das Vorgehen sei hier für den einfachsten Fall skizziert. Wir werden dabei von unserem Prognoseproblem erst einmal abgehen. Gegeben sei eine sogenannte endogene Variable y und mehrere sogenannte exogene Variable x_i . Die Termini *endogen* und *exogen* stehen in der Ökonometrie für folgenden Tatbestand: Eine Regressionsgleichung ist ein Modell zur Erklärung der links stehenden Größe. Diese zu erklärende Größe nennt man *endogen*, da ihre Werte von dem Modell „erzeugt“ werden. Die unabhängigen Variablen erhalten ihre Werte außerhalb des Modells

zugewiesen, daher werden sie als *exogene* (in bezug auf das Modell) Variable bezeichnet. Endogen und exogen ist eine Größe immer nur relativ zu einem Modell; diese Begriffe stellen keine absolute Klassifikation dar.

Man nimmt an, daß zum Zeitpunkt t die Beziehung

$$y_t = a_0 + a_1 x_{1t} + \dots + a_n x_{nt} + u_t \quad (9)$$

besteht. u_t ist dabei eine Zufallsvariable mit dem Erwartungswert 0. Das bedeutet, daß man davon ausgeht, daß im Mittel die Gleichung (9) unter Weglassung von u_t richtig ist.

Die *Methode der kleinsten Quadrate* erweist sich als ein geeignetes Hilfsmittel zur Bestimmung der unbekanntenen Koeffizienten a_j . Ausgangspunkt ist

$$\sum u_t^2 = \sum (y_t - a_0 - a_1 x_{1t} - \dots - a_n x_{nt})^2, \quad (10)$$

wobei über alle Zeitpunkte t summiert wird, für die Beobachtungen vorliegen. Die Methode der kleinsten Quadrate besagt für dieses Problem, daß die unbekanntenen a_j so zu bestimmen sind, daß der Ausdruck (10) minimiert wird. Die technischen Details seien hier unterdrückt.

Mit u_t ist auch y_t eine Zufallsgröße. Man kann nun die Varianz der Zufallsgröße in zwei Teile aufspalten, den sogenannten erklärten Anteil und den Rest. Mit *erklärtem Anteil* meint man den auf die Variabilität der exogenen Variablen zurückzuführenden Teil der Varianz. Diese Größe ist eine statistische Kenngröße für die jeweilige Regression. Durch Hinzunahme einer weiteren exogenen Variablen kann der Wert dieser Kenngröße nur wachsen. Man wird nun danach streben, solche Variablen in die Gleichung (9) aufzunehmen, die den erklärten Anteil möglichst groß machen, und bei der Frage der Hinzunahme einer weiteren Variablen die auswählen, die einen möglichst großen Zuwachs für den erklärten Anteil hervorruft. Man beachte aber, daß dies eine rein statistische Argumentation ist, die bei blinder Anwendung durchaus zu sachlich nicht haltbaren Erklärungsmodellen vom Typ (9) führen kann.

Die Übertragung des eben geschilderten Konzeptes auf das Prognoseproblem ist jetzt einfach. Man hat nur die Variable y_t mit der Prognose $\hat{x}_t(t+1)$ zu identifizieren. Es treten dann allerdings noch einige methodische Probleme auf, da die Methode der kleinsten Quadrate modifiziert werden muß, wenn beim Prognosesachverhalt unter den exogenen Variablen verzögerte endogene Variablen (y_{t-1} , y_{t-2} usw.) auftreten. Dieses Problem läßt sich lösen und hängt wohlbemerkt nur an dem

Wunsch, gewisse statistische Eigenschaften für die Schätzung der unbekanntenen Koeffizienten zu erhalten.

Prognose mit Hilfe eines Regressionsmodells bedeutet dann ein Vorgehen in zwei Schritten. Zuerst wird für den Ansatz (8) mit Hilfe von Vergangenheitswerten (man ersetzt dabei $\hat{x}_t(t+1)$ durch die tatsächlich beobachteten Werte) eine Koeffizientenschätzung durchgeführt. Im zweiten Schritt werden in die jetzt völlig festgelegte Gleichung (8) die „exogenen“ Größen für den Zeitpunkt $t+1$ (beobachtete: z. B. der Werbeaufwand, geschätzte: z. B. der Preis eines Konkurrenzgutes) eingesetzt. Es ergibt sich daraus die Prognose $\hat{x}_t(t+1)$. Für diese Prognose kann man aufgrund der statistischen Eigenschaften des Regressionsansatzes auch noch Konfidenzintervalle berechnen. Solche Intervalle können als Maß für die Güte der Prognose gelten.

