Diskussionsschriften

Prognose uni- und multivariater Zeitreihen

Manfred Deistler
Klaus Neusser

04-01

Januar 2004

Universität Bern
Volkswirtschaftliches Institut
Gesellschaftstrasse 49
3012 Bern, Switzerland
Tel: 41 (0)31 631 45 06
Web: www.vwi.unibe.ch

PROGNOSE UNI- UND MULTIVARIATER ZEITREIHEN

MANFRED DEISTLER AND KLAUS NEUSSER

1. Einführung

Zeitlich ablaufende zufällige Vorgänge können durch stochastische Prozesse modelliert werden, insbesondere ist es in diesem Rahmen möglich, Unsicherheit über die Zukunft zu beschreiben. Für stationäre Prozesse wurde bereits vor ca. 60 Jahren eine elegante Prognosetheorie von Kolmogorov [26] und Wiener [39] entwickelt. Ein weiterer wesentlicher Beitrag geht auf Kalman [25] zurück. Diese Theorie behandelt die lineare Kleinst-Quadrate-Prognose unter der Voraussetzung, dass die zweiten Momente des zugrundeliegenden Prozesses bekannt sind. In den meisten Fällen sind diese zweiten Momente jedoch nicht bekannt und müssen geschätzt werden, so dass das Prognoseproblem mit einem Identifikationsproblem einhergeht. Die Theorie der linearen Kleinst-Quadrate-Prognose stationärer Prozesse bei bekannten zweiten Momenten und die Theorie der Identifikation von AR-, ARMA- und Zustandsraumsystemen stellen die beiden Herzstücke der theoretischen Analyse des Prognoseproblems dar. Unsere Darstellung beschränkt sich auf diese lineare Kleinst-Quadrate-Prognose und die Identifikation von linearen dynamischen Systemen. Nichtlineare Prognosefunktionen und von den quadratischen abweichende Kostenfunktionen werden demnach nicht behandelt, wenn es nicht ausdrücklich erwähnt ist. Die Praxis hat gezeigt, dass diese linearen Ansätze auch bei offensichtlich nichtlinearen Mechanismen erstaunlich erfolgreich sind.

Date: 20. Januar 2004.

In der Praxis müssen bei der Entwicklung von Prognosealgorithmen der Verwendungszweck, die vorhandenen a priori-Informationen und die spezifischen Besonderheiten der Daten berücksichtigt werden. Was die Verwendung betrifft, so sind u. a. zu überlegen: Die Fristigkeit, die gewünschte Genauigkeit, die sich auch im Aufwand für die Modellierung niederschlägt, und der erforderliche Rechenaufwand. Im speziellen kann man zwei Extremfälle unterscheiden: Zum einen schnell verfügbare, relativ ungenaue Prognosen, bei denen auf eine detaillierte Modellierung der Daten weitgehend verzichtet wird. Solche Verfahren könnte man als "automatisierte Kurvenlineale" bezeichnen. Sie finden z. B. in der Absatzprognose Verwendung (siehe Abschnitt 8). Zum anderen Prognosen, bei denen eine möglichst hohe Genauigkeit erwünscht und daher eine detaillierte und zeitaufwendige Modellierung der Daten angezeigt ist. Ein Beispiel hierfür liefert die Prognose der Industrieproduktion (vgl. Abschnitt 7).

In vielen Fällen stehen zusätzlich a priori-Informationen zur Verfügung, die aber im Bereich der Wirtschaftswissenschaften, im Gegensatz zu vielen Anwendungen in den Naturwissenschaften oder den technischen Wissenschaften, oft unpräzise oder schwer quantifizierbar sind. Andererseits ist die Information aus den Daten in vielen Fällen alleine nicht ausreichend. In der Entscheidung über Art und Ausmaßder verwendeten a priori-Information zeigt sich ganz wesentlich die Kunst des Prognostikers. Abschnitte 7 und 8 bieten ein konkrete Beispiele für die bei der Prognose auftretenden Überlegungen und Vorgangsweisen.

2. Die Theorie der linearen Kleinst-Quadrate Prognose

In diesem Abschnitt behandeln wir die Theorie der linearen Kleinst-Quadrate (KQ)-Prognose für bekannte erste und zweite Grundgesamtheitsmomente. Wir nehmen an, dass die zu prognostizierende, n-dimensionale

Zeitreihe durch einen zugrundeliegenden Zufallsmechanismus, einen n-dimensionalen stochastischen Prozess $(X_t)_{t\in\mathbb{Z}}$ mit $\mathbb{R}^n$-wertigen Zufallsvariablen X_t und Indexmenge $\mathbb{Z}$, der Menge der ganzen Zahlen, erzeugt wird. Ferner sollen sämtliche Prozessvariablen endliche zweite Momente besitzen:

$$\mathrm{E}X_t'X_t < \infty \qquad \text{für alle t} \in \mathbb{Z} \tag{1}$$

dabei bezeichnet X_t einen n-dimensionalen Spaltenvektor und X_t' den zugehörigen transponierten Vektor. In diesem Fall existieren die Erwartungswerte $\mathrm{E}X_t = m$ und die Kovarianzmatrizen $\Gamma(t+\tau,t) = \mathrm{E}\left(X_{t+\tau} - \mathrm{E}X_{t+\tau}\right)\mathrm{E}\left(X_t - \mathrm{E}X_t\right)'$. Die Informationsgewinnung aus vergangenen und gegenwärtigen Werten $X_s,\ t-T \le s \le t$, zur Prognose eines zukünftigen Wertes $X_{t+\tau},\ \tau > 0$, erfolgt durch die Prognosefunktion (Prädiktor) $\mathrm{P}\left((X_s)_{t-T\le s\le t}\right)$. Dabei können endlich viele $(T < \infty)$ oder unendlich viele vergangene Werte $(T = \infty)$ berücksichtigt werden. Der Prädiktor hängt im allgemeinen vom Zeitpunkt t und dem Prognosehorizont τ ab. Um die Schreibweise zu erleichtern, verzichten wir auf die Indexierung des Prädiktors mit t bzw. τ. Die Werte $\widehat{X}(t,\tau) = \mathrm{P}\left((X_s)_{t-T\le s\le t}\right)$ der Prognosefunktion werden ebenfalls als Prädiktor bezeichnet. Zur konkreten Formulierung des Prognoseproblems muss festgelegt werden, was unter einer möglichst guten Annäherung des Prädiktors $\widehat{X}(t,\tau)$ an den zukünftigen Wert $X_{t+\tau}$ zu verstehen ist. Mit anderen Worten, es muss ein Gütekriterium festgelegt werden. Wir beschränken uns hier auf das Kleinst-Quadrate-Kriterium: Der Kleinst-Quadrate-Prädiktor (KQ-Prädiktor) aus einer gewissen Klasse von Prognosefunktionen minimiert

$$\mathrm{E}\left(X_{t+\tau} - \widehat{X}(t,\tau)\right)'\left(X_{t+\tau} - \widehat{X}(t,\tau)\right)$$

über diese Klasse.

Nimmt man die größtmögliche in diesem Kontext sinnvolle Klasse von Prognosefunktionen, also alle messbaren und quadratisch integrierbaren Funktionen der X_s, $t-T \leq s \leq t$, so erhält man den bedingten Erwartungswert

$$\mathrm{E}\left(X_{t+\tau}|X_t, X_{t-1}, \ldots\right).$$

Da die konkrete Berechnung des bedingten Erwartungswertes aber in vielen Fällen auf Schwierigkeiten stößt, beschränken wir uns auf lineare Prognosefunktionen und erhalten so lineare KQ-Prädiktoren. Die Beschränkung auf lineare Prognosefunktionen ist natürlich nur dann sinnvoll, wenn aufgrund der Struktur des Prozesses (X_t) die lineare KQ-Prognose und der bedingte Erwartungswert entweder zusammenfallen, wie im Gauss'schen Fall, oder zumindest nicht stark voneinander abweichen.

Das Problem der linearen KQ-Prognose für $X_{t+\tau}$ aus endlicher oder unendlicher Vergangenheit kann folgendermaßen formuliert werden: Man suche unter allen linearen Prognosefunktionen $\mathrm{P}\left((X_s)_{t-T\leq s\leq t}\right)$, also Linearkombinationen $b+\sum_{i=0}^{T} a_i X_{t-i}$ oder deren Grenzwerte, wobei b ein Element aus $\mathbb{R}^n$ und die a_i $n \times n$ Matrizen sind, diejenige, die

$$\mathrm{E}\left(X_{t+\tau} - \mathrm{P}\left((X_s)_{t-T\leq s\leq t}\right)\right)' \left(X_{t+\tau} - \mathrm{P}\left((X_s)_{t-T\leq s\leq t}\right)\right) \tag{2}$$

minimiert. Nach dem sogenannten Projektionssatz (siehe z. B. Brockwell und Davis [7]) ist der lineare KQ-Prädiktor $\widehat{X}(t,\tau)$ vollständig durch folgende Eigenschaften charakterisiert:

$$\widehat{X}(t,\tau) = \mathrm{P}\left((X_s)|t-T\leq s\leq t\right), \qquad \mathrm{P}\text{ linear} \tag{3}$$

$$\mathrm{E}\left(X_{t+\tau} - \widehat{X}(t,\tau)\right) X_s' = 0, \qquad t-T\leq s\leq t \tag{4}$$

$$\mathrm{E}X_{t+\tau} = \mathrm{E}\widehat{X}(t,\tau) \tag{5}$$

Zudem gibt es immer genau ein $\widehat{X}(t,\tau)$ mit diesen Eigenschaften. Der lineare KQ-Prädiktor existiert also immer und ist eindeutig. Die Bedingungen (4) und (5) besagen, dass der *Prognosefehler* $\left(X_{t+\tau} - \widehat{X}(t,\tau)\right)$

orthogonal zu den vergangenen Zufallsvariablen $X_s,\ t-T \leq s \leq t$ und zur Konstanten 1 steht. Anschaulich gesagt bedeutet der Projektionssatz, dass man die beste lineare Approximation für einen Punkt durch einen Teilraum als orthogonale Projektion des Punktes auf den Teilraum erhält. Das wichtigste Maß für die Qualität der Prognose ist die *Prognosefehlervarianz*

$$\Sigma_\tau = \mathrm{E}\left(X_{t+\tau} - \widehat{X}(t,\tau)\right)\left(X_{t+\tau} - \widehat{X}(t,\tau)\right)'$$

