

ÜBER DIE SCHÄTZUNG VON SPEKTRALDICHTEN

Immanuel M. BOMZE
Robert M. KUNST
Erhard RESCHENHOFER

Forschungsbericht/
Research Memorandum No. 184

Mai 1983

Die in diesem Forschungsbericht getroffenen Aussagen liegen im Verantwortungsbereich der Autoren und sollen daher nicht als Aussagen des Instituts für Höhere Studien wiedergegeben werden.

A B S T R A C T
=====

The investigation of spectral densities has become an indispensable part of time series analysis which in turn is one of the most common tools of quantitative social science. This paper concentrates on the highlights of current research in spectral estimation and does not intend to give a full survey of the known methods.

Following an overview of foundations, the second part presents two methods of parametric inference, namely recursive and robust estimation. The third section deals with nonparametric procedures, in particular with some weighted covariance estimators; a new lag window is introduced which is optimal in the sense of MISE.

The algorithms described in this paper have been implemented by the authors at the Institute of Advanced Studies. Representative outputs are listed in appendix A while appendix B is devoted to ladder forms on which all specified recursions are based.

Z U S A M M E N F A S S U N G

=====

Die Untersuchung von Spektraldichten hat sich zu einem unentbehrlichen Teil der Zeitreihenanalyse entwickelt, welche wiederum eines der gebräuchlichsten Instrumente der quantitativen Sozialwissenschaften darstellt. Diese Arbeit greift die wichtigsten Aspekte neuerer Forschung auf dem Gebiet der Spektralschätzung heraus, ohne einen vollständigen Überblick über die bekannten Methoden geben zu wollen.

Nach einer kurzen Darstellung der Grundlagen werden im zweiten Teil zwei Methoden parametrischer Inferenz vorgestellt, nämlich rekursive und robuste Schätzung. Der dritte Abschnitt behandelt nichtparametrische Verfahren, insbesondere einige gewichtete-Kovarianz-Schätzer; ein neues Lagfenster wird eingeführt, welches im Sinne von MISE optimal ist.

Die in dieser Arbeit beschriebenen Algorithmen wurden von den Verfassern am Institut für Höhere Studien implementiert. Beispielhafte Ausdrücke finden sich in Anhang A, während Anhang B den "ladder forms" gewidmet ist, auf denen alle angeführten Rekursionen beruhen.

INHALTSVERZEICHNIS

=====

I. Grundlagen	
I.1 Die WOLD-Zerlegung	4
I.2 Die Spektraldarstellung	6
I.3 ARMA Prozesse	8
II. Parametrische Verfahren : AR- und ARMA-Modelle	
II.1 Rekursive Schätzung... ..	9
II.1.1 Ordnungs- und Parameterschätzung von AR-Prozessen	10
II.1.2 Ordnungs- und Parameterschätzung von ARMA-Prozessen..	11
II.2 Robuste Spektralschätzung..	13
III. Nichtparametrische Verfahren..	23
Anhang A..	26
Anhang B..	30
Literatur.	34

I. GRUNDLAGEN

In diesem Kapitel wird die Theorie der stationären Prozesse - soweit sie für diese Arbeit relevant ist - in knapper Form dargestellt.

Für eine ausführlichere Diskussion sei auf die einschlägige Literatur verwiesen (DOOB(1953), ROZANOW(1967), HANNAN(1970)).

I.1 DIE WOLD ZERLEGUNG

Ein stochastischer Prozeß ist eine Folge von komplexwertigen Zufallsvariablen (bezüglich eines festen Wahrscheinlichkeitsmaßes P).

Ein stochastischer Prozeß

$$\underline{y} = (y_t)_{t \in \mathbb{Z}}$$

heißt (schwach) stationär, wenn sämtliche Momente erster und zweiter Ordnung existieren, d.h.

$$E y_t \in \mathbb{C}, E y_s \overline{y_t} \in \mathbb{C} \quad \forall s, t \in \mathbb{Z}$$

und invariant gegenüber Verschiebungen entlang der Zeitachse sind, d.h.

$$E y_t = E y_0, E y_s \overline{y_t} = E y_0 \overline{y_{t-s}} \quad \forall s, t \in \mathbb{Z}.$$

O.B.d.A. nehmen wir künftig $E y_0 = 0$ an.

Wegen

$$\int |y_t|^2 dP = E|y_t|^2 = E y_t \overline{y_t} \in \mathbb{C}$$

können die y_t als Elemente des Hilbertraumes \mathcal{H} der (bezüglich P) quadratisch integrierbaren komplexwertigen Funktionen aufgefaßt werden.

Der von den y_t , $t \in \mathbb{Z}$ erzeugte Unterraum von \mathcal{H} heißt Zeitbereich von y .

y heißt ergodisch, wenn

$$\lim_{T \rightarrow \infty} T^{-1} \sum_{k=0}^{T-1} y_{t+k} = 0$$

gilt.

Falls nicht ausdrücklich Gegenteiliges gesagt wird, sind Grenzwerte immer im Sinne der Konvergenz im quadratischen Mittel zu verstehen.

