

ANWENDUNG DER "STOCHASTISCHEN PROZESSE" ZUR PROGNOSE?

P. Naeve, Bielefeld

Zusammenfassung: Die klassische Zeitreihenanalyse räumt in Zeitreihenmodellen dem Zufall nur die Rolle des unerklärten Restes ein. Neuere Ansätze wenden sich der Modellierung dieses "Restes" zu. Die Theorie der schwach stationären Prozesse liefert die Grundlage der zahlreichen in der Literatur diskutierten Zeitreihenmodelle. Eine Diskussion der grundlegenden Annahmen dieser Theorie führt zu einem größeren Verständnis der Möglichkeiten und der Grenzen der Anwendung stochastischer Prozesse bei Prognoseproblemen. Außerdem wird eine bisher wenig diskutierte Klasse von Zeitreihenmodellen vorgestellt.

Summary: In a classical time series analyst's point of view randomness just plays the part of the unexplained rest when modeling time series. A new approach in time series analysis is concerned with building models for this residues. The numerous time series models discussed in the literature are based on the theory of weak stationary processes. Discussing the basical assumptions of this theory may lead to a deeper understanding of the opportunities and limits when applying stochastic processes for forecasting purposes. A new class of time series models is introduced which has not been studied very much yet.

1. Einleitung

In dieser Arbeit soll der Versuch gemacht werden, den Beitrag aufzuzeigen, den die Theorie der stochastischen Prozesse zur Zeitreihenanalyse leisten kann und könnte. Die Ausführungen beschränken sich dabei auf die sogenannte univariate Zeitreihenanalyse und beschäftigen sich lediglich mit dem Vorhersageproblem.

Es wird also die folgende Situation betrachtet:

Gegeben: Die Zeitreihe x_1, x_2, \dots, x_T

Gesucht: Die Prognose \hat{x}_{T+1} zum Zeitpunkt T für den Zeitpunkt T+1 mit $l \geq 1$

Gewünscht: \hat{x}_{T+1} soll möglichst gut x_{T+1} "treffen".

Univariate Zeitreihenanalyse bedeutet, daß keine weiteren Daten (Zeitreihen) explizit zur Lösung dieser Aufgabe herangezogen werden als die vorliegenden Werte x_1, \dots, x_T . Dies schließt aber nicht aus, daß andere Daten (Informationen über die "Entstehung" der vorliegenden Zeitreihe) implizit durchaus zur Lösung herangezogen werden können. Der gegen die univariate Zeitreihenanalyse oft erhobene Vorwurf des theorielosen "data mining" ist also nicht unvermeidbar. Ein sehr schönes Beispiel univariater Zeitreihenanalyse im eben erwähnten Sinne findet sich in einer Untersuchung von LENZ (1974) zur Kassenhaltung bei Banken.

Im Folgenden soll zuerst aufgezeigt werden, wo und wie die Theorie der stochastischen Prozesse in die Zeitreihenanalyse eingebracht wird. Es zeigt sich, daß die klassische Zeitreihenanalyse stark von Gedankengängen der Mechanik beeinflusst ist. Erst eine Neubetrachtung des "unerklärten Restes" öffnet einer Anwendung der Theorie stochastischer Prozesse die Tür. Dann wird gezeigt werden, daß die oft verwirrende Vielzahl von expliziten stochastischen Prozessen, die als Zeitreihenmodelle angeboten werden, sich auf das Konzept der Stationarität und die "Lösung" eines Schätzproblems zurückführen lassen. Danach sollen einige Probleme diskutiert werden, die sich aus den grundlegenden theoretischen Annahmen bei der Übertragung auf praktische Probleme ergeben. Diese Arbeit will nicht so verstanden sein, daß die Theorie der stochastischen Prozesse die Lösung des Prognoseproblems bringt. Stochastische Prozesse können aber sicher einen Beitrag zur Lösung leisten. Dogmatisches Beharren auf der Methode hat in der Geschichte noch nie zu langfristigen Erkenntnissen geführt.

Ein weiteres Anliegen dieser Arbeit ist erfüllt, wenn der Praktiker nicht mehr Methoden, die auf der Theorie stochastischer Prozesse fußen, unspezifisch wegen ihres zugegeben oft anspruchsvolleren mathematischen Gehalts ablehnt son-

dern gezielt durch Diskussion der grundlegenden Konzepte wie z.B. Stationarität, Linearität o.ä. nachzuweisen versucht, daß für sein konkretes Problem kein wesentlicher Lösungsbeitrag aus dieser Richtung zu erwarten ist.

2. Klassische Zeitreihenanalyse - oder gute Beispiele verderben die Sitten.

Die Arbeiten des Triumvirates Tycho de Brahe, Johannes Kepler und Isaac Newton zur Erklärung der Planetenbewegung: Brahe lieferte die Zeitreihen, Kepler ein Modell und Newton die erklärende Theorie - haben bewußt und unbewußt die Bemühungen um die Analyse ökonomischer Zeitreihen in ihren Anfängen beeinflußt. Nimmt man das 1941 erschienene Buch zur Zeitreihenanalyse von DAVIS (1941) so fällt ein der Mechanik entstammendes Vokabular auf. Begriffe wie Trägheit, Moment, Energie wurden dort im Zusammenhang mit Zeitreihen verwandt. Die klassischen Ansätze der Zeitreihenanalyse gehen im Prinzip immer von einer Darstellung

$$x_t = T_t \circ Z_t \circ S_t \circ R_t$$

aus, dabei wird \circ ersetzt durch

$$\circ \hat{=} \begin{cases} + \hat{=} \text{additives Modell} \\ \times \hat{=} \text{multiplikatives Modell} \\ + \times \hat{=} \text{gemischte Modelle} \end{cases}$$