Im Fall der Prognose aus endlicher Vergangenheit ($T < \infty$) ist der Prädiktor durch die Normalgleichungen (4) und (5) folgendermaßen bestimmt:

$$\begin{aligned}\begin{pmatrix} a_0 & \cdots & a_T \end{pmatrix} &\begin{pmatrix} \Gamma(t,t) & \cdots & \Gamma(t-T,t) \\ \vdots & \ddots & \vdots \\ \Gamma(t,t-T) & \cdots & \Gamma(t-T,t-T) \end{pmatrix} \\ &= \begin{pmatrix} \Gamma(t+\tau,t) & \cdots & \Gamma(t+\tau,t-T) \end{pmatrix} \qquad (6)\end{aligned}$$

$$\mathrm{E}X_{t+\tau} = b + \sum_{i=0}^{T} a_i \mathrm{E}X_{t-i} \qquad (7)$$

Zur Lösung des linearen KQ-Prognoseproblems reicht die Kenntnis der ersten und zweiten Momente aus. Man erhält die Lösung durch Inversion der Matrix $(\Gamma(i,j))_{i,j=t,\ldots,t-T}$, falls diese regulär ist. Man kann zeigen, dass das System (6) auch dann lösbar ist, wenn die Matrix singulär ist. In diesem Fall bestimmt jede Lösung $b, a_0, \ldots, a_T$ den gleichen Prädiktor.

Zur Lösung des Prognoseproblems werden im allgemeinen weitere Annahmen über die Struktur des Prozesses (X_t) getroffen. Diese sind, zumindest wenn die zweiten Momente geschätzt werden müssen, aus statistischen Gründen praktisch immer erforderlich. Oft wird angenommen, dass der stochastische Prozess *stationär* (im weiteren Sinn) ist,

d. h., es gilt zusätzlich zu Bedingung (1)

$$\begin{aligned} \mathrm{E}X_t &= m = \text{konstant}, && \text{für alle } t \in \mathbb{Z} \\ \Gamma(t+s,t) &= \gamma(s) && \text{hängt für alle } t, s \in \mathbb{Z} \text{ nicht von } t, \\ & && \text{sondern nur von } s \text{ ab.} \end{aligned}$$

Die Funktion γ heißt die Kovarianzfunktion des Prozesses.

In vielen Anwendungen kann der ursprüngliche Prozess in eine Summe aus einem Trend ($\mathrm{E}X_t = m_t$) und einer stationären Komponente zerlegt werden. Da die Prognose des Trends, bei bekannten ersten Momenten, trivial ist, beschränken wir uns auf die Darstellung der Prognose stationärer zentrierter, d. h. erwartungswertbereinigter Prozesse. In anderen Fällen kann durch Differenzenbildung Stationarität erreicht werden.

3. Die Prognose aus unendlicher Vergangenheit

Obwohl in der Praxis nur endliche Vergangenheit vorliegt, ist die Theorie der Prognose aus unendlicher Vergangenheit dennoch wichtig, da sie die tiefere Struktur des Problems aufzeigt. Die allgemeine Prognosetheorie stationärer Prozesse, die für den eindimensionalen Fall von Kolmogorov [26], Wiener [39] und Wold [40] und für den mehrdimensionaien Fall von z. B. Rozanov [33] behandelt wurde, wird heute in der Praxis weniger verwendet als ursprünglich angenommen. Wir werden diese Theorie daher hier nicht darstellen, sondern nur einige zentrale Ergebnisse festhalten.

Klassifiziert man stationäre Prozesse nach ihrem Prognoseverhalten, so kann man zwei Extreme herausgreifen: Die sogenannten *singulären* Prozesse lassen sich exakt aus unendlicher Vergangenheit prognostizieren:

$$\widehat{X}(t,\tau) = X_{t+\tau} \qquad \text{für alle } \tau > 0 \tag{8}$$

Die bei weitem wichtigste Klasse singulärer Prozesse sind endliche Summen von Sinus- und Cosinusfunktionen mit zufälligen Amplituden (harmonische Prozesse). Die *regulären* Prozesse sind charakterisiert durch:

$$\lim_{\tau\to\infty} \widehat{X}(t,\tau) = \mathrm{E}X_t = 0 \tag{9}$$

Für $\tau \to \infty$ kann also aus der Vergangenheit keine über den Erwartungswert hinausgehende Information zur linearen Prognose gewonnen werden.[1]

Nach dem Satz von Wold lässt sich jeder stationäre Prozess (X_t) eindeutig als Summe eines stationären regulären Prozesses (Y_t) und eines stationären singulären Prozesses (Z_t) darstellen:

$$X_t = Y_t + Z_t \tag{10}$$

wobei (Y_t) und (Z_t) unkorreliert sind und sich durch lineare Transformationen aus $X_s, s \le t$, ergeben. Man kann das Prognoseproblem für beide Komponenten getrennt lösen. In der Praxis kann man sich fast immer auf reguläre Prozesse beschränken, da der Einfluss harmonischer Prozesse durch eine einfache Regression berücksichtigt werden kann.

Jeder reguläre stationäre Prozess (X_t) besitzt eine Darstellung (Wold-Darstellung) der Form

$$X_t = \sum_{i=0}^{\infty} C_i\, \varepsilon_{t-i} \tag{11}$$

wobei (ε_t) weißes Rauschen ist, also

$$E\varepsilon_t = 0, \qquad E\varepsilon_s\varepsilon_t' = \begin{cases} \Sigma, & \text{für } s = t; \\ 0, & \text{für } s \neq t. \end{cases} \tag{12}$$

gilt und ε_t eine lineare Transformation von $X_s, s \le t$, ist. Die $n \times n$ Matrizen C_i sind quadratisch summierbar[2] und es kann die Normierung

[1] Von nun an soll Konvergenz als Konvergenz im quadratischen Mittel verstanden werden.

[2] Quadratisch summierbar bedeutet, dass $\sum_{i=0}^{\infty} \|C_i\|^2 < \infty$, wobei $\|C_i\|^2$ den größten Eigenwert von C_iC_i' bezeichnet.

$C_0 = I_n$ gewählt werden. Aus (11) kann man sehen, dass der lineare KQ-Prädiktor mit dem bedingten Erwartungswert $\mathrm{E}\,(X_{t+\tau}|X_t, X_{t-1}, \ldots)$ genau dann übereinstimmt, wenn gilt $\mathrm{E}\,(\varepsilon_{t+1}|\varepsilon_t, \varepsilon_{t-1}, \ldots) = 0$, wenn also (ε_t) nicht nur weißes Rauschen, sondern auch ein Martingaldifferenzenprozess ist.

Aufgrund der Wold-Darstellung gilt für den ersten Term auf der rechten Seite der folgenden Zerlegung

$$X_{t+\tau} = \sum_{i=\tau}^{\infty} C_i\, \varepsilon_{t+\tau-i} + \sum_{i=0}^{t-1} C_i\, \varepsilon_{t+\tau-i} \tag{13}$$

dass er eine lineare Funktion der $X_s, s \leq t$, ist, und für den zweiten Term

$$\mathrm{E}\left(\sum_{i=0}^{t-1} C_i\, \varepsilon_{t+\tau-i}\right) X_s' = 0, \quad \text{für alle } s \leq t.$$

Aufgrund des Projektionssatzes gilt somit:

$$\widehat{X}(t, \tau) = \sum_{i=\tau}^{\infty} C_i\, \varepsilon_{t+\tau-i} \tag{14}$$

und der Prognosefehler ist

$$\sum_{i=0}^{t-1} C_i\, \varepsilon_{t+\tau-i}$$

Damit ist jedoch das Prognoseproblem noch keineswegs gelöst. Erstens müssen die Matrizen C_i bestimmt werden, und zweitens müssen, um die Prognosefunktion zu erhalten, in (14) die ε_s durch $X_i, i \leq s$, ersetzt werden.

Wir führen folgenden wichtigen Begriff ein: Für jeden regulären stationären Prozess wird durch

$$f(\lambda) = \frac{1}{2\pi} \sum_{s=-\infty}^{\infty} \gamma(s) e^{\imath\lambda s} \quad \text{mit } \lambda \in [-\pi, \pi] \tag{15}$$

$$\gamma(s) = \int_{-\pi}^{\pi} e^{-\imath\lambda s}\, f(s)\, d\lambda \tag{16}$$

der Kovarianzfunktion γ umkehrbar eindeutig die *spektrale Dichte* $f : [-\pi, \pi] \longrightarrow \mathbb{C}^{n\times n}$ zu geordnet[3]. Die unendliche Summe (15) ist dabei im Sinne der Konvergenz im quadratischen Mittel zu verstehen. Aus (11) folgt

$$f(\lambda) = \frac{1}{2\pi}\left(\sum_{s=0}^{\infty} C_s e^{-\imath\lambda s}\right)\Sigma\left(\sum_{s=0}^{\infty} C_s e^{-\imath\lambda s}\right)^* \tag{17}$$

($*$ bedeutet konjungiert transponiert).