Wenn hier auch alle technischen Details der konkreten Berechnung unterdrückt wurden, so hat der Leser sicher das Gefühl bekommen, daß dieser Prognoseansatz viel aufwendiger ist als der im vorangehenden Abschnitt vorgestellte. Mit Sicherheit wird man ihn nicht für prognostische Massenprobleme verwenden, da er zu daten- und rechenaufwendig ist. Neu hinzugekommene Informationen – und das ist ja wohl in jeder Periode der Fall – verlangen zu ihrer Berücksichtigung eine fast völlig erneute Durchrechnung des Ansatzes. Man wird daher diesen Ansatz auf hochaggregierte Reihen beschränken, m. a. W. auf die Schnittstelle zwischen Unternehmen und ökonomischer Umwelt, d. h. zwischen makro- und mikroökonomischen Größen.

4. Spektralanalyse

Vordergründig scheinen sich Prognoseverfahren, die auf spektralanalytischen Verfahren beruhen, durch einen noch gewaltigeren mathematischen Aufwand von den bisher vorgestellten Methoden zu unterscheiden. Wesentlich ist aber ein anderer Aspekt. Die Anwendung der Spektralanalyse setzt ein ganz bestimmtes Modell für die Zeitreihen voraus. Dieses Modell unterscheidet sich grundlegend von den in den ersten beiden Abschnitten enthaltenen Vorstellungen über Zeitreihenmodelle. Es erscheint sinnvoll, den Leser nicht mit wuchtigen mathematischen Formeln und Herleitungen zu beeindrucken, sondern zu versuchen, die Grundideen herauszuarbeiten.

Ausgangspunkt ist die Vorstellung, daß eine Zeitreihe Realisation eines stochastischen Prozesses ist. Dieser Prozeß rückt in den Mittelpunkt der

Untersuchung. Man ist bestrebt, von der Realisation, d. h. von der Zeitreihe, auf die Struktur des Prozesses zurückzuschließen.

Ein stochastischer Prozeß ist ein „mathematischer Gegenstand“. Man mag ihn sich vereinfachend vorstellen als eine Verallgemeinerung des Begriffs „Zufallsvariable“. Unter Struktur eines Prozesses werden alle die ihn charakterisierenden Größen verstanden. Man erinnere sich, daß eine Zufallsvariable durch Erwartungswert, Varianz usw. charakterisiert wird. Die Struktur eines stochastischen Prozesses besteht unter anderem aus entsprechenden Verallgemeinerungen dieser Größen.

Das folgende Beispiel soll zeigen, daß das mathematische Modell des stochastischen Prozesses unter Umständen gut geeignet erscheint, um eine reale Situation zu beschreiben. Damit wäre dann auch die Anwendung spektralanalytischer Verfahren gerechtfertigt.

Stellen wir uns den Absatz-Manager einer großen Ladenkette vor. Will er den Absatz eines bestimmten Artikels X untersuchen, so kann er sich von jedem einzelnen Laden die Absatzzahlen (oder die Anteile, die dieser Artikel am Umsatz hat, um die Zahlen vergleichbar zu machen) ansehen, er hat also für jeden Laden eine Zeitreihe vorliegen. Wenn ihn nicht die vom allgemeinen Erscheinungsbild der Läden auffällig nach oben oder unten abweichenden interessieren, sondern er allgemein die Entwicklung des Absatzes dieses Artikels untersuchen möchte, dann zielt er doch gerade auf den Prozeß „Absatz des Artikels X“. Jede Absatzreihe für den Artikel X in den einzelnen Läden sagt sicher etwas über diesen Prozeß aus, ist aber nur eine mögliche Realisation. Es ist nun Aufgabe des Managers, aus allen diesen verschiedenen Realisationen das Gemeinsame, eben die Struktur des Absatzprozesses für den Artikel X, herauszufinden.

Die hier geschilderte Situation läßt sich durch das Modell „stochastischer Prozeß“ beschreiben. Der Zusatz „stochastisch“ drückt aus, daß die einzelnen Zeitreihen bei aller Gemeinsamkeit der Entwicklung doch individuelle – aber im Beispiel für den Manager uninteressante – Ausprägungen haben.