Als (linearen) Prognosefehler (bzw. Innovation) u_t (zum Zeitpunkt t) bezeichnet man die Differenz von y_t und der Projektion von y_t in die Vergangenheit, d.h. in den von den y_{t-1}, y_{t-2}, \dots aufgespannten Unterraum von \mathcal{H} .

Verschwinden alle Prognosefehler, dann heißt y deterministisch, sonst indeterministisch.

Ein indeterministischer Prozeß y heißt rein indeterministisch, wenn zu jedem Zeitpunkt t die von den

$$y_t, y_{t-1}, y_{t-2}, \dots \quad \text{bzw. den}$$

$$u_t, u_{t-1}, u_{t-2}, \dots$$

aufgespannten Unterräume von \mathcal{H} identisch sind.

Jeder indeterministische Prozeß läßt sich in eindeutiger Weise als Summe zweier unkorrelierter Prozesse, eines deterministischen und eines rein indeterministischen, darstellen.

Ist \underline{y} ein rein deterministischer Prozeß und \underline{u} sein Innovationsprozeß, dann existieren eindeutig bestimmte c_k , sodaß für alle $t \in \mathbb{Z}$

$$y_t = \sum_{k=0}^{\infty} c_k u_{t-k}$$

gilt.

I.2 DIE SPEKTRALDARSTELLUNG

Zu jeder Kovarianzfunktion

$$K: t \mapsto E y_0 \overline{y_t}, t \in \mathbb{Z}$$

eines stationären Prozesses \underline{y} existiert genau eine Verteilungsfunktion F mit

$$K(t) = \int_{[-\pi, \pi]} e^{i\lambda t} dF = \int_{[-\pi, \pi]} e^{i\lambda t} dF(\lambda), t \in \mathbb{Z}$$

und $F(\lambda) = 0$ für $\lambda \leq -\pi$, $F(\lambda) = K(0)$ für $\lambda \geq \pi$.

F heißt spektrale Verteilungsfunktion von \underline{y} .

\underline{y} ist genau dann ergodisch, wenn F an der Stelle 0 keine Sprungstelle besitzt.

Der Hilbertraum der (bezüglich F) quadratisch integrierbaren komplexwertigen Funktionen auf $[-\pi, \pi]$ heißt Spektralbereich von \underline{y} .

Die Funktion $e_t \mapsto y_t$, $t \in \mathbb{Z}$ läßt sich in eindeutiger Weise zu einer bijektiven Isometrie zwischen Spektralbereich und Zeitbereich fortsetzen.

Ein stationärer Prozeß \underline{y} ist also bis auf Isomorphie vollständig durch seine spektrale Verteilungsfunktion F bzw. - falls F absolut stetig ist - durch seine Spektraldichte $f = dF/d\lambda$ bestimmt.

Der Innovationsprozeß \underline{u} eines indeterministischen Prozesses ist weißes Rauschen, d.h.

$$Eu_t = 0, \quad Eu_s \overline{u_t} = \delta_{st} \sigma_u^2, \quad \delta_{st} = 1 \quad \text{für } s=t$$

und

$$\delta_{st} = 0 \quad \text{sonst}, \quad \sigma_u^2 > 0.$$

Weißes Rauschen ist rein indeterministisch und besitzt die Spektraldichte $\sigma_u^2 / (2\pi)$.

Ist \underline{y} indeterministisch, dann entspricht die Zerlegung von \underline{y} in einen deterministischen und einen rein indeterministischen Anteil (siehe I.1) genau der LEBESGUE Zerlegung von F in eine diskrete bzw. singuläre und eine absolut stetige Komponente.

Die dem rein indeterministischen Anteil zugeordneten c_k (siehe I.1) und die Varianz σ_u^2 seines Innovationsprozesses erfüllen und sind eindeutig bestimmt durch

$$(i) \quad c_0 = 1$$

$$(ii) \quad \sum_{k=0}^{\infty} |c_k|^2 < +\infty$$

$$(iii) \quad \sum_{k=0}^{\infty} c_k z^k \neq 0 \quad \text{für } |z| < 1$$

$$\text{und } (iv) \quad f = \frac{\sigma_u^2}{2\pi} \left| \sum_{k=0}^{\infty} c_k e^{-k} \right|^2.$$

I.3 ARMA PROZESSE

Ein stationärer Prozeß y heißt autoregressiver Moving Average (ARMA) Prozeß, wenn komplexe Zahlen $a_0, \dots, a_p, b_0, \dots, b_q$, von denen mindestens ein $a_k \neq 0$ und ein $b_l \neq 0$ sind, und ein weißes Rauschen u existieren, sodaß für alle $t \in \mathbb{Z}$

$$\sum_{k=0}^p a_k y_{t-k} = \sum_{l=0}^q b_l u_{t-l} \quad \text{gilt.}$$

O.B.d.A. können $p, q, a_0, \dots, a_p, b_0, \dots, b_q$ so gewählt werden, daß

(i) $a_0 = b_0 = 1, a_p, b_q \neq 0$ gilt,

(ii) die Polynome $A(z) = \sum_{k=0}^p a_k z^k$ und

$$B(z) = \sum_{l=0}^q b_l z^l \quad \text{relativ prim sind}$$

und (iii) alle Nullstellen von $A(z)$ außerhalb und keine Nullstellen von $B(z)$ innerhalb des Einheitskreises liegen.