Die Ermittlung der C_s, aus f, also aus den zweiten Momenten, bezeichnet man als die Faktorisierung der spektralen Dichte. Wir wollen uns mit diesem Problem hier in voller Allgemeinheit ebensowenig befassen wie mit der Ersetzung der ε_s durch $X_i, i \leq s$, also der Herleitung der Prognoseformel.

4. AR- und ARMA-Prozesse

Die bei weitem wichtigste Klasse von regulären stationären Prozessen sind die ARMA-Prozesse (X_t). Sie sind als stationäre Lösung der stochastischen Differenzengleichung

$$\sum_{i=0}^{p} A_i X_{t-i} = \sum_{i=0}^{q} B_i \varepsilon_{t-i} \tag{18}$$

definiert. Dabei sind A_i und B_i $n \times n$ Matrizen und (ε_t) ist weißes Rauschen mit $\Sigma = \mathrm{E}\varepsilon_t\varepsilon_t'$. Für $q = 0$ erhält man als Spezialfall die autoregressiven Prozesse (AR-Prozesse), für $p = 0$ die moving average Prozesse (MA-Prozesse).

Wir nehmen weiter an:

$$\det \Sigma > 0 \tag{19}$$

$$\det \sum_{i=0}^{\infty} A_i z^i \neq 0, \qquad \text{für alle } z \in \mathbb{C} \text{ mit } |z| \leq 1, \tag{20}$$

$$\det \sum_{i=0}^{\infty} B_i z^i \neq 0, \qquad \text{für alle } z \in \mathbb{C} \text{ mit } |z| \leq 1, \tag{21}$$

[3] Die spektrale Dichte ist λ-fast überall bestimmt

wobei $\mathbb{C}$ den Körper der komplexen Zahlen bezeichnet. Die große Bedeutung der ARMA-Prozesse in der Praxis liegt vor allem in zwei Tatsachen begründet. Zum ersten kann jeder reguläre stationäre Prozess durch geeignete Wahl der Ordnungen p und q beliebig genau durch ARMA-Prozesse approximiert werden. Zum zweiten hängen die zweiten Momente von ARMA-Prozessen, bei gegebenen p und q, nur von endlich vielen Parametern ab. Der Parameterraum ist daher für gegebene Ordnungen p und q endlichdimensional, wodurch die Schätzung bedeutend vereinfacht wird.

Die Annahme (20) erlaubt es, die Lösung des ARMA-Systems (18) zu schreiben als

$$X_t = \sum_{i=0}^{\infty} C_i \varepsilon_{t-i}, \tag{22}$$

wobei die C_i bestimmt sind durch

$$C(z) = \sum_{i=0}^{\infty} C_i z^i = A^{-1}(z)B(z) \tag{23}$$

mit $A(z) = \sum_{i=0}^{p} A_i z^i$ und $B(z) = \sum_{i=0}^{q} B_i z^i$. Die Stabilitätsbedingung (20) sichert somit auch die Stationarität der Lösung.

Die spektrale Dichte von (X_t) ist wegen (22) und (23) gegeben durch:

$$f(\lambda) = \frac{1}{2\pi} A^{-1}(e^{-\imath\lambda}) B(e^{-\imath\lambda})\, \Sigma\, B^*(e^{-\imath\lambda}) A^{*^{-1}}(e^{-\imath\lambda}) \tag{24}$$

Die spektrale Dichte repräsentiert in gewissem Sinne die äußeren, d. h. die aus den Beobachtungen direkt schätzbaren Eigenschaften des Prozesses. Sie kann konsistent geschätzt werden. Die Modellparameter A_i, B_i und Σ stellen hingegen die innere Struktur des Systems dar. Im allgemeinen können A_i, B_i und Σ nicht ohne weitere Annahmen aus $f(\lambda)$ eindeutig bestimmt werden. Das ist das sogenannte Identifizierbarkeitsproblem.

Dieses Identifizierbarkeitsproblem kann in zwei Schritte zerlegt werden. Im ersten geht es um die Eindeutigkeit von $C(z)$ und Σ bei gegebenem $f(\lambda)$; im zweiten um die Eindeutigkeit von $A(z)$ und $B(z)$ bei

gegebenem $C(z)$ (siehe dazu Abschnitt 5). Für die Prognose ist primär der erste und einfachere Schritt wichtig. Unter unseren Annahmen, insbesondere unter (21), ist $C(z)$ eindeutig bis auf Nachmultiplikation mit einer konstanten nichtsingulären Matrix bestimmt (vgl. [19] und [21]). Durch eine einfache Normierung, wie z. B.

$$C_0 = I_n \tag{25}$$

können die C_i eindeutig festgelegt werden.

Aus (18) und (21) folgt:

$$\varepsilon_t = \sum_{i=0}^{\infty} D_i X_{t-i} \tag{26}$$

mit $\sum_{i=0}^{\infty} D_i z^i = B^{-1}(z)A(z)$. Die ε_t sind also lineare Funktionen der $X_s, s \leq t$, die Lösung (22) ist somit eine *Wold* Darstellung (11). Aus (14) und (26) folgt daher:

$$\widehat{X}(t,\tau) = \sum_{i=\tau}^{\infty} C_i \sum_{j=0}^{\infty} D_j X_{t+\tau-i-j} \tag{27}$$

Für bekannte C_i und daher bekannte D_i ist somit die Prognosefunktion durch (27) gegeben. Zur Bestimmung der C_i und daher der D_i aus $f(\lambda)$ siehe z. B. [33, Kapitel 1.10] oder [21]. Heute werden die A_i, B_i und Σ meist direkt geschätzt.

Die Prognosefehlervarianz ist gegeben durch:

$$\Sigma_\tau = \sum_{i=0}^{\tau-1} C_i \, \Sigma \, C_i^* \tag{28}$$

Für AR-Prozesse ist die Prognose besonders einfach. Nimmt man $A_0 = B_0 = I_n$ an, so ist

$$X_t = -A_1 X_t - \ldots - A_p X_{t-p} + \varepsilon_t \tag{29}$$

Aus dem Projektionssatz folgt:

$$\widehat{X}(t,1) = -A_1 X_{t-1} - \ldots - A_p X_{t+1-p} \tag{30}$$

da $\mathrm{E}\,\varepsilon_{t+1}X_s' = \mathrm{E}\,\varepsilon_{t+1}\left(\sum_{i=0}^{\infty} C_i\varepsilon_{s-i}\right)' = 0$ für $s \leq t$. Mit dem gleichen Argument gilt

$$\widehat{X}(t,2) = -A_1\widehat{X}(t,1) - A_2X_t - \ldots - A_pX_{t+2-p} \tag{31}$$

usw. Beim AR-Prozess bestimmen die A_i also sehr direkt die Prognoseformel.

Man kann den Prädiktor aus gegebenen oder geschätzten Koeffizienten eines ARMA-Modells durch Koeffizientenvergleich aus $A(z)C(z) = B(z)$ blockrekursiv berechnen:

$$\begin{aligned} &A_0\,C_0 = B_0 \\ &A_1\,C_0 + A_0\,C_1 = B_1 \\ &\qquad \ldots \\ &A_p\,C_{i-p} + A_{p-1}\,C_{i-p+1} + \cdots + A_0\,C_i = 0 \quad \text{für } i \geq \max(p, q+1) \end{aligned} \tag{32}$$

Die C_i erhält man aus der Lösung dieses (unendlichen) linearen Gleichungssystems. Analog können auch die D_i in (27) bestimmt werden:

$$\begin{aligned} &B_0\,D_0 = A_0 \\ &B_1\,D_0 + B_0\,D_1 = A_1 \\ &\qquad \ldots \\ &B_q\,D_{i-q} + B_{q-1}\,D_{i-q+1} + \cdots + B_0\,D_i = 0 \quad \text{für } i \geq \max(q, p+1) \end{aligned} \tag{33}$$

Wegen (20) bzw. (21) konvergieren die C_i bzw. D_i für $i \longrightarrow \infty$ geometrisch gegen 0. Für viele praktische Fälle ist diese Konvergenz sogar sehr schnell, so dass man oft mit relativ kleinen "Anfangsausschnitten" aus (32) und (33) eine hinreichend genaue Approximation für den Prädiktor aus unendlicher Vergangenheit erhält.

Eine zweite Möglichkeit der praktischen Berechnung ist die Verallgemeinerung der in (30) und (31) beschriebenen Iteration auf den ARMA-Fall. Durch Bildung der Prädiktoren auf beiden Seiten von (18) erhält

man wegen der Linearität der Projektion:

$$\sum_{i=0}^{p} A_i \widehat{X}(t, \tau - i) = \sum_{i=0}^{q} B_i \hat{\varepsilon}(t, \tau - i) \tag{34}$$

wobei

$$\hat{\varepsilon}(t, \tau - i) = \begin{cases} \varepsilon_{t+\tau-i}, & \text{für } \tau - i \leq 0; \\ 0, & \text{für } \tau - i > 0. \end{cases} \tag{35}$$

$$\widehat{X}(t, \tau - i) = X(t + \tau - i), \qquad \text{für } \tau - i \leq 0 \tag{36}$$

Ersetzt man die ε_s, in (34) dann aus (18) durch $X_{s-i},\ i \geq 0$, und $\varepsilon_{s-i},\ i > 0$, so erhält man auf diese Weise eine Approximation für den Prädiktor. Dies ist nur eine Approximation, da die unbekannten Anfangswerte bei dieser Prozedur gleich Null gesetzt wurden. Für genügend großes t ist ihr Einfluss allerdings gering.