Die Komponenten bedeuten

$$T_t \hat{=} \text{Trend (Trägheit der Zeitreihe)}$$

$$Z_t \hat{=} \text{zyklische Komponente}$$

$$S_t \hat{=} \text{Saison}$$

$$R_t \hat{=} \text{unerklärter Rest}$$

Diese Darstellung wird damit begründet, daß eben anders als im Falle des Planetensystems mit dem alles fast erdrückenden Einfluß der Sonne im ökonomischen System viele verschiedene

Kräfte am Wirken sind. Die einzelnen Komponenten finden eine zwischen reiner Beschreibung und substanzwissenschaftlicher Deduktion liegende Begründung. Oft wird die eine als der Rest bei der Bestimmung der anderen definiert.

Die Denkweise der klassischen Zeitreihenanalyse läßt sich auch gut an einer ihrer Antworten auf das Prognoseproblem erkennen, nämlich dem Verfahren des "Exponentiellen Glättens" (exponential smoothing).

Bezeichnet man die Prognose zum Zeitpunkt t für den Zeitpunkt $t+1$ mit $\hat{x}_t(1)$ dann lautet die bekannte Prognoseformel

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

Dies kann man umformen zu

$$\hat{x}_t(1) = \hat{x}_{t-1}(1) + \alpha(\hat{x}_{t-1}(1) - x_{t-1})$$

also: neue Prognose = alte Prognose verändert um einen Bruchteil der Differenz zwischen alter Prognose und beobachtetem Zeitreihenwert.

Das Auftreten von $\hat{x}_{t-1}(1)$ - Beibehalten der alten Prognose - entspricht der Vorstellung von Trägheit der Zeitreihen, der zweite Teil berücksichtigt Änderungen - aber keine drastischen, da in praktischen Anwendungen mit α in der Nähe von Null gearbeitet wird.

Es geht hier nicht um pro und contra exponential smoothing (mir ist auch bekannt, daß man vieles mit exponential smoothing in der Praxis machen kann und daß man auch daran viel interessante Mathematik (z-Transformation) betreiben kann) sondern es war nur als ein Beispiel für die Denkweise der klassischen Zeitreihenanalytiker gedacht.

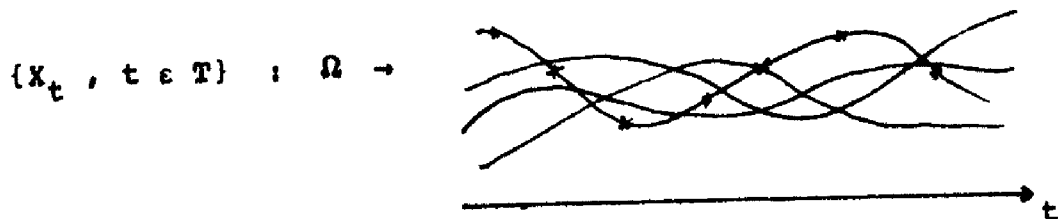
3. Stochastische Prozesse - oder der Zufall überdacht.

In der klassischen Zeitreihenanalyse taucht die Größe R_t eigentlich immer in der Form des unerklärten Restes auf - manchmal werden auch einige Annahmen gemacht wie $E(R_t) = 0$ oder $E(R_t R_s) = \sigma^2 \delta_{ts}$. Wenn man heute Zeitreihenanalyse sagt, dann geht man - zumindest an einer Universität - davon aus, daß eine Zeitreihe die Realisation eines stochastischen Prozesses ist.

Was heißt das ?

In der Literatur findet sich die folgende Definition (DOOB (1953)) "We define a stochastic process as any family of random variable $\{X_t, t \in T\}$."

Bei einem stochastischen Prozeß wird also jedem Ausgang eines Zufallsexperimentes nicht eine reelle Zahl - wie bei einer Zufallsvariablen - sondern eine Funktion von t zugeordnet. Anschaulich stellt sich die Situation wie in der folgenden Graphik dar.



Heranziehen der Theorie der stochastischen Prozesse zur Zeitreihenanalyse bedeutet also, daß die vorliegende Zeitreihe (~~ist~~) als eine aus einer Menge möglicher Realisationen angesehen wird. Der stochastische Prozeß steht für das Ensemble aller möglichen Realisationen.

Für viele Anwender mag in diesem Verlust der "Individualität" ihrer Zeitreihe schon eine erste Grenze der Anwendung der Theorie der stochastischen Prozesse liegen.

Die oben angeführte Definition ist recht weit. Man sucht daher nach Größen, mit deren Hilfe man Prozesse beschreiben und klassifizieren kann. Eine große Klasse von Prozessen stellen die sogenannten schwach stationären Prozesse dar. Ein schwach stationärer Prozeß ist dadurch charakterisiert, daß

$$E(X_t) = m \quad \text{für alle } t$$

$$E((X_t - m)(X_s - m)) = \gamma_\tau \quad \text{mit } \tau = t - s \quad \text{für alle } t, s$$

gilt.

Schwach stationäre Prozesse sind in der Literatur sehr ausführlich untersucht und seit den Arbeiten von WIENER (1949) und KOLMOGOROFF (1941) weiß man auch, wie man optimale (im Sinne minimaler quadratischer Fehler) Prognosen zu bilden hat.