Bei den auf dem Konzept des stochastischen Prozesses fußenden Zeitreihenverfahren stehen fast immer die sogenannten Kovarianzfunktionen der Prozesse im Mittelpunkt. Die Kovarianzfunktion gehört zu den charakterisierenden Größen eines Prozesses, die seine Struktur ausmachen.

Die Kovarianzfunktion eines Prozesses ist eine Verallgemeinerung des Begriffs „Kovarianz“ zweier Zufallsvariablen. Die Kovarianz zweier Zufallsvariablen gibt bekanntlich Aufschluß über den (linearen) Zusam-

menhang der beiden Größen. Ein stochastischer Prozeß läßt sich nun auffassen als eine geordnete „Menge“ von Zufallsvariablen. Greift man zwei Zufallsvariablen heraus, kann man für sie die Kovarianz bestimmen. Führt man dies für alle möglichen Paare durch, so erhält man eine „Menge“ von Kovarianzen – diese Menge bezeichnet man als Kovarianzfunktion. Das folgende Bild soll diesen Vorgang verdeutlichen

$$\begin{array}{ccccccc} X_t & X_{t+1} & \dots & X_s & X_{s+t} \\ \hline & | & & | & & & | \end{array}$$

Der stochastische Prozeß X besteht aus der „Menge“ der Zufallsvariablen X_i , $-\infty \leq i \leq +\infty$. Man bildet nun Paare, z. B. X_t und X_s , X_{t+1} und X_{s+1} , aber auch X_t und X_{t+1} usw. und bestimmt jeweils die zugehörige Kovarianz, sie sei mit $C(t, s)$, $C(t+1, s+1)$, $C(t, t+1)$ bezeichnet. Alle diese $C(i, j)$, $-\infty \leq i, j \leq +\infty$ bilden dann die Kovarianzfunktion. Der Begriff Menge wurde aus mathematischen – hier nicht zu erläuternden – Gründen in Anführungszeichen gesetzt. Die unterschiedlichen Analyseverfahren unterscheiden sich nur darin, welche Aspekte der Kovarianzfunktion sie in den Vordergrund stellen und welche zusätzlichen Annahmen sie für den stochastischen Prozeß treffen.

Für alle der Spektralanalyse zuzurechnenden Verfahren sind die nachfolgenden Dinge grundlegend. Der stochastische Prozeß ist schwach stationär. Diese mathematische Forderung bedeutet für den Anwender, daß sein reales System so beschaffen sein muß, daß der Zusammenhang zwischen der Prozeßvariablen zu verschiedenen Zeitpunkten (z. B. Absatz des Artikels zum Zeitpunkt t_1 und zum Zeitpunkt t_2) nur von der zeitlichen Differenz $\tau = t_2 - t_1$, nicht aber von der absoluten Lage auf der Zeitachse abhängig ist. Das bedeutet, daß die Kovarianzen $C(t, s)$ und $C(t+s, s+1)$ gleich sind, da die Zeitpunkte t und s die gleiche Differenz $s - t$ zueinander haben wie $t+1$ und $s+1$. Genauso ist $C(t, t+1) = C(s, s+1) = C(1)$.

Für die Kovarianzfunktion $C(\tau)$ eines schwach stationären Prozesses X_t gilt dann – wobei weitere Annahmen und mathematische Details zugunsten der Grundlinie der Argumentation hier und weiterhin unterdrückt werden – die Darstellung

$$C(\tau) = \int_{-\pi}^{\pi} e^{i\tau\omega} f(\omega) d\omega. \quad (11)$$

Die Funktion $f(\omega)$ heißt Spektraldichte – oft kurz als Spektrum bezeichnet. i steht hier für $\sqrt{-1}$. Für den Prozeß X_t selbst ergibt sich aus (11) die Repräsentation

$$X_t = \int_{-\pi}^{\pi} e^{it\omega} dZ(\omega). \quad (12)$$

Zwischen den Größen $f(\omega)d\omega$ und $dZ(\omega)$ lassen sich Beziehungen aufstellen. Weiterhin gilt

$$f(\omega) = \sum_{\tau=-\infty}^{-\infty} e^{-i\tau\omega} C(\tau). \quad (13)$$

Die Formeln (11) bis (13) sind mathematisch ableitbare Konsequenzen der Forderung, der stochastische Prozeß sei schwach stationär. Mathematisch gesehen charakterisieren diese Beziehungen einen schwach stationären Prozeß.