Betrachten wir als Beispiel ein einfaches ARMA-System mit $n = 1,\ p = q = 1$:

$$X_t + aX_{t-1} = \varepsilon_t + b\varepsilon_{t-1} \tag{37}$$

für die Einschrittprognose gilt:

$$\begin{aligned} \widehat{X}(t, 1) &= -aX_t + b\varepsilon_t \\ &= -aX_t + bX_t + abX_{t-1} - b^2\varepsilon_{t-1} \\ &= (-a + b)X_t + (-a + b)(-b)X_{t-1} + \ldots \\ &\qquad + (-1)^{t-1}ab^tX_0 + (-1)^t b^{t+1}\varepsilon_0 \end{aligned} \tag{38}$$

und wir setzen $X_0 = 0$ und wie bereits zuvor erwähnt $\varepsilon_0 = 0$. Die Zweischrittprognose erhält man aus

$$\widehat{X}(t, 2) = (-a + b)\hat{X}(t, 1) + abX_t + \ldots$$

usw.

5. Die Schätzung der Prädiktoren für ARMA-Systeme

In der Praxis sind die C_i bzw. f unbekannt und müssen aus den Beobachtungen geschätzt werden. Dadurch kommt zu dem vorher beschriebenen Prognosefehler im Fall bekannter zweiter Momente ein weiterer Prognosefehler hinzu. Die Komponente des Prognosefehlers kann bei endlichen Stichproben erheblich sein, geht aber bei Konsistenz der Schätzer mit wachsender Stichprobe gegen null.

Das Schätzproblem ist der vielleicht schwierigste Teil des Prognoseproblems. Bei ARMA-Modellen liegt es nahe, nicht die C_i oder f direkt zu schätzen, sondern A_i, B_i und Σ und, falls unbekannt, p und q. Das bietet den Vorteil, dass für vorgeschriebenes p und q der Parameterraum Teilmenge des Euklidischen Raumes ist, was die Schätzung sehr vereinfacht. Wir behandeln zunächst den Fall, dass p und q a priori bekannt sind.

Das Problem der Schätzung ist sehr eng mit dem zweiten Schritt des Identifizierbarkeitsproblems verzahnt, also dem Problem der eindeutigen Festlegung von A_i und B_i[4]. Dieses Identifizierbarkeitsproblem ist im mehrdimensionalen Fall bedeutend schwieriger zu lösen als im eindimensionalen Fall. Die wichtigsten Beiträge hierzu stammen von Hannan [18]. Einfache, jedoch nicht ganz allgemeine, hinreichende Bedingungen zur Identifizierbarkeit sind:

$$A_0 = B_0 = I_n \tag{39}$$

$$A(z) \text{ und } B(z) \text{ sind relativ linksprim} \tag{40}$$

$$(A_p, B_q) \text{ hat Rang } n \tag{41}$$

[4] Hannan and Deistler [21] bzw. Ljung [29] geben eine ausführliche Darstellung der hier zusammengefassten Ergebnisse

Bedingung (40) bedeutet, dass alle gemeinsamen Linksteiler von $A(z)$ und $B(z)$ unimodular sind, das ARMA-System daher keine künstlich aufgeblähte Dynamik besitzt.[5]

Das am häufigsten verwendeten Schätzverfahren erhält man aus der Gauss'schen Likelihood Funktion L_T:

$$-\frac{2}{T}\ln L_T(\theta) = \frac{1}{T}\ln\det\Gamma_T(\theta) + \frac{1}{T}\mathbf{X}_T'\Gamma_T^{-1}(\theta)\mathbf{X}_T. \tag{42}$$

Die Optimierung dieser Funktion ergibt den Maximum Likelihood (ML-Schätzer) Schätzer. Dabei enthält der Parametervektor θ die freien, d. h. unabhängig wählbaren Parameter in A_i, B_i und Σ in einer bestimmten Anordnung, und wir wählen folgende Notation:

$$\mathbf{X}_T := \begin{pmatrix} x_1 \\ \vdots \\ x_T \end{pmatrix} \qquad \Gamma_T(\theta) := \begin{pmatrix} \gamma_\theta(0) & \gamma_\theta(1) & \cdots & \gamma_\theta(T-1) \\ \gamma_\theta'(1) & \gamma_\theta(0) & \cdots & \gamma_\theta(T-2) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_\theta'(T-1) & \gamma_\theta'(T-2) & \cdots & \gamma_\theta(0) \end{pmatrix}$$

wobei $\gamma_\theta(s)$ die Kovarianzfunktion eines ARMA-Prozesses mit dem Parametervektor θ darstellt.

In den meisten Fällen erhält man eine Approximation des ML-Schätzers durch die numerische Optimierung der Funktion L_T. Unter den Identifizierbarkeitsannahmen lässt sich mit einigen weiteren Voraussetzungen zeigen, dass der ML-Schätzer konsistent und asymptotisch normal ist (vgl. [21]).

Betrachten wir nun den Fall, dass p und q unbekannt sind. Hier gibt es zwei Fehlermöglichkeiten: entweder sind die p und q zu groß ("overfitting") oder p oder q sind zu klein ("underfitting") gewählt. Im Fall von "overfitting" ist der ML-Schätzer i.a. nicht mehr konsistent für die A_i und B_i, wohl aber für die C_i. Mit anderen Worten, bei

[5] Eine Polynommatrix $U(z)$ heißt unimodular, falls $\det U(z) = \text{konstant} \neq 0$. Im Fall (37) würde aufgeblähte Dynamik bedeuten $a = b \neq 0$. Siehe Hannan und Deistler [21] für eine eingehende Diskussion.

"overfitting" geht wohl die Konsistenz für die wahren Parameter A_i und B_i verloren, der ML-Schätzer konvergiert aber gegen die wahre Äquivalenzklasse, d. h. gegen die Menge aller A_i und B_i, die nach (23) die wahren C_i ergeben. Die C_i und damit die Prognosefunktion werden konsistent geschätzt. Das Schätzproblem ist also für die Prognose in diesem Sinne gutmütiger als für die Parameter A_i und B_i.

Wir erläutern diesen Sachverhalt anhand des eindimensionalen ARMA-Systems mit der Spezifikation (37). Gilt für das wahre System $a = b = 0$, so konvergiert der ML-Schätzer im allgemeinen nicht gegen $a = b = 0$, sondern nur gegen die Gerade $a = b$ mit der Einschränkung $|a| < 1$ und $|b| < 1$. Für die Gerade $a = b$ gilt $\sum C_i z^i = A^{-1}(z)B(z) = 1$. Der ML-Schätzer konvergiert also gegen die wahren C_i ($C_0 = 1; C_i = 0, i > 0$).

Zwar muss beim "overfitting" ein Verlust an asymptotischer Effizienz in Kauf genommen werden. Zudem wirft die Unbestimmtheit der Parameterschätzung auch numerische Probleme auf, die jedoch bei rein autoregressiven Prozessen nicht auftreten. Die Schätzer der "überzähligen" Parameter konvergieren in diesem Fall gegen Null. Außerdem ist der ML-Schätzer bei AR-Modellen vom Kleinst-Quadrate-Typ. Die Schätzformel ist daher im Gegensatz zum allgemeinen ARMA-Fall explizit gegeben und daher schnell auswertbar. Deshalb werden AR-Modelle oft bevorzugt (vgl. Parzen [32]), zumal auch sie reguläre stationäre Prozesse beliebig genau approximieren können. Um jedoch eine bestimmte Güte der Approximation zu erreichen, müssen beim reinen AR-Modell im allgemeinen viel mehr Parameter geschätzt werden als beim ARMA-Modell, was einen Nachteil des AR-Ansatzes darstellt.

Bei "underfitting" konvergieren die ML-Schätzer gegen die besten Prädiktoren, die der eingeschränkte Parameterraum zulässt.

Aufbauend auf den Arbeiten von Akaike [1] wurden vollautomatisierte Verfahren zur Schätzung von p und q entwickelt (siehe Hannan und

Deistler [21]). Die Grundidee ist dabei die folgende: Der ML-Schätzer tendiert insofern zum "overfitting", als die Werte der Likelihoodfunktion für größeres p und q größer oder zumindest gleich sein werden. Aus diesem Grund liegt es nahe, die Likelihoodfunktion oder den ML-Schätzer für Σ mit einem Korrekturterm zu versehen, der von der Anzahl der freien Parameter und damit von p und q abhängt. Eine wichtige Klasse von Kriterien ist von der Form:

$$A(p,q) = \ln \det \hat{\Sigma}_T(p,q) + n^2(p+q)\frac{C(T)}{T} \tag{43}$$

wobei $\hat{\Sigma}_T(p,q)$ der ML-Schätzer von Σ für gegebenes p und q ist. Die Funktion $C(T)$ beschreibt den "trade-off" zwischen der Güte der Anpassung des Systems an die Daten und der Komplexität oder genauer der Dimension des Parameterraums. Konkrete Wahlen sind:

$$C(T) = 2 \qquad \text{(AIC-Kriterium)}$$

$$C(T) = \ln T \qquad \text{(BIC-Kriterium)}$$

Die Schätzer von p und q werden durch Minimierung von $A(p,q)$ über einen bestimmten endlichen ganzzahligen Bereich gewonnen. Es kann gezeigt werden, dass das BIC-Kriterium unter allgemeinen Voraussetzungen konsistent ist, während das AIC-Kriterium asymptotisch zur Überschätzung neigt (vgl. [20]).

Eine Alternative zur Verwendung von Informationskriterien besteht darin, diejenige Spezifikation zu wählen, die die Summe der Quadrate der "out-of-sample" Einschrittprognosefehler minimiert.