Im Folgenden gehören alle betrachteten stationären Prozesse zur Klasse der schwach stationären Prozesse.

4. Modelle für schwach stationäre Prozesse

Ausgangspunkt für unsere Überlegungen ist die folgende Darstellung eines allgemeinen linearen stationären Prozesses:

$$(\star) \quad X_t = \sum_{i=0}^{\infty} g_i \varepsilon_{t-i}$$

$$\text{bzw. } \varepsilon_t = \sum_{i=0}^{\infty} h_i X_{t-i},$$

dabei ist ε_t ein unkorrelierter Prozeß mit

$$E(\varepsilon_t) = 0 \quad \text{und} \quad E(\varepsilon_t \varepsilon_s) = \sigma_\varepsilon^2 \delta_{ts} \quad .$$

Die Rechtfertigung für diese Darstellung beruht auf einem Satz von WOLD (1938), der besagt, daß sich jeder schwach stationäre Prozeß X_t als Summe von zwei Komponenten darstellen läßt, wobei die eine Komponente deterministisch ist, d.h. bei Kenntnis der Vergangenheit $(-\infty, T)$ ohne Fehler vorhergesagt werden kann, während die zweite Komponente gerade die Darstellung (*) hat. Weiterhin betrachten wir den allgemeinen nicht linearen Prozeß:

$$(\star\star) \quad X_t = \sum_{i=0}^{\infty} g_t \varepsilon_{t-1} + \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} g_{ij} \varepsilon_{t-i} \varepsilon_{t-j} \\ + \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} g_{ijk} \varepsilon_{t-i} \varepsilon_{t-j} \varepsilon_{t-k} \dots$$

$$\text{bzw.} \quad \varepsilon_t = \sum_{i=0}^{\infty} h_i X_{t-i} + \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} h_{ij} X_{t-i} X_{t-j} \\ + \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} h_{ijk} X_{t-i} X_{t-j} X_{t-k} + \dots$$

wobei ε_t jetzt unabhängig identisch verteilte Zufallsvariablen sind. Die Darstellung ($\star\star$) geht auf N.Wiener zurück. Sie wird auch als Volterra Entwicklung bezeichnet.

Aus den Darstellungen (*) und ($\star\star$) ist sofort zu erkennen, daß sie für praktische Anwendungen so nicht geeignet sind, da die Kenntnis (oder die Erlangung der Kenntnis - sprich Schätzung) von unendlich vielen Koeffizienten g_i , g_{ij} usw. nötig ist.

Man sucht nun nach möglichst einfachen (Prinzip der parsimony) Parametrisierungen - Beschreibungen der Sachverhalte (*) bzw. ($\star\star$) mit endlicher Anzahl von Parametern.

Die Antwort für lineare Modelle in der Form des AR-, MA- oder ARMA-Prozeß ist seit BOX und JENKINS (1970) wohl allgemein bekannt.

$$\text{AR}(p) : X_t - \Phi_1 X_{t-1} - \dots - \Phi_p X_{t-p} = \varepsilon_t$$

$$MA(q) \quad X_t = \epsilon_t + \theta_1 \epsilon_{t-1} + \dots + \theta_q \epsilon_{t-q}$$

$$ARMA(p,q) \quad X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = \epsilon_t + \theta_1 \epsilon_{t-1} + \dots + \theta_q \epsilon_{t-q}$$

Einige Beispiele sollen zeigen, daß die "Box-Jenkins Modelle" immer auf die "Formen" von (*) zurückgeführt werden können.

Beispiel 1: AR-Prozeß

$$X_t = a X_{t-1} + \epsilon_t$$

$$X_{t-1} = a X_{t-2} + \epsilon_{t-1}$$

$$\text{Einsetzen ergibt:} \quad X_t = \epsilon_t + a \epsilon_{t-1} + a^2 X_{t-2}$$

$$X_{t-2} = a X_{t-3} + \epsilon_{t-2}$$

$$\text{Einsetzen ergibt:} \quad X_t = \epsilon_t + a \epsilon_{t-1} + a^2 \epsilon_{t-2} + a^3 X_{t-3}$$

Nach wiederholter Durchführung dieses Einsetzungsvorganges ergibt sich:

$$X_t = \sum_{i=0}^{\infty} a^i \epsilon_{t-i} = \sum_{i=0}^{\infty} g_i \epsilon_{t-i} \quad \text{mit} \quad g_i = a^i$$

$$\epsilon_t = X_t - a X_{t-1} = \sum_{i=0}^{\infty} h_i X_{t-1} \quad \text{mit} \quad h_0 = 1 \quad h_1 = a \quad h_i \equiv 0 \quad i > 2$$