(iii) ist notwendig und hinreichend dafür, daß u der Innovationsprozeß von y ist.

Unter den Bedingungen (i)-(iii) sind die Ordnung (p, q) und die Parameter $a_0, \dots, a_p, b_0, \dots, b_q$ eindeutig bestimmt.

Ein ARMA Prozeß ist rein indeterministisch und seine Spektraldichte ist gegeben durch

$$f = \frac{\sigma_u^2}{2\pi} \left| \frac{\sum_{l=0}^q b_l e^{-j\omega l}}{\sum_{k=0}^p a_k e^{-j\omega k}} \right|^2$$

Ein ARMA Prozeß der Ordnung $(p,0)$ heißt autoregressiver (AR) Prozeß, und ein ARMA Prozeß der Ordnung $(0,q)$ heißt Moving Average (MA) Prozeß.

Jeder Prozeß y mit absolut stetiger spektraler Verteilungsfunktion, insbesondere also jeder rein indetermi-

nistische Prozeß, kann durch AR, MA oder ARMA Prozesse approximiert werden, d.h. es existiert eine Folge

$(y^{(k)})_{k \in \mathbf{N}}$ von AR, MA oder ARMA Prozessen, sodaß gilt:

$$\lim_{k \rightarrow \infty} y^{(k)}_t = y_t \quad \forall t \in \mathbb{Z}$$

Für einen AR Prozeß gelten die YULE-WALKER-Gleichungen:

$$\sum_{k=0}^p a_k K(j-k) = \begin{cases} \sigma_u^2 & \text{für } j=0 \\ 0 & \text{für } j>0 \end{cases}$$

II. PARAMETRISCHE VERFAHREN: AR- UND ARMA-MODELLE

Einen parametrischen Schätzer für die einer Zeitreihe y_1, \dots, y_T zugrundeliegende Spektraldichte erhält man durch Schätzen eines AR- bzw. ARMA-Modelles und anschließendem Einsetzen der geschätzten Parameter in die Spektraldichtenformel aus I.3.

Im folgenden Abschnitt beschränken wir uns auf reellwertige Prozesse.

II.1 REKURSIVE SCHÄTZUNG

Zur Veranschaulichung der im folgenden auftretenden Rekursionen sei auf Anhang B verwiesen (siehe dazu auch FINDLEY (1981)).

II.1.1 Ordnungs- und Parameterschätzung von AR-Prozessen

Seien y_1, \dots, y_T Beobachtungen eines ergodischen AR Prozesses \underline{y} der Ordnung $p_0 (\leq P)$, und erfülle der Innovationsprozeß \underline{u} von \underline{y} die Bedingungen

$$E u_t^4 < +\infty$$

$$E(u_t | u_{t-1}, u_{t-2}, \dots) = 0 \quad \forall t \in \mathbb{Z}$$

$$E(u_t^2 | u_{t-1}, u_{t-2}, \dots) = \sigma_{\underline{u}}^2,$$

dann sind $p \in \{0, \dots, P\}$ und Lösungen

$$a_1(p), \dots, a_p(p), \sigma_{\underline{u}}^2(p) \text{ des Gleichungssystems}$$

$$1. \hat{K}(j) + \sum_{k=1}^p a_k(p) \hat{K}(j-k) = \delta_{j0} \sigma_{\underline{u}}^2(p), \quad j = 0 \dots p$$

konsistente Schätzer für die Ordnung p_0 und die Parameter a_1, \dots, a_{p_0} von \underline{y} , falls sie das

BIC-Kriterium

$$\log \sigma_{\underline{u}}^2(p) + p \frac{\log T}{T}$$

minimieren (siehe auch HANNAN & QUINN (1979)).

$\sigma_{\underline{u}}^2(p)$ und $a_1(p), \dots, a_p(p)$ können für $p=0 \dots P$ mit Hilfe der DURBIN (1960)-LEVINSON (1947)-Rekursionen

$$a_p(p) = - \sum_{k=0}^{p-1} a_k(p-1) \hat{K}(p-k) / \sigma_{\underline{u}}^2(p-1)$$

$$a_k(p) = a_k(p-1) + a_p(p) a_{p-k}(p-1), \quad k=1 \dots p-1$$

$$a_0(p) = 1$$

$$\sigma_{\underline{u}}^2(p) = (1 - a_p(p)^2) \sigma_{\underline{u}}^2(p-1)$$

auf ökonomische Weise aus den empirischen Autokovarianzen

$$\hat{K}(l) = \frac{1}{T-l} \sum_{t=1}^{T-l} y_t y_{t+l}, \quad l=0 \dots p$$

berechnet werden.