6. ARMAX-Modelle und bedingte Prognose

Bei vielen Anwendungen hängen die endogenen Variablen (X_t) noch von exogenen Variablen oder Inputs (Z_t) ab. Dann erweitert man das ARMA-System (18) zu einem ARMAX-System:

$$\sum_{i=0}^{p} A_i X_{t-i} = \sum_{i=0}^{r} E_i Z_{t-i} + \sum_{i=0}^{q} B_i \varepsilon_{t-i}, \tag{44}$$

wobei E_i $n \times m$ Matrizen sind und A_0 nicht unbedingt gleich I_n sein muss. Analog spricht man von einem ARX-System, wenn $q = 0$ und $B_0 = I_n$ ist. Die "klassischen" linearen ökonometrischen Systeme sind ARX-Systeme.

Bei ARMAX-Systemen wird oft die Frage nach der *bedingten Prognose* für die endogenen Variablen, gegeben die exogenen Variablen, gestellt, z. B. um die Auswirkungen wirtschaftspolitischer Instrumente auf ökonomische "Zielvariablen" zu prognostizieren. Der lineare KQ-Prädiktor für $X_{t+\tau}$ gegeben $X_s,\ s \leq t$, und $Z_s,\ s \leq t+\tau$, ist die lineare KQ-Approximation von $X_{t+\tau}$ durch diese gegebenen Variablen.

Wir nehmen an, dass (Z_t) nichtstochastisch ist und (20) sowie (21) erfüllt sind. Der Erwartungswert von $X_{t+\tau}$ ist dann

$$\mathrm{E}X_{t+\tau} = \sum_{i=0}^{\infty} T_i Z_{t+\tau+i} \tag{45}$$

mit

$$\begin{aligned} T(z) &= \sum_{i=0}^{\infty} T_i z^i = A^{-1}(z)E(z) \\ E(z) &= \sum_{i=0}^{r} E_i z^i \end{aligned} \tag{46}$$

und dieser Erwartungswert kann aus der Kenntnis der A_i, E_i und $Z_s, s \leq t+\tau$, berechnet werden. Die Abweichungen $X_{t+\tau} - \mathrm{E}X_{t+\tau}$ vom Erwartungswert genügen einem ARMA-System (18) und können mit der in Abschnitt 4 behandelten Theorie getrennt prognostiziert werden.

Die Annahme, dass (Z_t) nichtstochastisch ist, wurde hier nur aus Gründen der Einfachheit der Notation gesetzt. Die Argumentation verläuft analog, falls $(X_t', Z_t')'$ stationär ist und $\mathrm{E}Z_s\varepsilon_t' = 0$ für alle s und t gilt.

Im "klassischen" ökonometrischen Fall genügt $X_t - \mathrm{E}X_t$ einem autoregressiven Prozess. Der lineare KQ-Prädiktor ist daher:

$$\widehat{X}(t,1) = -A_0^{-1}\left(\sum_{i=1}^{p} A_i X_{t+1-i} + \sum_{i=0}^{r} E_i Z_{t+1-i}\right) \tag{47}$$

und die Mehrschrittprognosen erhält man durch iteratives Einsetzen analog zu (31).

Wieder ist die Schätzung der A_i, B_i, E_i und Σ in (44) der komplizierteste Schritt zur tatsächlichen Ermittlung der Prädiktoren. Der eigentlichen Schätzung ist wiederum ein Identifizierbarkeitsproblem vorgelagert. Im ARMAX-Fall ist die "strukturelle" Identifizierbarkeit, also die Identifizierbarkeit der ökonomisch direkt interpretierbaren Parameter aufgrund von a priori-Information, besonders wichtig. Hier lassen sich meist aus der a priori-Kenntnis, dass gewisse Variable in gewissen Gleichungen nicht erscheinen, Nullrestriktionen an die Elemente von A_i und E_i ableiten. Diese a priori-Information reduziert in vielen praktischen Fällen die Dimension des Parameterraumes (durch Überidentifikation) bedeutend und erhöht dadurch die asymptotische Effizienz, ja sie macht oft bei einer relativ großen Zahl von Gleichungen und relativ kurzen Zeitreihen die Schätzung erst möglich.

Zur Schätzung selbst empfehlen sich wieder der ML-Schätzer oder geeignete Approximationen - praktisch alle ökonometrischen Schätzverfahren lassen sich als Approximation an den ML-Schätzer auffassen. Die negative logarithmierte Likelihoodfunktion ist in diesem Fall proportional zu

$$-\frac{2}{T}\ln L_T(\theta) = \frac{1}{T}\ln\det\Gamma_T(\theta) + \frac{1}{T}\left(\mathbf{X}_T^a - \mathbf{Z}_T^e\right)'\Gamma_T^{-1}\left(\mathbf{X}_T^a - \mathbf{Z}_T^e\right) \quad (48)$$

Dabei sind $\mathbf{X}_T^a$, $\mathbf{Z}_T^e$ bzw. ε_T^b $nT \times l$ Vektoren, bei denen der t-te Block $\sum A_i X_{t-1}$, $\sum E_i Z_{t-i}$ bzw. $\sum B_i \varepsilon_{t-i}$ ist; und $\Gamma_T(\theta) = \mathrm{E}\varepsilon_T^b(\varepsilon_T^b)'$. Die ML-Schätzer sind unter allgemeinen Annahmen konsistent und asymptotisch normal.

Die dynamische Spezifikation des ARMAX-Systems, also die Festlegung der maximalen Lag-Längen p, r und q, erfolgt analog zum ARMA-Fall. In vielen Anwendungen will man noch zusätzlich die Inputs bzw.

die exogenen Variablen datengetrieben auswählen.Dabei wird die Inputselektion aus einer vorgegebenen Liste von $\tilde{m}$ a-priori Kandidaten gleichzeitig mit der dynamischen Spezifikation, etwa mittels eines der in Gleichung (43) angegebenen Informationskriterien, durchgeführt. Dabei wird oft das Kriterium über alle Teilmengen der Menge aller Input Kandidaten optimiert, so dass sich ohne Dynamik schon $2^{\tilde{m}}$ unterschiedliche Inputkombinationen ergeben. Klarerweise soll die Zahl der Spezifikationen über die das Informationskriterium optimiert werden soll in einem vernünftigen Verhältnis zur Stichprobengröße stehen, um "overfitting" zu vermeiden; mit anderen Worten bei kleinen Stichproben macht es wenig Sinn ein Informationskriterium über zu viele Spezifikationsmöglichkeiten zu optimieren.

In wirtschaftspolitischen Anwendungen ist es oft nicht von vornherein klar, welche Variablen endogen und welche exogen sind. Wir geben folgende Definition von Exogenität: Sei (Y_t) der Prozess aller beobachtbaren Variablen, wobei noch nicht notwendigerweise zwischen endogenen und exogenen Variablen unterschieden worden ist, gegeben eine beliebige Partitionierung (eventuell nach Umordnung der Komponenten) $(Y_t) = (X_t', Z_t')'$, dann heißt (Z_t) exogen für (X_t), falls die Projektion von X_t auf $(Z_s)_{s\in\mathbb{Z}}$ nicht von $(Z_s)_{s>t}$ abhängt. Zukünftige Werte des Prozesses (Z_t) verbessern diese Approximation nicht. Diese Definition der Exogenität entspricht dem Kausalitätskonzept von Granger [16] (siehe [34] und [35]). Diese Definition kann direkt für einen Test (Kausalitätstest) verwendet werden, bei dem man in der Regression

$$X_t = \sum_{i=-N}^{N} \beta_i Z_{t-i} + u_t$$

die Koeffizienten $\beta_i,\ i < 0$, z. B. mit einem F-Test, auf Null testet. Es gibt eine umfangreiche Literatur über alternative Methoden (vgl. [15]).

Die praktische Bedeutung des Kausalitätstests für die Prognose besteht darin, dass man das Prognoseproblem in zwei Schritte zerlegen

kann. Zuerst werden die exogenen Variablen aus ihrer eigenen Vergangenheit prognostiziert. Im zweiten Schritt werden die endogenen Variablen bedingt vorhergesagt, wobei für die exogenen Variablen die Prädiktoren der ersten Stufe eingesetzt werden. Bei gegebenen zweiten Momenten ist diese Prognose gleich der unbedingten Prognose für (Y_t). Weißman jedoch, dass (Z_t) exogen ist, so hat man weniger Parameter zu schätzen und damit die Möglichkeit eines Effizienzgewinns.

7. Die Prognose gesamtwirtschaftlicher Grössen

In diesem Abschnitt wird ein makroökonomisches Prognosemodell für Deutschland besprochen. Dieses konkrete Beispiel soll aufzeigen, welche zusätzlichen inhaltlichen wie statistischen Überlegungen angestellt werden müssen, um ein aussagekräftiges Prognosemodell zu entwickeln. Wir beschränken uns auf folgende vier Zeitreihen: die Industrieproduktion Y_t, den Verbraucherpreisindex P_t, das Zinssatzdifferential zwischen dem Monatsgeldmarktsatz und der Umlaufsrendite R_t, sowie der Arbeitslosenrate U_t.[6] Diese Zeitreihen stehen nicht nur im Mittelpunkt makroökonomischer Forschung, sondern beanspruchen auch das Interesse einer breiteren Öffentlichkeit. Wir haben das Zinssatzdifferential anstelle eines Zinssatzes verwendet, da die Differenz zwischen kurz- und langfristigem Zinssatz wichtige Informationen über den zukünftigen Verlauf der wirtschaftlichen Aktivität sowie der

[6] Alle Daten wurden der Zeitreihen-Datenbank der Deutschen Bundesbank mit Internetadresse http://www.bundesbank.de/stat/zeitreihen/index.htm entnommen. Die genaue Beschreibung der ausgewählten Reihen ist wie folgt: die arbeitstäglich bereinigte Industrieproduktion mit Basisjahr 1995 (Code: UXNI63), der Verbraucherpreisindex mit Basisjahr 2000 (Code: UUFA01), der Zinssatz für Monatsgeld am Frankfurter Bankplatz (Code: SU0104), die ungewogene Umlaufsrendite von Bundeswertpapieren mit Restlaufzeit von 9 bis 10 Jahre (Code: WX3950) und die Arbeitslosenquote bezogen auf alle zivilen Erwerbspersonen (Code: UUCC02).