Beispiel 2: MA-Prozeß

$$X_t = \epsilon_t + b \epsilon_{t-1}$$

$$X_{t-1} = \epsilon_{t-1} + b \epsilon_{t-2} \quad \Rightarrow \quad \epsilon_{t-1} = X_{t-1} - b \epsilon_{t-2}$$

$$X_t = \epsilon_t + b X_{t-1} - b^2 \epsilon_{t-2}$$

$$X_{t-2} = \epsilon_{t-2} + b \epsilon_{t-3} \quad \Rightarrow \quad \epsilon_{t-2} = X_{t-2} - b \epsilon_{t-3}$$

$$X_t = \epsilon_t + b X_{t-1} - b^2 X_{t-2} + b^3 \epsilon_{t-3}$$

Durch fortgesetzte Wiederholung dieses Ersetzungsprozesses erhält man

$$X_t = \epsilon_t + \sum_{i=1}^{\infty} (-1)^{i+1} b^i X_{t-i} \quad \Rightarrow \quad \epsilon_t = X_t - \sum_{i=1}^{\infty} (-1)^{i+1} b^i X_{t-i}$$

also die Darstellung:

$$X_t = \epsilon_t + b \epsilon_{t-1} = \sum_{i=0}^{\infty} g_i \epsilon_{t-i} \quad \text{mit} \quad g_0 = 1 \quad g_1 = b \quad g_i \equiv 0 \quad i > 2$$

$$\epsilon_t = X_t - \sum_{i=1}^{\infty} (-1)^{i+1} b^i X_{t-i} = \sum_{i=0}^{\infty} (-b)^i X_{t-i} = \sum_{i=0}^{\infty} h_i X_{t-i} \quad \text{mit} \quad h_i \equiv (-b)^i$$

Beispiel 3: ARMA-Prozess

$$X_t - a X_{t-1} = \varepsilon_t + b \varepsilon_{t-1}$$

$$X_t = a X_{t-1} + \varepsilon_t + b \varepsilon_{t-1}$$

$$X_{t-1} = a X_{t-2} + \varepsilon_{t-1} + b \varepsilon_{t-2}$$

$$\text{Einsetzen ergibt: } X_t = \varepsilon_t + b \varepsilon_{t-1} + a \varepsilon_{t-1} + ab \varepsilon_{t-2} + a^2 X_{t-2}$$

$$X_{t-1} = a X_{t-2} + \varepsilon_{t-2} + b \varepsilon_{t-3}$$

$$\text{also dann: } X_t = \varepsilon_t + (b+a)\varepsilon_{t-1} + a(b+a)\varepsilon_{t-2} + a^2 b \varepsilon_{t-3} + a^3 X_{t-3} .$$

Nach fortgesetzter Wiederholung dieses Vorganges ergibt sich:

$$X_t = \varepsilon_t + \sum_{i=1}^{\infty} (b+a)a^{i-1} \varepsilon_{t-i} = \sum_{i=0}^{\infty} g_i \varepsilon_{t-i} \quad \text{mit } g_0 = 1 \quad g_1 = (b+a)a^{i-1}$$

bzw.

$$\varepsilon_t = X_t - a X_{t-1} - b \varepsilon_{t-1}$$

$$\varepsilon_{t-1} = X_{t-1} - a X_{t-2} - b \varepsilon_{t-2}$$

$$\text{Einsetzen ergibt: } \varepsilon_t = X_t - a X_{t-1} - b X_{t-1} + ab X_{t-2} + b^2 \varepsilon_{t-2}$$

$$\varepsilon_{t-2} = X_{t-2} - a X_{t-3} - b \varepsilon_{t-3}$$

$$\text{und damit: } \varepsilon_t = X_t - (a+b)X_{t-1} + b(a+b)X_{t-2} - ab^2 X_{t-3} - b^3 \varepsilon_{t-3} .$$

Nach fortgesetzter Wiederholung dieses Vorganges gelangt man zu:

$$\varepsilon_t = X_t + \sum_{i=1}^{\infty} (-1)^i b^{i-1} (a+b) X_{t-i} = \sum_{i=0}^{\infty} h_i X_{t-i}$$

$$\text{mit } h_0 = 1 \quad h_i = (-1)^i b^{i-1} (a+b) \quad i > 1 .$$

Man kann nun zwei Positionen beziehen:

1. Man glaubt es liegt in der Realität ein MA(q) , AR(p) oder ARMA(p,q) vor.
2. Man glaubt nur an das Vorliegen eines schwach stationären Prozesses und nimmt den jeweiligen MA(q) , AR(p) bzw. ARMA(p,q) als Approximation an MA(∞) bzw. AR(∞) .

Wenden wir uns jetzt dem stationären nicht-linearen Modell zu. Bilineare Modelle stellen eine Möglichkeit zur Parametrisierung von nicht linearen Modellen dar. Eine Einführung in diese Prozeßklasse gibt das Buch von GRANGER und ANDERSEN (1978).

$$X_t = \sum_{i=1}^p \phi_i X_{t-i} + \sum_{i=0}^q \theta_i \varepsilon_{t-i} + \sum_{i=1}^Q \sum_{j=1}^P \beta_{ij} \varepsilon_{t-i} X_{t-j}$$

Bemerkungen:

1. Dies ist nicht der allgemeine Fall sondern lediglich der sogenannte homogene Fall.
2. Offensichtlich sind für $\beta_{ij} \equiv 0$ alle ARMA-Modelle enthalten.

Klassifikation:

1. Fehlt der ARMA-Teil so spricht man vom vollständig bilinearen Modell.
2. Diagonalmodell

$$\beta_{ij} = \begin{cases} \beta_i & i = j \\ 0 & \text{sonst} \end{cases}$$

3. Superdiagonalmodell

$$\beta_{ij} = \begin{cases} \beta_{ij} & i \leq j \\ 0 & i > j \end{cases}$$

4. Subdiagonalmodell

$$\beta_{ij} = \begin{cases} 0 & i < j \\ \beta_{ij} & i \geq j \end{cases}$$

Die Namensgebung wird einsichtig, wenn man das Modell als

$$X_t = \varepsilon' B X$$

schreibt mit

$$\epsilon = \begin{pmatrix} \epsilon_{t-1} \\ \epsilon_{t-2} \\ \vdots \\ \epsilon_{t-p} \end{pmatrix} \quad X = \begin{pmatrix} X_{t-1} \\ X_{t-2} \\ \vdots \\ X_{t-Q} \end{pmatrix} \quad B = (\beta_{ij})$$

Einige Beispiele sollen demonstrieren, daß bilineare Modelle auf die Form (***) zurückzuführen sind.