II.1.2. Ordnungs- und Parameterschätzung von ARMA Prozessen

Seien y_1, \dots, y_T Beobachtungen eines ergodischen ARMA Prozesses \underline{y} der Ordnung (p_0, q_0) mit $p_0 \leq P$, $q_0 \leq Q$; erfüllt der Innovationsprozeß \underline{u} von \underline{y} die Bedingungen

$$\begin{aligned} E u_t^4 &< +\infty \\ E(u_t | u_{t-1}, u_{t-2}, \dots) &= 0 \quad \forall t \in \mathbb{Z} \\ E(u_t^2 | u_{t-1}, u_{t-2}, \dots) &= \sigma_{\underline{u}}^2 \quad \text{und} \end{aligned}$$

liegt keine der Nullstellen $z_l, l=1, \dots, q_0$ von $B(z)$ am Einheitskreis, dann sind die OLS-Residuen \hat{u}_t einer Autoregression von y_t auf y_{t-1}, \dots, y_{t-L} ,

mit $L = [c \log T], \quad c = \frac{1}{2 \log \max_{1 \leq l \leq q_0} |z_l|},$

konsistente Schätzer für die Innovationen u_t . Weiters sind dann (p, q) und OLS-Lösungen

$$\begin{aligned} &a_1(p, q), \dots, a_p(p, q); b_1(p, q), \dots, b_q(p, q); \sigma_{\underline{u}}^2(p, q) \\ \text{von } y_t &= - \sum_{k=1}^p a_k(p, q) y_{t-k} + \sum_{l=1}^q b_l(p, q) \hat{u}_{t-l} + u_t(p, q) \end{aligned}$$

konsistente Schätzer für Ordnung (p_0, q_0) ,

Parameter a_1, \dots, a_p ; b_1, \dots, b_q und Innovationsvarianz σ_u^2 von \underline{y} , falls sie das BIC-Kriterium

$$\log \sigma_u^2(p, q) + (p+q) \frac{\log T}{T}$$

minimieren. (Siehe HANNAN & RISSANEN (1982) und die zugehörige Bemerkung (1983)).

Beschränkt man sich bei dieser Minimierung auf solche (p, q) , für die $p=q$ gilt, so erhält man einen konsistenten Schätzer für $\max\{p_0, q_0\}$. Die entsprechenden Schätzer für Innovationsvarianz und Parameter bleiben konsistent, selbst wenn $p_0 \neq q_0$ ist. Für den Fall $p=q$ ist die Berechnung der

$$a_1(p, q), \dots, a_p(p, q); b_1(p, q), \dots, b_q(p, q); \sigma_u^2(p, q)$$

aus den empirischen Kovarianzmatrizen

$$\hat{C}(l) := \frac{1}{T-l} \sum_{t=1}^{T-l} \begin{bmatrix} y_t \\ -\hat{u}_t \end{bmatrix} [y_{t+l}, -\hat{u}_{t+l}] \quad l=0, \dots, \max\{P, Q\}$$

mit Hilfe modifizierter WHITTLE(1963) - Rekursionen möglich:

$$B_p(p) = -D(p-1) G(p-1)^{-1}; \quad F_p(p) = -D(p-1)' g(p-1)^{-1}$$

$$g(p) = (I - B_p(p) F_p(p)) g(p-1), \quad G(p) = (I - F_p(p) B_p(p)) G(p-1)$$

$$g(0) = G(0) = \hat{C}(0);$$

$$B_l(p) = B_l(p-1) + B_p(p) F_{p-1}(p-1), \quad F_l(p) = F_l(p-1) + F_p(p) B_{p-1}(p-1) \\ l = 1, \dots, p-1$$

$$B_0(p) = F_0(p) = I,$$

$$D(p) = \sum_{l=0}^p B_p(p) \hat{C}(p+1-l)$$

$$a_1(p, q) = [F_1(p)]_{1,1}, \quad b_1(p, q) = [F_1(p)]_{1,2}, \quad l = 1 \dots p$$

$$\sigma_u^2(p, q) = [G(p)]_{1,1}.$$

II.2 ROBUSTE SPEKTRALSCHÄTZUNG

In den letzten Jahrzehnten griff die Suche nach robusten Methoden immer stärker um sich und erreichte schließlich selbst die entlegensten Grenzgebiete der Statistik. Auch in der Spektralanalyse existieren mehrere Ansätze zur Robustifizierung, von denen speziell einer im folgenden näher erläutert werden soll.

Die Problemstellung der robusten Statistik läßt sich am ehesten so eingrenzen:

Einem herkömmlichen Verfahren liegt eine Modellklasse M zugrunde, aus der es anhand der Daten ein nach einem festgelegten Kriterium optimales Modell m auswählt. Für die Dauer eines Gedankenspieles erweitert der Robustifizierer M zu einer Oberklasse M' . Angenommen, m' wäre nun das beste Modell und liege nicht in M , aber es gebe in M ein Modell \bar{m} , welches m' am "nächsten" liege. Soll man dann \bar{m} oder m verwenden, wie erhält man Verfahren, die \bar{m} statt m auswählen, wie weit können \bar{m} und m auseinanderliegen usw. ?