Für den Anwender des Modells „stochastischer Prozeß“ ergibt sich aber aus den obigen Beziehungen eine Reihe von Möglichkeiten. Wie ein Blick auf die Formeln (11) und (13) zeigt, besteht ein enger Zusammenhang zwischen der Kovarianzfunktion und dem Spektrum – mathematisch gesprochen: sie bilden ein sogenanntes Paar von Fouriertransformierten. Das bedeutet zwar, daß das Spektrum nicht mehr Information über den Prozeß enthält, als schon in der Kovarianzfunktion enthalten ist. Es zeigt sich aber, daß die Informationen im Spektrum anders aufbereitet sind, so daß sie in vielen Fällen beim Übergang von der Kovarianzfunktion zum Spektrum leichter verarbeitbar werden.

Verfahren der Zeitreihenanalyse sind fast immer Manipulationen an den Zeitreihendaten. Das Spektrum ist ein empfindlicher Indikator zur Feststellung von Veränderungen in der Prozeßstruktur, die durch solche Manipulationen hervorgerufen werden können. Es sei hier nur an den bekannten Slutzky-Effekt erinnert.

Die Darstellung (12) für den stochastischen Prozeß eröffnet – obwohl zuerst eine rein mathematische Konsequenz der Beziehung (11) – durch geeignete Interpretation den Weg für eine Vielzahl von Verfahren. Man kann (12) interpretieren als eine Zerlegung des Prozesses X_t in Komponenten der Form $e^{it\omega} dZ(\omega)$. Diese Komponenten bezeichnet man auch als Schwingungen mit der (Kreis-)Frequenz ω , wobei Phase und Amplitude Zufallsvariable sind. Die Zerlegung (12) des Prozesses X_t in Komponenten unterscheidet sich grundlegend von den bekannten Ansätzen, eine Zeitreihe als Summe eines Trends T_t , einer zyklischen Komponente Z_t (z. B. Konjunktur), eines Saisonanteils S_t und eines unerklärten Restes u_t aufzufassen, d. h.

$$x_t = T_t + Z_t + S_t + u_t. \quad (14)$$

Die Größen $e^{it\omega} dZ(\omega)$ sind mathematische Konstrukte. Sie bieten aber die Möglichkeit, einen Versuch – wie z. B. durch die Formel (14) ausgedrückt – zu verwirklichen. Liegt doch das Problem bei dem klassischen Ansatz (14) darin, daß man nicht in der Lage ist, die einzelnen Komponenten unabhängig voneinander zu definieren – man kann dies sehr schön in dem Buch von Davis² nachlesen. Durch geeignete Zusammenfassung der Komponenten $e^{it\omega} dZ(\omega)$ ist es jetzt möglich, in objektiver Weise eine Zeitreihe in Größen wie kurz-, mittel- oder langfristige Komponenten u. ä. zu zerlegen.

Auch auf die Beziehungen zwischen zwei und mehr stochastischen Prozessen läßt sich das hier nur kurz vorgestellte Konzept des Spektrums ausdehnen. Es ergeben sich eine Reihe von diese Beziehungen charakterisierenden Größen wie Phase, Kohärenz usw., die eine detaillierte Untersuchung des Zusammenhangs zweier Prozesse möglich machen. So kann man z. B. ergründen, ob die Entwicklung zweier Prozesse kurzfristig in der gleichen Weise gekoppelt ist wie mittel- und langfristig.

Auf nähere Einzelheiten kann hier aus Platzgründen nicht eingegangen werden. Auch die eigentlichen auf spektralanalytischen Überlegungen beruhenden Prognoseverfahren können aus dem gleichen Grund nicht vorgestellt werden. Kurz muß aber unbedingt noch darauf hingewiesen werden, daß sich trotz des mathematisch geschlossenen Theoriegebäudes, aus dem die spektralanalytischen Verfahren kommen, bei der Anwendung große numerische und statistische Probleme ergeben können. Es erscheint daher bei der Verwendung entsprechender Methoden angebracht, einen Experten heranzuziehen und nicht nur Zeitreihen durch die entsprechenden Computerprogramme zu schicken.

Als Einführung in die hier angeführte Theorie der schwach stationären Prozesse sei auf die Literatur^{3, 5, 9} verwiesen.