Inflationsrate enthält (siehe [13] und [12]). Alle Daten stehen monatlich ab Januar 1992 bis Oktober 2003 zur Verfügung und umfassen daher nur die Zeit nach der deutschen Wiedervereinigung. Als eigentlichen Schätzzeitraum wurde nur die Zeit bis zum Dezember 2001 gewählt. Die restliche Zeitspanne von Januar 2002 bis Oktober 2003 dient zur Evaluation der Prognosegüte des Modells.

Charakteristisch für viele ökonomische Zeitreihen ist das Vorhandensein eines Trends und/oder eines Saisonmusters. Die Zeitreihen sind daher nicht stationär, so dass die in den vorangegangenen Abschnitten dargestellte Theorie nicht unmittelbar anwendbar ist. Um die Natur des Trends und des Saisonmusters feststellen zu können, insbesondere ob diese deterministisch oder stochastisch sind, sind in den letzten beiden Jahrzehnten eine Vielzahl von Tests entwickelt worden, auf die aus Platzmangel nicht weiter eingegangen werden kann (siehe etwa [5] oder [23]). In jedem Fall empfiehlt es sich, soweit dies möglich ist, sowohl den Trend als auch die Saison zu modellieren und nicht mechanisch die Daten durch ein Saisonbereiningungsprogramm (z.B. Census X12) zu glätten. Diese Programme können nicht nur die Eigenschaften der Zeitreihen in oft unsystematischer Weise verfälschen, sondern sind, da sie im Kern einen unendlichen zweiseitigen Filter approximieren, für die Evaluation und Erstellung von Prognosemodellen ungeeignet. Eine geeignete Transformation der Variablen besteht darin, die Wachstumsraten gegenüber dem Vorjahresmonat zu bilden, also $\Delta_{12} \ln Y_t = \ln Y_t - \ln Y_{t-12}$ bzw. $\Delta_{12} \ln P_t = \ln P_t - \ln P_{t-12}$ zu modellieren. Eine Transformation des Zinssatzdifferentials und der Arbeitslosenquote ist nicht angezeigt. Der zu modellierende Prozess besteht daher aus $(X_t) = \big((\Delta_{12} \ln Y_t, \Delta_{12} \ln P_t, R_t, U_t)'\big)$.

Ausgangspunkt der Modellierung bildet ein vektorautoregressives (VAR) Modell der Ordnung 2.[7] Dabei zeigt sich, dass die saisonalen Effekte durch die obigen Transformationen nicht vollständig berücksichtigt werden.[8] Es werden daher neben der Konstanten noch zwei Dummy-Variable, eine für Dezember und eine für Januar, im Modell berücksichtigt. Die Anwendung der Informationskriterien (43) zur Bestimmung der Ordnung des VAR Modells ergibt ein uneinheitliches Bild. Während das AIC-Kriterium eine Ordnung von 12 wählt, bevorzugt das BIC-Kriterium ein VAR der Ordnung 1. Da ein VAR Modell der Ordnung 12 aufgrund der relativ kleinen Stichprobe und der damit verbundenen vielen statistisch insignifikanten Koeffizienten wenig Sinn macht, wird ein VAR Modell der Ordnung 2 inklusive Konstante und den beiden Dummy-Variablen spezifiziert, wobei zusätzlich noch U_{t-12} in jeder Gleichung und $\Delta_{12} \ln Y_{t-12}$ nur in der Gleichung für die Produktion berücksichtigt werden. Diese pragmatische Vorgangsweise scheint sich in der Praxis zu bewähren (siehe [27]). Da das gesamte Modell relativ viele Parameter umfasst, wird auf eine Auflistung der einzelnen Koeffizienten verzichtet. Stattdessen werden die Zusammenhänge anhand folgender Abbildung verdeutlicht. Dabei zeigen die durchgezogenen (strichlierten) Pfeile Zusammenhänge an, die aufgrund eines F-Test am 5(10)-Prozentniveau statistisch signifikant sind.

Die Prognosegüte des Modells wurde wie folgt untersucht. Zuerst wurde das Modell für den Beobachtungszeitraum Januar 1992 bis Dezember 2001 geschätzt. Anschließend wurden mit diesem Modell Prognosen für die nächsten 24 Monate erstellt. Abbildung 2 vergleicht die Prognosewerte mit den Realisationen, die allerdings nur bis Oktober 2003 zur Verfügung stehen. Diese Abbildung zeigt, dass das Wachstum

[7] In der Sprechweise der vorigen Abschnitte entspricht dies einem AR Modell mit maximaler Verzögerung von $p = 2$.

[8] Kointegrationsbeziehungen (siehe [24] und [31]) zwischen den hier verwendeten Variablen werden aus theoretisch-ökonomischen Überlegungen ausgeschlossen.

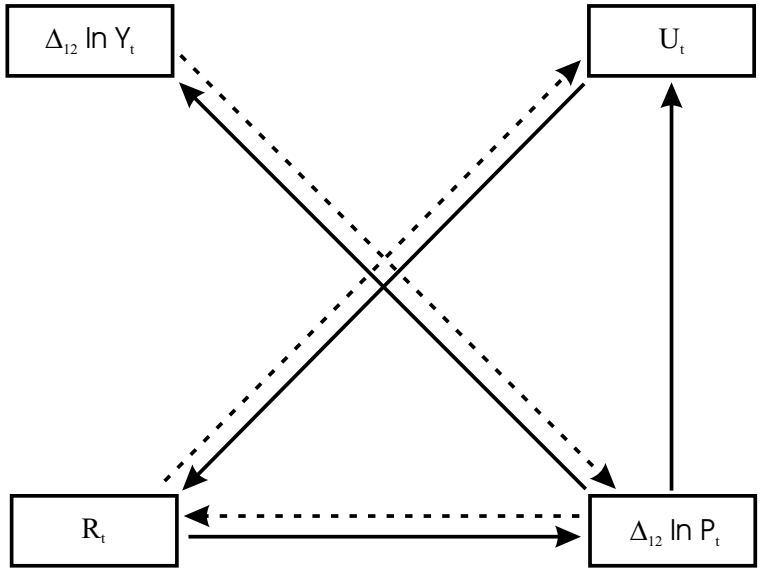

ABBILDUNG 1. Zusammenhänge zwischen den Variablen

der Industrieproduktion im Jahre 2002 relativ gut prognostiziert wird. Der Aufschwung bricht jedoch im Jahr 2003 zusammen, was durch das Modell nicht erfasst wird. Die Prognose der Inflationsrate sowie des Zinssatzdifferentials fällt wenig befriedigend aus, da sowohl die Richtung als auch das Niveau der Entwicklung verfehlt werden.

Im nächsten Schritt wird der Schätzzeitraum um ein Monat, den Januar 2002, erweitert und das so adaptierte Modell für eine neuerliche Prognose, diesmal von Februar 2002 bis Dezember 2003, verwendet. Danach wurde der Schätzzeitraum um ein weiteres Monat erweitert, die Koeffizienten des Modells wiederum an die neue Information angepasst und Prognosen für die nächsten Monate berechnet. Insgesamt können auf diese Weise 22 Einschritt-, 21 Zweischritt-, und schließlich 10 Zwölfschrittprognose erstellt werden. Die Güte der Prognosen wird anhand der Wurzel des mittleren quadrierten Prognosefehlers (Root Mean Squared Error, RMSE) überprüft. Bei einem Prognosehorizont von τ Monaten und einem Überprüfungszeitraum von N Monaten ist

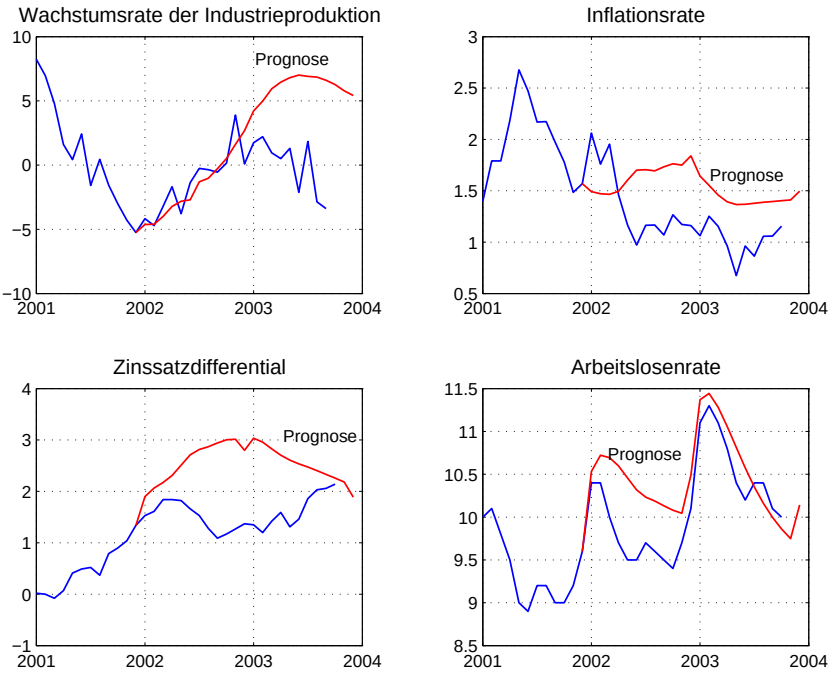

ABBILDUNG 2. Prognose und Realisation

der RMSE für die Variable $x_t = X_{it}$ für den Prognosehorizont τ definiert durch:

$$RMSE(\tau) = \sqrt{\frac{1}{N+1-\tau} \sum_{i=0}^{N-\tau} \left(x_{t+\tau+i} - \hat{x}(t+i,\tau)\right)^2} \qquad (49)$$

Da für die Evaluation der Prognose nur jener Zeitraum verwendet wird, der nicht zur Identifikation und Schätzung des Modells diente, spricht man von einer "out-of-sample" Evaluation.