Damit haben wir das Gerüst für eine Formalisierung der robusten Spektralschätzung im engeren Sinne vorgegeben. Unser M setzt sich aus bestimmten stationären Prozessen zusammen, etwa AR-, MA-, ARMA-Prozessen oder solchen, die jene Regularitätsannahmen erfüllen, die wir jeweils für unsere Schätzmethode benötigen. Die erweiterte Modellklasse M' läßt dann auch die Möglichkeit von Ausreißern offen, d.h. von Daten, die ein vom Großteil der Stichprobe abweichendes Verhalten aufweisen (diese Definition ist ungenau und wird später korrigiert). Schätzen wir nun das Spektrum eines Prozesses in M , so wird sich das Ergebnis der Schätzung bei Vorliegen derartiger Ausreißer i.a. ändern. Die robuste Spektralschätzung geht von dem Axiom aus, daß die dabei entstehende Schätzung weniger wünschenswert ist als das

Spektrum desjenigen Prozesses, der nach Elimination bzw. Korrektur der Ausreißer entstehen würde, und ist bemüht, dieses Spektrum des zugrundeliegenden Prozesses zu gewinnen.

Die Aufgabenstellung der Identifizierung des zugrundeliegenden Prozesses bzw. seiner Charakteristika steht in engem Zusammenhang mit der Spezifizierung geeigneter Ausreißermodelle. Erst die Festlegung des Ausreißerverhaltens macht eine Annäherung an den zugrundeliegenden Prozeß überhaupt möglich. In der robusten Zeitreihenanalyse sind vor allem zwei Ausreißermodelle in Gebrauch, die in ihrer Urform auf FOX(1972) zurückgehen:

1. "innovations outliers" (IO): der Innovationsprozeß (u_t) ist bösartig verteilt. Solange er - und damit auch i.a. der beobachtete Prozeß (y_t) - endliche Varianzen aufweist, ergibt sich in der Spektralschätzung kein theoretisches Problem, lediglich ein schätztechnisches. Wachsen aber die zweiten Momente unbeschränkt, so liegt ein Spektrum im klassischen Sinn nicht mehr vor.

Die Betrachtung dieses Falles, gerade mit unendlichen Varianzen, wurde durch die Analyse von Börsenkursen und anderen spekulativen Preisen durch MANDELBROT inspiriert.

2. "additive outliers" (AO): die Daten selbst sind unverläßlich, die beobachteten Variablen x setzen sich aus einem dem jeweils spezifizierten Modell entsprechenden Prozeß y (d.h. etwa ARMA mit maximal 20 Lags) und einem Störprozeß s zusammen:

$$x_t = y_t + s_t \quad \text{und z.B.} \quad \sum_{k=0}^p A_k y_{t-k} = \sum_{k=0}^q B_k u_{t-k}$$

Die Störungen s_t werden als unkorreliert angenommen. Derartige Modelle sind meist dann gemeint, wenn die Suche nach dem zugrundeliegenden Prozeß y angesprochen wird.

Viele naturwissenschaftliche Prozesse rechtfertigen die Verwendung von AO-Modellen, die das Verhalten von Störungen in Meßinstrumenten oder das Auftreten von Tippfehlern wiedergeben. Besondere Schwierigkeiten folgen aus zeitlicher Korrelation der Störgrößen (sogenannte "patchy outliers"). Das Problem von Spektralschätzungen unter IO-Modellen fand bisher kaum Beachtung. Die hauptsächlichen Gründe hierfür wurden bereits erwähnt. Es sollte aber doch darauf hingewiesen werden, daß in diesem Fall parametrische und nichtparametrische Schätzprozeduren wesentlich andere Ergebnisse liefern können. Werden AR-Modelle in die Spektralschätzung eingebaut, so ist bekannt, siehe etwa HANNAN & KANTER (1977), daß Standardverfahren selbst bei unendlichen Varianzen die AR-Parameter konsistent schätzen und somit ein "Spektrum" produzieren. Die nähere Analyse des ARMA-Falles steht noch aus. Das entstehende Gebilde ist natürlich dann kein Spektrum im üblichen Sinn, da wir uns nicht mehr im L_2 bewegen und damit ein Großteil der theoretischen Grundlagen zusammenbricht, es handelt sich dabei wohl auch nicht um die einzige denkmögliche Verallgemeinerung des Spektrums auf L_p -Räume; trotzdem könnte man argumentieren, daß es Informationen auf brauchbare und noch genauer zu interpretierende Weise aufbereitet. Demgegenüber liefern alle Prozeduren, die das Spektrum über die zweiten Momente des Prozesses bestimmen, i.a. überhaupt kein Spektrum, weil die Schätzungen für diese Momente divergieren.

Das Problem von Spektralschätzungen unter AO-Modellen ist das Grundthema des Artikels von KLEINER & al. (1979), die zunächst feststellen, daß derartige Ausreißer üble Verzerrungen der Schätzer zur Folge haben.

Grundlage der nichtparametrischen Spektralschätzung ist das Periodogramm, eine empirische FOURIER-Transformierte der Kovarianzfunktion. Ein einziger Ausreißer

der Größe A ändert dieses Periodogramm um den Term

$$\frac{A^2}{n} + \frac{A}{n} \cdot 2\operatorname{Re}(X(\lambda)\exp(i2\pi\lambda))$$

(λ ist hier die Auswertungsstelle des Periodogramms, n die Größe der Stichprobe, X die FOURIER-Transformierte der Daten x_t).