Beispiel 4: Diagonal-Modell

$$X_t = \epsilon_t + a \epsilon_{t-1} X_{t-1}$$

$$X_{t-1} = \epsilon_{t-1} + a \epsilon_{t-2} X_{t-2}$$

$$\text{Einsetzen ergibt: } X_t = \epsilon_t + a \epsilon_{t-1} + a^2 \epsilon_{t-1} \epsilon_{t-2} X_{t-2}$$

$$X_{t-2} = \epsilon_{t-2} + a \epsilon_{t-3} X_{t-3}$$

$$\text{Einsetzen ergibt: } X_t = \epsilon_t + a \epsilon_{t-1} \epsilon_{t-2} + a^2 \epsilon_{t-1} \epsilon_{t-2} \epsilon_{t-2} + a^2 X_{t-2}$$

Nach fortgesetzter Durchführung dieses Ersetzungsprozesses ergibt sich:

$$X_t = \epsilon_t + a \epsilon_{t-1} \epsilon_{t-1} + a^2 \epsilon_{t-1} \epsilon_{t-2} \epsilon_{t-2} + a^3 \epsilon_{t-1} \epsilon_{t-2} \epsilon_{t-2} \epsilon_{t-3} + \dots$$

$$= \sum_{j=0}^{\infty} a^j \epsilon_{t-j} \prod_{i=1}^j \epsilon_{t-i}$$

$$= \sum_{i=0}^{\infty} g_i \epsilon_{t-i} + \sum_{i=1}^{\infty} \sum_{j=0}^{\infty} g_{ij} \epsilon_{t-i} \epsilon_{t-j} + \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} g_{ijk} \epsilon_{t-i} \epsilon_{t-j} \epsilon_{t-k} \dots$$

$$g_i = \begin{cases} 1 & i=0 \\ 0 & \text{sonst} \end{cases} \quad g_{ij} = \begin{cases} a & i=j=1 \\ 0 & \text{sonst} \end{cases} \quad g_{ijk} = \begin{cases} a^2 & i=1 \quad j=2 \quad k=2 \\ 0 & \text{sonst} \end{cases}$$

analog ergibt sich für ϵ_t :

$$\epsilon_t = X_t - a \epsilon_{t-1} X_{t-1}$$

$$\epsilon_{t-1} = X_{t-1} - a \epsilon_{t-2} X_{t-2}$$

$$\text{Einsetzen ergibt: } \epsilon_t = X_t - a X_{t-1} X_{t-1} + a^2 \epsilon_{t-2} X_{t-1} X_{t-2}$$

$$\epsilon_{t-2} = X_{t-2} - a \epsilon_{t-2} X_{t-3}$$

Einsetzen ergibt: $\epsilon_t = X_t - a X_{t-1} X_{t-1} + a^2 X_{t-2} X_{t-1} X_{t-2} - a^3 \epsilon_{t-3} X_{t-1} X_{t-2} X_{t-3}$

Fortsetzung dieses Ersetzungsprozesses führt zu:

$$\begin{aligned} \epsilon_t &= \sum_{j=0}^{\infty} (-a)^j X_{t-j} \prod_{i=1}^j X_{t-i} \\ &= \sum_{i=0}^{\infty} h_i X_{t-1} + \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} h_{ij} X_{t-1} X_{t-j} + \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} h_{ijk} X_{t-1} X_{t-j} X_{t-k} \dots \\ h_i &= \begin{cases} 1 & i=0 \\ 0 & \text{sonst} \end{cases} \quad h_{ij} = \begin{cases} -a & i=j=1 \\ 0 & \text{sonst} \end{cases} \quad h_{ijk} = \begin{cases} a^2 & i=1 \quad j=2 \quad k=2 \\ 0 & \text{sonst} \end{cases} \end{aligned}$$

Beispiel 5: Superdiagonal-Modell

$$X_t = \epsilon_t + a X_{t-2} \epsilon_{t-1}$$

$$X_{t-2} = \epsilon_{t-2} + a X_{t-4} \epsilon_{t-3}$$

Einsetzen ergibt: $X_t = \epsilon_t + a \epsilon_{t-1} \epsilon_{t-1} + a^2 X_{t-4} \epsilon_{t-1} \epsilon_{t-3}$

$$X_{t-4} = \epsilon_{t-4} + a X_{t-6} \epsilon_{t-5}$$

Einsetzen ergibt:

$$X_t = \epsilon_t + a \epsilon_{t-1} \epsilon_{t-2} + a^2 \epsilon_{t-1} \epsilon_{t-3} \epsilon_{t-4} + a^3 X_{t-6} \epsilon_{t-1} \epsilon_{t-3} \epsilon_{t-5}$$