Die Ergebnisse dieser Übung sind in Tabelle 1 zusammengefasst. Insgesamt bestätigt die Tabelle, dass die Prognose makroökonomischer Variablen schwierig ist. Dies trifft vor allem auf die Prognose der Wachstumsrate der Industrieproduktion zu. Die Gründe für diese unbefriedigende Ergebnisse sind vielfältig (siehe [22]). Diese reichen von Messfehlern in den Daten über die Auswahl der Variablen, die Instabilität und Unzuverlässlichkeit aggregierter makroökonomischer Zusammenhänge

	Prognosehorizont (in Monaten)						
Variable	1	2	3	4	6	8	12
$\Delta_{12} \ln Y$	1.77	1.96	2.15	2.62	3.31	3.76	4.63
$\Delta_{12} \ln P$	0.26	0.35	0.45	0.52	0.63	0.68	0.73
R	0.23	0.39	0.47	0.55	0.72	0.82	0.88
U	0.18	0.30	0.39	0.43	0.43	0.44	0.42

TABELLE 1. Prognosegüte (RMSE) für verschiedene Prognosehorizonte

(siehe etwa Lucas-Kritik [30]) bis hin zu Fehlspezifikationen der Modelle.[9] Bei VAR Modellen wirkt sich besonders die Überparametrisierung negativ auf die Prognosegüte aus (siehe [14] und [27]). Eine Methode zur Lösung dieses Problems besteht darin, statistisch nicht signifikante Koeffizienten zu eliminieren. Dieses einfache heuristische Verfahren kann die Prognosegüte bereits deutlich verbessern. Eine andere bewährte Methode besteht darin, Bayesianische Verfahren zu verwenden, um a-priori-Information besser berücksichtigen zu können (siehe [28] und [36]). Ausser rein statistischen Verfahren kann auch die ökonomische Theorie zur Gewinnung von a-priori-Information beitragen. Diese ist jedoch bezüglich der Dynamik der Variablen meist wenig informativ.

8. ABSATZPROGNOSE

Prognosen gesamtwirtschaftlicher Größen, wie sie im vorigen Abschnitt besprochen wurden, waren lange Zeit die in der Ökonometrie dominierenden Prognosen. In den letzten zwanzig Jahren haben jedoch ökonometrische Prognosemethoden und Modelle im Finanzmarkt- und Unternehmensbereich stark an Bedeutung gewonnen.

[9] Abschnitt 8 diskutiert weitere Problemfelder und mögliche Lösungsvorschläge.

Die Finanzmarktökonometrie hat durch die große Nachfrage nach empirischen Analysen und Entscheidungsfundierung auf diesem Sektor einen enormen Aufschwung genommen; diese Nachfrage wurde ihrerseits durch die steigende Bedeutung der Finanzmärkte und durch die steigende Komplexität ihrer Produkte ausgelöst. Darauf soll aber in diesem Beitrag nicht eingegangen werden.

Vielleicht nicht ganz so spektakulär wie im Finanzmarktbereich, aber doch deutlich an Gewicht gewinnend sind ökonometrische Prognose- und Analysemethoden im Unternehmensbereich. Zunehmende Verfügbarkeit von Daten, etwa in Data Warehouses und wachsender Konkurrenzdruck sind wesentliche Anreize, auch anspruchsvollere Methoden einzusetzen. Wir betrachten hier die Absatzprognose, die einen wesentlichen Input, etwa für das Supply- Chain Management (siehe z.B. Aviv [4]), bildet.

Bei der Erstellung eines Prognosemodells sind folgende Fragen zu klären bzw. Punkte zu berücksichtigen (siehe z.B. Überwimmer und Deistler [37] und Wehlig [38]):

- Wie viele Zeitreihen sollen prognostiziert werden? Bei Bedarfsprognosen für Warenwirtschaftssysteme, wo oft 10000 und mehr Artikel betrachtet werden, ist vollständige Automatisierung anzustreben. Sind wenige Zeitreihen zu prognostizieren, so kann die Genauigkeit des Prognosemodells durch “fine tuning” erhöht werden.
- Liegen sehr viele Zeitreihen vor, so stellt sich die Frage, ob diese in Cluster mit gleichem Prognoseverhalten unterteilt werden sollen.
- Sollen die Prognosewerte der Zeitreihen direkt verwendet werden oder werden sie z.B. noch von Experten korrigiert? Eine

interessante Entwicklung sind sogenannte hybride Prognosesysteme, die z.B. auf ARX Systemen basierenden Zeitreihenprognosen mit wissensbasierten Systemen, bei denen zusätzliche Erfahrungsregeln verwendet werden, kombinieren.

- Sind die Kostenfunktionen stark asymmetrisch, d.h. die Folgekosten von positiven Prognosefehlern unterscheiden sich stark von denen betragsgleicher negativer Prognosefehler, dann werden in gewissen Fällen sogenannte lin-lin Kostenfunktionen verwendet, die aus zwei linearen Ästen bestehen. In diesem Fall erhält man den optimalen Prädiktor als ein Quantil der bedingten Verteilung, das durch den Quotienten der Steigungen der beiden Geradenäste der Kostenfunktion gegeben ist (siehe Christoffersen und Diebold [8]).
- Welche Kalendereffekte (z.B. Feiertags-, Weihnachts- oder Ostereffekte) und Saisoneffekte sind relevant? Derartige Effekte können durch Dummyvariablen berücksichtigt werden. Ein anderer Zugang sind variable, z.B. tagesspezifische Lags (siehe Deistler et.al. [10]).
- Soll der Trend getrennt modelliert werden?
- Sind Ausreisser vorhanden?
- Welche Variablen beeinflussen die zu prognostizierende Variable? Typische Kandidaten für Inputs sind: Preisreduktionen, Preisreduktionen bei Konkurrenzprodukten, Werbung und Promotionsmassnahmen.
- Ein spezielles Problem bilden in der Beobachtungszeit neu eingeführte Produkte.

In Ueberwimmer und Deistler [37] und Wehling [38] wurden Analyse- und Prognosemodelle für Absätze von Markenartikeln aus Wochendaten über einem Beobachtungszeitraum von zwei oder drei Jahren entwickelt. Der Prognosezeitraum beträgt eine Woche; in Wehling [38] lag das Hauptaugenmerk auf der Vollautomatisierung der Prozedur.

Die Basismodelle waren dabei univariate (genauer ein Output, mehrere Inputs) ARX Modelle. Die Spezifikation der Dynamik und die Inputselektion wurden, wie zuvor beschrieben, mit Informationskriterium oder dem out-of-sample Einschrittprognosefehler durchgeführt. Um eine zeitaufwändige Untersuchung aller Teilmengen der Liste der Kandidaten für die prädeterminierten Variablen (also der Inputs und der verzögerten Outputs) zu vermeiden, wurde der sogenannte An-Algorithmus (An und Gu [2]) verwendet. Dieser Algorithmus sucht in einer "intelligenten" Weise über eine Teilmenge der Menge aller möglichen Spezifikationen. In vielen Fällen sind verzögerte Inputs wichtig, um nicht allen Inputs die gleiche Dynamik aufzuprägen. In der überwiegenden Mehrzahl der Fälle ist es wichtig, auch die letzten noch verfügbaren Daten zur Identifikation zu benützen, die Parameter werden daher laufend neu geschätzt. Ebenso wurde die Spezifikation der Modelle in regelmässigen Abständen erneut vorgenommen. Dies wurde sowohl mit einem gleitenden ("moving") oder einem sich erweiternden ("extending") Fenster vorgenommen (s.a. Deistler und Hamann [11]). Dadurch wird auch der Effekt von Variationen der Parameter abgeschwächt.

Für die Validierung der Prognosemodelle sind folgende Schritte wichtig:

- Der quadratische Korrelationskoeffizient, berechnet aus dem out-of-sample- Einschnittprognosefehler. Dabei ist wichtig, dass für Parameterschätzung und Spezifikation nur Daten verwendet werden, die vor dem Prognosezeitpunkt liegen. Insbesondere ist der

Vergleich mit dem "no change predictor" als einfachsten Benchmark wichtig.