Als Abhilfe empfehlen KLEINER & al. gemischte Prozeduren, die zunächst parametrisch AR- oder ARMA-Modelle schätzen und die entstehenden Residuen einer üblichen nichtparametrischen Spektralschätzung (geglättetes Periodogramm) unterwerfen. Der erste Schritt kann dann mittels einer robusten Prozedur durchgeführt werden - bei Autoregressionen stehen ja prinzipiell alle robusten Regressionsverfahren zur Verfügung - und auch die Residuen für den zweiten Schritt können so aufbereitet werden, daß sie nach Möglichkeit "ausreißerfrei" sind. Das Problem, die nichtparametrischen Spektralschätzer selbst zu robustifizieren, wird dabei umgangen.

Für die Schätzung der Spektraldichte ergibt sich somit die auf TUKEY zurückgehende Formel

$$\hat{f}(\lambda) = \frac{\hat{g}(\lambda)}{|H_p(\lambda)|^2} \quad \text{mit} \quad H_p(\lambda) = 1 - \sum_{j=1}^p \hat{\phi}_j \exp(i2\pi j\lambda)$$

Hier ist λ die Auswertungsstelle der Funktion, p die vorgegebene oder mittels eines handelsüblichen Kriteriums (z.B. BIC) vor-geschätzte Ordnung eines passenden AR-Prozesses $\hat{\phi}_j$ sind die mithilfe einer noch näher zu präzisierenden Prozedur geschätzten Parameter des AR-Prozesses, \hat{g} ist ein Schätzer für die Spektraldichte der Residuen unter den Parametern $\hat{\phi}_j$ im Sinne eines geglätteten Periodogramms. Eigentlich sollten für \hat{g} nicht die beobachteten Residuen

$$x_t - \sum_{j=1}^p \hat{\phi}_j x_{t-j} \quad ,$$

sondern diejenigen des zugrundeliegenden Prozesses verwendet werden. y_t ist aber nicht beobachtbar, Schätzungen \hat{y}_t sind notwendig. Das weiter unten beschriebene Verfahren liefert solche \hat{y}_t , im abschließenden Beispiel werden sie aber aus programmtechnischen Gründen nicht zur Konstruktion von \hat{g} herangezogen, dieser Ausdruck wird vielmehr konstant angenommen. Dies ist immer dann plausibel, wenn man darauf vertraut, daß das parametrische Modell den zugrundeliegenden Prozeß schon hinreichend gut bis auf weißes Rauschen erklärt. Selbstverständlich sind alle diese Überlegungen sofort auf den ARMA-Fall generalisierbar.

Zur Bestimmung der $\hat{\phi}_j$ stehen prinzipiell, wie schon erwähnt, alle robusten Regressionsmethoden zur Verfügung. Simulationen, etwa in DENBY & MARTIN (1979), zeigen hiebei, daß die sogenannten GM-(generalized maximum-likelihood-type) Schätzer unter einer breiten Auswahl von AO- und IO-Modellen die besten Eigenschaften aufweisen. Der GM-Schätzer ist definiert als Lösung des Minimierungsproblems

$$\sum_{t=1}^{T-p} W(y_t) \rho(y_{t+p} - \phi_1 y_{t+p-1} - \dots - \phi_p y_t) \rightarrow \min$$

Hier sind die ϕ_j die GM-Schätzer und ρ sowie W sind spezielle Gewichtsfunktionen. $\rho(x) = x^2$ und $W(x) = \text{const}$ ergibt den üblichen OLS-Schätzer, das Verfahren erfüllt seinen eigentlichen Zweck aber erst mit asymptotisch gegen 0 gerichteten Gewichtsfunktionen. Daneben sind für $\rho(x)$ auch Funktionen wie die sogenannte HUBER-Funktion

$$\rho_H(x) = \begin{cases} x^2 & \text{für } |x| \leq 1 \\ |x| & \text{für } |x| \geq 1 \end{cases}$$

in Betracht zu ziehen. Zur Lösung des Minimierungsproblems stehen iterative Computeralgorithmen zur Verfügung, die auf der näherungsweise Auflösung nichtlinearer Gleichungen beruhen. Mehrere Erwägungen lassen allerdings alle diese Prozeduren gegenüber den robusten Filtern unterlegen erscheinen. So fallen zum Beispiel Schätzwerte für den

zugrundeliegenden Prozeß nicht direkt ab, die Verwendung zweier Gewichtsfunktionen - gegenüber nur einer beim robusten Filtern - schafft neue Unsicherheiten bezüglich der Effizienz und der im folgenden geschilderte robuste Filter ist prinzipiell programmtechnisch schneller und angenehmer.

Die zweite von KLEINER & al. (1979) vorgeschlagene Methode ist also der robuste Filter, eine Modifikation des KALMAN-Filters und ursprünglich zur Elimination von Störsignalen erschaffen - d.h. mit dem Ziel der Glättung und nicht zur Strukturidentifikation mittels Glättung. Im einzelnen geht die Sache so vor sich:

- 1) aus den gegebenen Daten $x_t, t=1...T$ werden mittels OLS in

$$x_t = \sum_{j=1}^p \hat{\phi}_j x_{t-j} + u_t$$

Schätzwerte $\hat{\phi}_j, j=1...p$ ermittelt. Diese Schätzwerte dienen als Startwerte für die Iteration.