Nach fortgesetzter Durchführung dieses Ersetzungsvorganges erhält man:

$$\begin{aligned} X_t &= \sum_{j=0}^{\infty} a^j \epsilon_{t-2j} \prod_{i=1}^j \epsilon_{t-2i+1} \\ &= \sum_{i=0}^{\infty} g_i \epsilon_{t-1} + \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} g_{ij} \epsilon_{t-1} \epsilon_{t-j} + \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} g_{ijk} \epsilon_{t-1} \epsilon_{t-j} \epsilon_{t-k} \dots \\ g_i &= \begin{cases} 1 & i=0 \\ 0 & \text{sonst} \end{cases} \quad g_{ij} = \begin{cases} a & i=1, j=2 \\ 0 & \text{sonst} \end{cases} \quad g_{ijk} = \begin{cases} a^2 & i=1 \quad j=3 \quad k=4 \\ 0 & \text{sonst} \end{cases} \end{aligned}$$

bzw. für die ϵ_t - Komponente:

$$\epsilon_t = X_t - a X_{t-2} \epsilon_{t-1}$$

$$\epsilon_{t-1} = X_{t-1} - a X_{t-3} \epsilon_{t-2}$$

Einsetzen ergibt: $\epsilon_t = X_t - a X_{t-1} X_{t-2} + a^2 X_{t-2} X_{t-3} \epsilon_{t-2}$

$$\epsilon_{t-2} = X_{t-2} - a X_{t-4} \epsilon_{t-3}$$

Einsetzen ergibt: $\epsilon_t = X_t - a X_{t-1} X_{t-2} + a^2 X_{t-2} X_{t-2} X_{t-3} - a^3 X_{t-2} X_{t-3} X_{t-4} \epsilon_{t-3}$

Fortgesetzte Durchführung dieses Ersetzungsvorganges führt zu:

$$\begin{aligned} \varepsilon_t &= \sum_{j=0}^{\infty} (-a)^j X_{t-j} \prod_{i=1}^j X_{t-1-i} \\ &= \sum_{i=0}^{\infty} h_i X_{t-i} + \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} h_{ij} X_{t-i} X_{t-j} + \dots \\ h_i &= \begin{cases} 1 & i=0 \\ 0 & \text{sonst} \end{cases} \quad h_{ij} = \begin{cases} -a & i=1 \quad j=2 \\ 0 & \text{sonst} \end{cases} \end{aligned}$$

Beispiel 6: Subdiagonal Modell

$$X_t = \varepsilon_t + a X_{t-1} \varepsilon_{t-2}$$

$$X_{t-1} = \varepsilon_{t-1} + a X_{t-2} \varepsilon_{t-3}$$

$$\text{Einsetzen ergibt: } X_t = \varepsilon_t + a \varepsilon_{t-1} \varepsilon_{t-2} + a^2 X_{t-1} \varepsilon_{t-2} \varepsilon_{t-3}$$

$$X_{t-2} = \varepsilon_{t-2} + a X_{t-3} \varepsilon_{t-4}$$

$$\text{Einsetzen ergibt: } X_t = \varepsilon_t + a \varepsilon_{t-1} \varepsilon_{t-2} + a^2 \varepsilon_{t-2} \varepsilon_{t-2} \varepsilon_{t-3} + a^3 X_{t-3} \varepsilon_{t-2} \varepsilon_{t-3} \varepsilon_{t-4}$$

usw. man erkennt, daß eine ähnliche Struktur wie in den vorstehenden Beispielen entsteht. g_{ijk} usw. sind für genau eine Indexkombination $\neq 0$.

Beispiel 7: Der einfache bilineare Prozeß

$$X_t = a X_{t-1} + \varepsilon_t + b X_{t-1} \varepsilon_{t-1}$$

führt schon zu einer bedeutend komplizierteren nicht linearen Struktur, wie man schon nach wenigen Schritten des Einsetzungsprozesses erkennt

$$X_{t-1} = a X_{t-2} + \varepsilon_{t-1} + b X_{t-2} \varepsilon_{t-2}$$

$$\begin{aligned} \text{Einsetzen ergibt: } X_t &= \varepsilon_t + a \varepsilon_{t-1} + b \varepsilon_{t-1} \varepsilon_{t-1} + ab X_{t-2} \varepsilon_{t-1} \\ &\quad + a^2 X_{t-2} + ab X_{t-2} \varepsilon_{t-2} + b^2 X_{t-2} \varepsilon_{t-1} \varepsilon_{t-2} \end{aligned}$$

Einsetzen von $X_{t-2} = a X_{t-3} + \varepsilon_{t-2} + b X_{t-3} \varepsilon_{t-3}$ ergibt:

$$\begin{aligned} X_t &= \varepsilon_t + a \varepsilon_{t-1} + a^2 \varepsilon_{t-2} + b \varepsilon_{t-1} \varepsilon_{t-1} + ab \varepsilon_{t-1} \varepsilon_{t-2} \\ &\quad + ab \varepsilon_{t-2} \varepsilon_{t-2} + b^2 \varepsilon_{t-1} \varepsilon_{t-2} \varepsilon_{t-2} + a^2 X_{t-3} + a^2 b X_{t-3} \varepsilon_{t-2} \\ &\quad + ab^2 X_{t-3} \varepsilon_{t-1} \varepsilon_{t-2} + ab X_{t-3} \varepsilon_{t-1} + a^2 b X_{t-3} \varepsilon_{t-3} \\ &\quad + ab^2 X_{t-3} \varepsilon_{t-1} \varepsilon_{t-3} + ab^2 X_{t-3} \varepsilon_{t-2} \varepsilon_{t-3} + b^4 X_{t-3} \varepsilon_{t-1} \varepsilon_{t-2} \varepsilon_{t-3} \end{aligned}$$