5. Prognosemodelle

Allen bisher vorgestellten Prognoseverfahren ist eigen, daß sie im Grunde abgetrennt von dem Prozeß der Verarbeitung der Prognose, d. h. dem Ziehen von Konsequenzen, den nachfolgenden Entscheidungen und auch abgetrennt von den Entscheidenden ablaufen. Dies mag noch angebracht sein bei den sich zyklisch wiederholenden Prognosen zur Steuerung eines Lagers; werden aber die Prognosen für Entschei-

dungen im strategischen und taktischen Bereich benötigt, dann kann das isolierte Vorgehen Nachteile haben.

Im wesentlichen sind alle hier vorgestellten Verfahren auf längerfristige Erscheinungen ausgerichtet, auch wenn sie – wie z. B. die Methode des Exponential Smoothing – ausdrücklich als kurzfristiges Prognoseverfahren klassifiziert werden. Denn den Verfahren ist gemeinsam, daß sie kurz andauernde Vorgänge – wie z. B. Sonderaktionen im Absatzbereich – so gut wie gar nicht oder nur verzerrt berücksichtigen können.

Eine adäquate Antwort in dieser Situation können Prognosemodelle sein. Diese Modelle gestatten nicht nur durch die Bereitstellung einer Vielzahl von Verfahren und betriebswirtschaftlicher Komponenten (z. B. Budgetbeschränkungen, Lagerkapazitäten, Marketingaktivitäten usw.) eine zutreffendere Beschreibung und Analyse der konkreten Absatzsituation, sondern durch sie können – insbesondere wenn sie in dialogfähiger Form auf einer Rechenanlage implementiert wurden – die Entscheidenden in den Prognoseprozeß integriert werden. Einen Schritt in diese Richtung stellen die bei Lewandowski⁶ beschriebenen Systeme FORSYS und MAVIS dar.

Eine zusammenfassende Darstellung von Ansätzen zu Prognosemodellen, die den obigen Vorstellungen entsprechen, würde den Umfang dieser Arbeit sprengen. Es erscheint auch viel sinnvoller, statt einer Klassifikation „technischer“ Einzelheiten einige Ausführungen mehr grundsätzlicher Art über Modelle und ihre Anwender zu machen.

Die folgenden Ausführungen beruhen auf einem 1970 erschienenen Artikel von John D. C. Little⁷.

Er führt aus: “The big problem with management science models is that managers practically never use them.” Die Meinung, daß das Problem nicht darin liegt, solche Modelle zu entwerfen, sondern sie zum praktischen Einsatz zu bringen, dürfte man auch bezüglich des deutschsprachigen Raumes teilen.

Little untersucht die Gründe für diese Situation und entwickelt aus der Analyse ein Konzept für eine Modellphilosophie: “If we want a manager to use a model, we should make it his, an extension of his ability to think about and analyze his operation. This puts special requirements on design and will often produce something rather different from what a management scientist might otherwise build. . . . A decision calculus will be defined as a model-based set of procedures for processing data and judgements to assist a manager in his decision making.” Und: “A model that is to be used by a manager should be simple, robust,

easy to control, adaptive, as complete as possible, and easy to communicate with.”

Little illustriert seine Ausführungen an einem umfangreichen Beispiel aus dem Werbebereich. Dadurch läuft er allerdings Gefahr, daß seine Überlegungen als nur für diesen Bereich gültig angesehen werden. Vielleicht ist dies ein Grund dafür, daß der Aufsatz bisher relativ wenig Einfluß auf andere Arbeiten gefunden hat.

Natürlich sind die Adjektiva „simple, robust, usw.“ in jedem Anwendungsbereich anders zu präzisieren. Insbesondere der Begriff „simple“ bedarf sorgfältiger Behandlung. So mancher mathematische Gegenstand, den wir heute als einfach genug ansehen, um ihn schon im Wege der neuen Mathematik den Kindern in der Grundschule zuzumuten, galt bei seiner „Erfindung“ als alles andere als einfach; ein Blick in alte Bücher bringt da manch überraschende Entdeckung. Es gilt also, den oben zitierten Satz nicht absolut, sondern *relativ* zu nehmen.