- 2) $mr/0.6745$, wobei mr der Median der absoluten Residuen aus dem vorigen Schritt ist, wird als Skalierung \hat{s} verwendet.

- 3) mittels der Formel

$$\hat{y}_t = \sum \hat{\phi}_j \hat{y}_{t-j} + \hat{s} \psi\left(\frac{x_t - \sum \hat{\phi}_j \hat{y}_{t-j}}{\hat{s}}\right)$$

werden neue Daten generiert, wobei die ersten p Werte unverändert bleiben. Die Formel ist rekursiv, und so werden sukzessive $\hat{y}_{p+1}, \dots, \hat{y}_T$ berechnet. Auf ψ wird noch eingegangen.

- 4) aus den neuen Daten $\hat{y}_t, t=1...T$ werden durch OLS neue Schätzwerte $\hat{\phi}_j$ in

$$\hat{y}_t = \sum \hat{\phi}_j \hat{y}_{t-j} + u_t$$

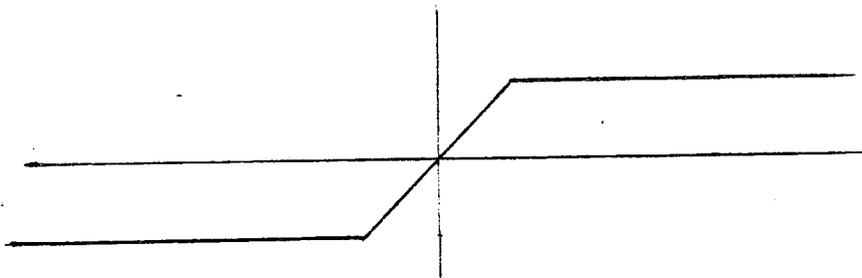
ermittelt. Wir gehen zu Schritt 2 zurück und iterieren

bis zur Konvergenz. Die Konvergenz der Prozedur ist theoretisch nicht sichergestellt.

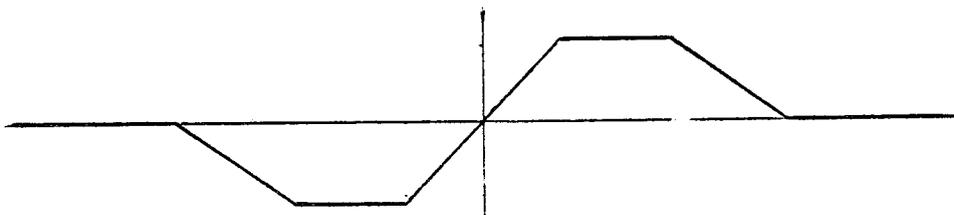
Bei ψ handelt es sich um eine Funktion, die der Ableitung der bei den GM-Schätzern erwähnten Gewichtsfunktion ρ entspricht. Dementsprechend erhält man für $\psi(x)=x$ immer die Originaldaten zurück, es ergibt sich der OLS-Schätzer und die Iteration wird überflüssig. Sinnvoll ist hier wieder HUBERS $\psi (= \rho'_H)$

$$\psi_H(x) = \begin{cases} \text{sgn}(x) & \text{für } |x| \geq 1 \\ x & \text{für } |x| \leq 1 \end{cases}$$

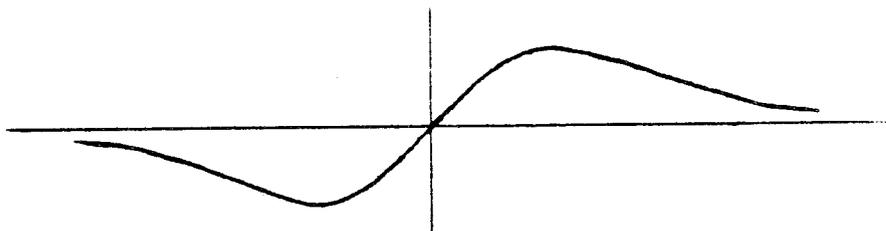
Daneben stehen ψ 's in Gebrauch, die großen $|x|$ 0 zuzuordnen ("hard redescenders") oder für die zumindest $\psi(x) \rightarrow 0$ für $|x| \rightarrow \infty$ gilt ("soft redescenders"). Schematisch haben diese drei Typen von Funktionen folgende Form:



HUBERS ψ



hard redescender



soft redescender

Alle diese Modifikationen unterstützen den Effekt, bei großen Abweichungen der Originaldaten x_t von ihren Prognosewerten $\sum \hat{\phi}_j \hat{y}_{t-j}$ diese Daten durch ihre Prognosewerte oder zumindest durch ein gewichtetes Mittel zu ersetzen. Damit ist auch eine einfache Motivation für das Verfahren gegeben: große Abweichungen zwischen Daten und Prognose werden als Ausreißer angesehen; im Gegensatz zu einem IO-Modell, in dem die Daten korrekt sind, werden die korrigierten Daten dann zur weiteren Prognose herangezogen.