Fortgesetztes Wiederholen führt zu:

$$\begin{aligned} x_t &= \epsilon_t + a \epsilon_{t-1} + a^2 \epsilon_{t-1} + \dots \\ &+ b \epsilon_{t-1} \epsilon_{t-1} + ab \epsilon_{t-2} \epsilon_{t-2} + a^2 b \epsilon_{t-3} \epsilon_{t-3} + \dots \\ &+ ab \epsilon_{t-1} \epsilon_{t-2} + a^2 b \epsilon_{t-2} \epsilon_{t-3} + a^3 b \epsilon_{t-3} \epsilon_{t-4} \dots \quad \text{usw.} \end{aligned}$$

5. Prognoseformeln

Wählt man als Kriterium den minimalen quadratischen Fehler

$$E((\hat{x}_{T+1} - x_{T+1})^2)$$

so ergibt sich die optimale Prognose als

$$E(x_{t+1} \mid x_1, \dots, x_T) .$$

Damit ergibt sich für die ARMA-Modelle

$$\hat{x}_{T+1} = \sum_{i=1}^p \phi_i \tilde{x}_{T+1-i} + \sum_{j=0}^q \theta_j \tilde{\epsilon}_{T+1-i}$$

mit

$$\tilde{x}_{T+1-i} = \begin{cases} x_{T+1-i} & \text{für } i \geq 1 \\ \hat{x}_{T+1-i} & \text{sonst} \end{cases}$$

$$\tilde{\epsilon}_{T+1-i} = \begin{cases} \epsilon_{T+1-i} & \text{für } i \geq 1 \\ 0 & \text{sonst} \end{cases}$$

d.h. um die optimale Vorhersage zu erhalten, ersetze in

$$x_{T+1} = \sum_{i=1}^{\infty} \phi_i x_{T+1-i} + \sum_{i=0}^{\infty} \theta_i \epsilon_{T+1-i}$$

die gegenwärtigen und vergangenen Größen durch ihre beobachteten Werte und die zukünftigen Werte durch ihre optimale Prognose. Für das bilineare Modell lautet die Prognoseformel

$$\hat{x}_{T+1} = \sum_{i=1}^p \phi_i \tilde{x}_{T+1-i} + \sum_{i=0}^q \theta_i \tilde{\epsilon}_{T+1-i} + \sum_{i=1}^p \sum_{j=1}^Q \beta_{ij} \epsilon_{T+1-i} \tilde{x}_{T+1-j}$$

wobei \tilde{x}_{T+1-i} und $\tilde{\varepsilon}_{T+1-i}$ wie oben definiert sind

$$\varepsilon_{T+1-i} X_{T+1-j} = \begin{cases} \varepsilon_{T+1-i} X_{T+1-j} & i \geq 1 \wedge j \geq 1 \\ \varepsilon_{T+1-i} \hat{X}_{T+1-j} & i \geq 1 \wedge j < 1 \\ E(\varepsilon_{T+1-i} X_{T+1-j}) & i < 1 \wedge j < 1 \\ 0 & \text{sonst} \end{cases}$$

Die Interpretation der Prognoseformel ist entsprechend der im linearen Fall.

6. Probleme

6.1. Das Problem der Invertierbarkeit

In die Prognoseformeln geht die nicht beobachtbare Größe ε_t ein. Wie erhält man ε_t ?

Wir hatten geschrieben

$$\varepsilon_t = \sum g_i X_{t-i}$$

$$\varepsilon_t = \sum g_i X_{t-i} + \sum \sum g_{ij} X_{t-i} X_{t-j} + \dots$$

Wenn diese Darstellung gilt, heißt der durch (*) und (**)
dargestellte Prozeß invertierbar.

Mit Hilfe des Backward shift operators B lassen sich ARMA-Modelle in kompakter Form angeben. B angewandt auf x_t liefert als Ergebnis x_{t-1} , B^2 steht für zweimaliges Anwenden von B usw.

$$\text{Sei} \quad \Phi(B) = 1 - \Phi_1 B - \dots - \Phi_p B^p$$

$$\text{und} \quad \Theta(B) = 1 + \Theta_1 B + \dots + \Theta_q B^q$$

dann ist $\Phi(B)X_t = \Theta(B)\varepsilon_t$ die allgemeine Formel eines ARMA-Prozesses. Ein ARMA-Prozeß ist nur genau dann invertierbar, wenn $\Theta(B)$ als Polynom in B nur Wurzeln außerhalb des Einheitskreises hat.

Beispiel:

$$\text{MA}(1) \quad X_t = \epsilon_t + a \epsilon_{t-1} = (1+aB)\epsilon_t$$

$$\text{Wir hatten gefunden} \quad \epsilon_t = X_t - \sum (-a)^{i+1} X_{t-i}$$

d.h. für $|a| < 1$ liegt die Wurzel von $(1+aB)$ außerhalb des Einheitskreises, die Gewichte, mit denen die X_{t-i} in ϵ_t eingehen, nehmen mit wachsender Entfernung vom Zeitpunkt t ab.

$$\hat{\epsilon}_t = X_t - a \hat{\epsilon}_{t-1}$$

Mit einer Anfangsschätzung $\hat{\epsilon}_0$ läßt sich dann ϵ_t schätzen.

Für bilineare Prozesse hat man noch keine allgemeine Bedingung für das Vorliegen von Invertierbarkeit gefunden. Aber auch hier bedeutet die Forderung nach Invertierbarkeit Beschränkung für die Koeffizienten β_{ij} .