Auf zwei Komponenten der Little'schen Modellphilosophie sei hier noch kurz eingegangen. Im Zusammenhang mit der Diskussion des Begriffs „completeness“ führt er aus: “An important aid to completeness is the incorporation of subjective judgement. . . . at any given point in time, subjective estimates will be valuable for quantities that are currently difficult to measure or which cannot be measured in the time available before a decision must be made.” Begriffe wie „objektiv“ und „intersubjektiv nachprüfbar“ haben – auch in der Ausbildung – den auf ein falsches Verständnis zurückgehenden Effekt gehabt, daß alles Subjektive und Nicht-Quantitative bei „Modellbauern“ verpönt ist. Wenn nur deutlich ist, an welcher Stelle diese Dinge in das Modell eingehen, dann können sie durchaus einen wesentlichen Beitrag zur Erklärung und Beherrschung einer nicht nur deterministisch-quantitativen ökonomischen Realität bilden.

Den vielleicht entscheidenden Punkt bei der Frage nach der Anwendung (nicht der Anwendbarkeit) von Modellen spricht Little bei seinen Ausführungen über „easy to communicate with“ an. “Every effort should be made to express input requests in operational terms. The internal parameterization of the model can be anything, but the requests to the user for data should be in his language. Thus, coefficients and constants without clear operational interpretation are to be discouraged. Let them be inferred by the computer from inputs that are easier for the user to work with.”

Diese Ausführungen dürfen nicht als ein Rezept verstanden werden, mit dessen Hilfe – gleichsam wie mit einer reißerischen Verpackung ein

Produkt – das Modell an den Anwender gebracht werden kann. Sie sind Ausdruck einer ganz bestimmten Haltung. Erst wenn ich auch in der Lage bin, mich im Sprachraum einer anderen, an der Problemlösung beteiligten Person auszudrücken, kann ich sicher sein, das Problem verstanden und mit gewisser Wahrscheinlichkeit auch etwas Sinnvolles zu seiner Lösung beigetragen zu haben. Die oft anzutreffende Betonung der mathematischen Struktur und der Nützlichkeit von Abstraktionen ist vielfach nichts anderes als ein künstliches, mathematisch-technisches *l'art pour l'art*.

Folgt man den Ausführungen Little's, so wird man bald feststellen, daß „good on-line models are user-instructing and introduce a person the issues of the problem and the model much faster than would otherwise be possible“. Der Einsatz von Modellen bringt schon bald die Erkenntnis, daß man mehr über die Entscheidungssituation wissen sollte – oder wie Little sagt: „One of the most evident consequences of the experience to date has been that a model is a stone in the shoe for better data.“

Dieser Abschnitt wollte nicht über bestehende Prognosemodelle referieren, sondern die Aufmerksamkeit auf die mehrfach zitierte Arbeit von Little lenken¹⁰. Erst wenn sich alle Beteiligten über die dort angerissenen Grundfragen des Modelleinsatzes einig sind, ist es sinnvoll, sich den „technischen“ Arbeiten der Modellkonstruktion hinzugeben. Andernfalls wird das Ergebnis vom praktischen Problemlösungsstandpunkt mit hoher Wahrscheinlichkeit unbefriedigend sein.

Anmerkungen

- 1 Brown, R. G.: Smoothing, Forecasting and Prediction of Discrete Time Series, Englewood Cliffs, N. J., 1963.
- 2 Davis, H. T.: The Analysis of Economic Time Series. San Antonio 1963.
- 3 Granger, C. W. J., u. M. Hatanaka: Spectral Analysis of Economic Time Series. Princeton 1964.
- 4 Knuth, D. E.: The Art of Programming Vol. 1. 2nd Edition Reading, Mass., 1973.
- 5 König, H., u. J. Wolters: Einführung in die Spektralanalyse ökonomischer Zeitreihen. Meisenheim am Glan 1972.
- 6 Lewandowski, R.: Prognose- und Informationssysteme und ihre Anwendungen Bd. 1. Berlin, New York 1974.
- 7 Little, J. D. C.: Models and Managers: The Concept of a Decision Calculus. In: Management Science 16 (1970), No. 8, S. B-466ff.
- 8 Mertens, P. (Herausgeber): Prognoserechnung. Würzburg 1973.
- 9 Naeve, P.: Spektralanalytische Methoden zur Analyse ökonomischer Zeitreihen. Würzburg 1969.
- 10 Vgl. auch die Übersetzung dieses Aufsatzes in dem vorliegenden Buch: „Modelle und Manager: das Konzept eines Decision Calculus“.