- Tests auf Schiefe der Fehler: Starke Schiefe kann ein Indikator für Ausreisser sein. t- oder F-Tests für die Parameter bringen andererseits wenig zusätzliche Information; auf AIC oder BIC basierte Verfahren lassen sich als Sequenzen von Likelihood Ratio Tests interpretieren.
- Werbeeinflüsse werden oft durch sogenannte Adstocks beschrieben (s. Überwimmer und Deistler [37]). Hier ist zu untersuchen, ob dadurch die Dynamik der Werbeeinflüsse ausreichend beschrieben wird.
- Um zu untersuchen, ob die Parameter langsam mit der Zeit variieren oder ob Strukturbrüche vorliegen, kann man, um einen ersten Hinweis zu erhalten, adaptive Schätzverfahren verwenden, die den Zeitpfad der Parameter "tracken". Ein einfaches Schätzverfahren besteht darin, die zurückliegenden Daten mit einem geometrischen Vergessensfaktor zu gewichten, wobei dieser Faktor seinerseits durch "grid search" bestimmt wird. Aus den Zeitpfaden der Parameter kann man auch Hinweise über die Validität der Spezifikation erhalten.
- Eventuell auftretende Strukturbrüche können durch Tests untersucht werden. Ebenso sind Strukturbrüche in der Spezifikation von Interesse, so könnten z.B. gewisse Inputs nur in speziellen Regimen wirksam sein.
- Schliesslich ist zu prüfen, ob Nichtlinearitäten die Prognosequalität verbessern. Dies kann ganz allgemein durch den Vergleich mit auf neuronalen Netzen basierenden Prognosen (siehe z.B. Wehlig [38]) oder durch Erweiterungen des ARX Ansatzes geschehen. Übliche Erweiterungen sind etwa (siehe Ueberwimmer

und Deistler [37]) das additive Hinzufügen von Interaktionstermen (z.B. das Produkt aus Werbung und Preisreduktion), STARX Modelle (Granger und Teräsvirta [17]) oder Asymmetrieterme für die Werbewirkung.

In gewissen Fällen ist eine multivariate Modellierung angebracht, etwa wenn Produkte eng verwandt sind und ihre Absätze starke gemeinsame Bewegungen aufweisen. Bei "normalen"multivariaten (oder Vektor-) ARX Systemen kommt man dabei mit der Anzahl der gemeinsam zu modellierenden Produkte sehr rasch an Grenzen. Aus diesem Grunde werden oft Faktor Modelle "vorgeschaltet"(siehe Deistler et.al. [9] und Deistler et.al. [10]):

$$X_t = \Lambda f_t + u_t \tag{50}$$

wobei f_t r-dimensionale Faktoren sind, mit $r \ll n$, die dann mit ARX Modellen prognostiziert werden. Λ ist die $n \times r$ Faktorladungsmatrix. Das Faktormodell kann z.B. aus einer Hauptkomponentenanalyse stammen oder einer Rang-reduzierten Regression entsprechen (siehe Anderson [3] und Deistler und Hamann [11]

Die Prognosequalitäten die mit den von uns erstellten Modellen erzielt wurden, sind bei verschiedenen Produkten deutlich verschieden. Wir haben in fast allen Fällen zumindest keine Verbesserungen der Prognose mit neuronalen Netzen festgestellt; durch Interaktionsterme und STARX Modelle konnten zum Teil jedoch Verbesserungen erzielt werden.

Wir haben somit einen Zugang und eine Methodik zur Erstellung von Prognosemodellen für Absatzdaten in groben Zügen beschrieben. Dieser Zugang lässt sich auf andere Unternehmensdaten (in Zeitreihenform) übertragen. Im Kern dieses Zuganges stehen ARX Modelle, ihre Identifikation und die auf ihnen basierende Prognose. Anwendungen

finden derartige Prognosen etwa für die Lagerhaltung und die Produktionsplanung.

Literatur

[1] H. Akaike, *A New Look at the Statistical Model Identification*, IEEE Transactions on Automatic Control AC-19 (1974), S. 716 ff.

[2] H. An und L. Gu, *On Selection of Regression Variables*, Acta Mathematica Applicatae Sinica 2 (1985), S. 27 ff.

[3] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*, 2. Auflage, John Wiley and Sons 1984.

[4] Y. Aviv, *A Time-Series Framework for Supply-Chain Inventory Management*, Operations Research 51 (2003), S. 210 ff.

[5] A. Banerjee, J. Dolado, J. W. Galbraith und D. F. Hendry, *Co-Integration, Error-Correction, and the Econometric Analysis of Non-Stationary Data*, Oxford, 1993.

[6] B.S. Bernanke, *Alternative Explanations of the Money-Income Correlation*, Carnegie-Rochester Conference Series on Public Policy 25 (1986), S. 49 ff.

[7] P.J. Brockwell und R.A. Davis, *Time Series: Theory and Methods* New York, 1987.

[8] P. F. Christoffersen und F. X. Diebold, *Optimal Prediction under Asymmetric Loss* Economic Theory 13 (1997), S. 808 ff.

[9] M. Deistler, W. Fraißler, G. Petritsch und W. Scherrer, *Ein Vergleich von Methoden zur Kurzfristprognose elektrischer Last* Archiv für Elektrotechnik 71 (1988), S. 389 ff.

[10] M. Deistler, I. Käfer, W. Scherrer und M. Überwimmer, *Prediction of Daily Power Demand* in: B. Bitzer (Hrsg.): Proceedings of the 3rd European IFS Workshop, 2 (2000), S. 52 ff.

[11] M. Deistler und E. Hamann, *Identification of Factor Models for Forecasting Returns*, Forschungsbericht, Institut für Ökonometrie, Operations Research und Systemtheorie, TU Wien 2003.

[12] M. Dotsey, *The Predicitve Content of the Interest Rate Term Spread for Future Economic Growth*, Federal Reserve Bank of Richmond Economic Quarterly 84 (1998), S. 31 ff.

[13] A. Estrella und F. S. Mishkin, *The Predictive Power of the Term Structure of Interest Rates: Implications for the European Central Bank*, European Economic Review 41 (1997), S. 1375ff.

[14] R. Fair, *An Analysis of the Accuracy of Four Macroeconometric Models*, Journal of Political Economy 87 (1979), S. 701 ff.

[15] J. Geweke, *Inference and Causality in Economic Time Series Models*, in: Griliches, Z. und Intriligator, M.D. (Hrsg.), Handbooks of Econometrics, Vol. 2, Amsterdam 1984, S. 1101 ff.

[16] C.W.J. Granger, *Investing Causal Relations by Econometric Models and Cross-Spectral Methods*, Econometrica 37 (1969), S. 424 ff.

[17] C. W. J. Granger und T. Teräsvirta, *Modelling Economic Relationships*, Oxford University Press 1993.

[18] E.J. Hannan, *The Identification of Vector Mixed Autoregressive-Moving Average Systems*, Biometrika 57 (1969), S. 223 ff.

[19] E.J. Hannan, *Multiple Time Series*, New York 1970.

[20] E.J. Hannan, *The Estimation of the Order of an ARMA Process*, Annals of Statistics 8(1980), S. 1071 ff .

[21] E.J. Hannan und M. Deistler, *The Statistical Theory of Linear Systems*, New York 1988.

[22] D. F. Hendry und M. P. Clements, *Economic Forecasting: Some Lessons from Recent Research*, Mimeo 2002.

[23] S. Hylleberg, *Seasonality in Regresssion*, Orlando, Florida, 1986.

[24] S. Johansen, *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*, Oxford, 1995.

[25] R.E. Kalman, *A New Approach to Linear Filtering and Prediction Problems*, Journal of Basic Engineering, Transactions of the American Society of Mechanical Engineers D 82 (1960), S. 35 ff.

[26] A.N. Kolmogorov, *Sur l'interpolation et extrapolation des suites stationaires*, Comptes rendus hebdomadaires des scéances de l'Académie des Sciences 208(1939), S. 208 ff.

[27] R. Kunst und K. Neusser, *A Forecasting Comparison of Some VAR Techniques*, International Journal of Forecasting 2 (1986), S. 447 ff.

[28] R. B. Litterman, *Forecasting With Bayesian Vector Autoregressions – Five Years of Experience*, Journal of Business and Economic Statistics 4 (1986), S. 25 ff.

[29] L. Ljung, *System Identification, Theory for the User*, New Jersey 1987.

[30] R. E. Lucas, *Econometric Policy Evaluation: A Critique*, Journal of Monetary Economics, 2 (1976), Supplement Carnegie-Rochester Conference Series on Public Policy, S.19 ff.

[31] K. Neusser, *Testing the Long-Run Implications of the Neoclassical Growth Model*, Journal of Monetary Economics 27 (1991), S.3 ff.

[32] E. Parzen, *Multiple Time Series: Determining the Order of Approximating Autoregressive Schemes*, in: Krishnaiah, P. (Hrsg.), Multivariate Analysis IV, Amsterdam 1977, S. 283 ff.

[33] Y.A. Rozanov, *Stationary Random Processes*, San Francisco 1967.

[34] C.A. Sims, *Money, Income, and Causality*, American Economic Review 62 (1972), S. 540 ff.

[35] C.A. Sims, *Macroeconomics and Reality*, Econometrica 48 (1980), S. 1 ff.

[36] C. A. Sims, *A Nine-Variable Probabilistic Macroeconomic Forecasting Model*, in: J.H. Stock und M. W. Watson (Hrsg.), Business Cycles, Indicators, ans Forecasting, Chicago 1993, S. 179 ff.

[37] M. Überwimmer und M. Deistler, *Modelling Effects of Advertising for Non-Durable Goods* Mimeo 2003.

[38] S. Wehling, *Prognosemodelle für Warenwirtschaftssysteme im Hinblick auf einen vollständige Automatisierung*, Diplomarbeit, Institut für Ökonometrie, Operations Research und Systemtheorie, TU Wien 2003.

[39] N. Wiener, *Extrapolation, Interpolation and Smoothing of Stationary Time Series with Engineering Applications*, New York 1949.

[40] H. Wold, *A Study in the Analysis of Stationary Time Series*, Uppsala 1938.

Institut für Ökonometrie, Operations Research und Systemtheorie, Technische Universität Wien, Argentinierstrasse 8, A-1040 Wien, Österreich

E-mail address: Manfred.Deistler@tuwien.ac.at

Volkswirtschaftliches Institut, Universität Bern, Gesellschaftsstrasse 49, CH-3012 Bern, Schweiz

E-mail address: klaus.neusser@vwi.unibe.ch