Wenn man aus substanzwissenschaftlichen Gründen Vorstellungen über die Größe der Koeffizienten hat, kann dies zu Widersprüchen mit den Invertierbarkeitsbedingungen führen.

6.2. Stationarität

Für MA-Modelle $\text{MA}(q)$ gibt es keine Probleme, sie sind immer stationär.

Bei $\text{AR}(p)$ und $\text{ARMA}(p,q)$ -Modellen muß $\Phi(B)$ nur Wurzeln außerhalb des Einheitskreises haben, damit der Prozeß stationär ist. Bei bilinearen Prozessen hat man noch keine allgemeine Bedingung an die Koeffizienten formulieren können, sondern nur für spezielle Modelle.

Es gilt aber auch hier, daß man mit aus substanzwissenschaftlichen Argumentationen gefundenen Koeffizienten in Widerspruch zu den Stationaritätsbedingungen kommen kann.

Eine Transformation zur Erzwingung von Stationarität ist $\nabla X_t = X_t - X_{t-1}$ d.h. $\nabla = 1 - B$. Dann kann man die obige Theorie statt auf X_t auf

$$Y_t = (1-B)^d X_t = \nabla^d X_t$$

anwenden.

Für lineare Modelle führt dies zu den sogenannten ARIMA-

Modellen. Für den bilinearen Fall liegen entsprechenden Arbeiten noch nicht vor.

6.3. Identifikation

Wie unterscheidet man den linearen und den nicht-linearen Fall und wie ermittelt man p, q bzw. p, q, P, Q ?

BOX & JENKINS (1970) bieten für den linearen Fall ein relativ ausgearbeitetes Konzept an. Es wurde inzwischen zum "Automatischen Box Jenkins Verfahren" weiterentwickelt.

Die bilinearen Prozesse stehen erst am Anfang ihrer Aufbereitung für praktische Anwendungen. GRANGER, ANDERSEN (1978) zeigen z.B. das bilineare Prozesse von der Box-Jenkins-Technik entweder als weißes Rauschen klassifiziert werden oder irrtümlich als $MA(q)$ - Prozeß identifiziert werden. Ein erstes Diskriminierungsmittel scheint die Analyse von X_t^2 mit Hilfe der Box-Jenkins-Technik zu sein. Ist nämlich X_t weißes Rauschen, so wird auch X_t^2 als solches identifiziert, während in den von Granger und Andersen untersuchten Fällen bei irrtümlich als solches klassifiziertes X_t der Prozeß X_t^2 nicht als weißes Rauschen akzeptiert wurde.

6.4. Schätzprobleme

Schon das ARMA-Modell $\Phi(B)X_t = \Theta(B)\epsilon_t$ $\epsilon_t = \{\Theta(B)\}^{-1} \Phi(B)X_t$ führt bei der Minimierung von $\sum \hat{\epsilon}_j^2$ zu einem nicht-linearen Kleinstquadrat-Problem, das bilineare Modell ändert an dieser Situation nichts.

7. Lohnt sich die Heranziehung der Theorie stochastischer Prozesse ?

Es eröffnen sich eine Fülle von konkreten Modellen

- so führt z.B. das ARIMA-Modell $(1-B)X_t = \epsilon_t + (1-a)\epsilon_{t-1}$ mit seiner optimalen Prognose gerade zur Formel des exponential smoothing.
- Man kann also vielleicht auf weitere "quasiheuristische" Prognoseformeln kommen.

Die Theorie der stochastischen Prozesse führt nicht nur zu vielen Problemen wie der vorstehende Abschnitt zu zeigen scheint, sondern bietet auch gleichzeitig eine Erweiterung des Instrumentariums für Identifikations- und Schätzprobleme. Es gilt nämlich

$$\gamma_{\tau} = \int_{-\pi}^{\pi} e^{i\tau\omega} f(\omega) d\omega$$

Die sogenannte spektrale Dichte $f(\omega)$ eröffnet einen Weg die bisher in der Dimension Zeit behandelten Phänomene in die Dimension Frequenz zu überprüfen. Diese Transformation erzeugt zwar keine neuen Informationen aber stellt vorhandene anders dar, so daß sie zur Lösung gewisser Probleme leichter herangezogen werden können. Man sollte kein Purist sein und nur Prognosen, die mit Hilfe der Theorie stochastischer Prozesse gewonnen werden, gelten lassen - der richtige Weg liegt vielleicht in der kombinierten Prognose.

Literaturverzeichnis

- BOX, G.E.P. und JENKINS, G.M.: Time Series analysis, forecasting and control, San Francisco 1970
- DAVIS, H.T.: The analysis of economic time series, Cowles Commission Monograph No. 6
- DOBB, J.L.: Stochastic processes, New York 1953
- GRANGER, C.W.J. und ANDERSEN, A.B.: An introduction to bilinear time series models, Vandenhoeck u. Ruprecht, Göttingen 1978
- KOLMOGOROFF, N.: Interpolation und Extrapolation von stationären zufälligen Folgen. In: Bulletin de l'académie de sciences de URSS, Ser. Math. 5(1941)
- LENZ, H.J.: Zeitreihenanalyse- und Prognoseverfahren zur Optimierung der Kassenhaltung bei autokorrelierten Zahlungsprozessen, Berlin 1974
- WIENER, N.: Extrapolation, interpolation and smoothing of stationary time series, New York 1949
- WOLD, H.: A study in the analysis of stationary time series, Upsala 1938