Das Problem der Erweiterung auf den ARMA-Fall ist hier ein rein technisches, während es bei Regressionsprozeduren (M,GM) fundamentaler Natur ist. Es geht dabei nur darum, die Prognoseausdrücke für autoregressive Modelle durch ARMA-Prognosen zu ersetzen und Skalierungen entsprechend zu modifizieren. Bei MARTIN & al. (1981) findet sich ein Algorithmus, der genau dieses leistet. Die Umsetzung auf den Computer ist allerdings etwas aufwendig, da unter anderem Startwerte für Iterationen als Lösungen nichtlinearer Gleichungssysteme bestimmt werden müssen.

Auf der Grundlage des eben beschriebenen robusten Filterverfahrens wurde an der Rechenanlage des Institutes für Höhere Studien ein interaktives Fortran-Programm geschrieben, welches mit

```
  @use 8, dfile
  @add isa*meth.runi
  @xqt
```

aufgerufen ist. Auf dfile sollten die Daten stehen, in beliebigem vernünftigen Format, in gleichmäßiger Form und ohne führende Leerzeilen oder Header. Bevor die eigentlichen Rechenschritte erfolgen, findet ein Dialog über den Bildschirm z.B. in folgender Form statt (Meldungen des Systems zur Demonstration in Blockschrift):

NUMBER OF OBSERVATIONS ?
116 (Format i3)
HOW MANY LAG COEFFICIENTS ? (Ordnung des AR-Modells)
20 (Format i2)
VARIABLES FORMAT ? (Format der Daten auf dfile)
(2x,4f15.4)
DO YOU WANT TO USE FIRST DIFFERENCES ?
yes

Jetzt folgen die Iterationsschritte des robusten Filters.
Nach dem zehnten Schritt oder nach Konvergenz erscheint:

DO YOU WANT TO CONTINUE ?
no
DO YOU WANT TO SEE THE SPECTRUM ?
yes

An dieser Stelle erscheinen zwei Spektre: das autoregressive Spektrum der ungefilterten Daten (erster Iterationsschritt) und das autoregressive Spektrum der gefilterten Daten (Koeffizienten aus dem letzten Iterationsschritt), in beiden Fällen das logarithmierte Spektrum. Die nächsten Anweisungen werden verständlich, wenn man berücksichtigt, daß "meth" zur Prognose ökonomischer Zeitreihen erschaffen wurde:

DO YOU USE QUARTERLY DATA ?
yes
TIME OF FIRST VALUE ?
1954q1
oder z.B.:
DO YOU USE QUARTERLY DATA ?
no
TIME OF FIRST VALUE ?
0001 (Format i4)

Jetzt folgt ein tabellarischer Ausdruck der Originaldaten sowie der gefilterten (korrigierten) Daten. Im Anschluß daran können Prognosen aufgrund der korrigierten Daten und der geschätzten autoregressiven Koeffizienten veranstaltet werden. Themengemäß wollen wir uns aber auf die Spektralschätzungsfunktion des Programms konzentrieren und mit

DO YOU WANT TO MAKE FORECASTS ?

no

abschließen.

"meth" gebraucht den bereits beschriebenen robusten autoregressiven Filteralgorithmus, die Ψ -Funktion ist von der HUBER-Form. Die genaue Ordnung der Autoregression muß vorgegeben werden, die optimale Ordnung für ein A0-Modell ist natürlich unklar. Trotzdem dürften auf ungestörte Modelle ausgerichtete Ordnungsschätzungen mittels BIC oder ähnlichen Kriterien gute Hinweise bieten. Die Frage "do you use quarterly data ?" wird nur zur Dokumentation der Daten gestellt, eine Saisonbereinigung findet nicht statt. Eine derartige Saisonbereinigung würde sich auf die Prognoseerstellung komplizierend auswirken, neben den Unsicherheiten, die ein Saisonbereinigungsfiler in die robuste Prozedur bringen würde.

Das Programm wurde anhand österreichischer Bruttoinlandsprodukt-(BIP-)Daten 1954-1982 in Anspruch genommen. Da es sich hierbei um Quartalsdaten handelt, mußten in erster Linie Vielfache von 4 als mögliche Ordnungen der Autoregression p in Betracht kommen. Das Verhalten der Spektra bei unterschiedlichem p (12,16,20,24,28) legte 24 nahe. Bei größerem p verschwimmen die Unterschiede zwischen Original- und robustem Spektrum, bei kleinerem p glättet das robuste Spektrum vor allem die langfristigen Schwingungskomponenten vollkommen aus. In allen Fällen wurde vor der Inanspruchnahme des robusten Filters

die Transformation auf erste Differenzen durchgeführt, da das BIP einen klaren, beinahe linearen Trend aufweist. Nach sechs Iterationsschritten konvergiert das Verfahren.

Die im Anhang A abgebildeten Spektra zeigen deutliche Spitzen bei der Halbjahres- und der Jahresschwingung, das Saisonmuster dominiert die übrigen Frequenzen. Das robuste Spektrum beläßt von den vier mittleren langwelligen Spitzen im Originalspektrum lediglich eine, während die kurzfristigen Schwingungen deutlicher hervortreten. Die aufgelisteten Daten zeigen die größten Unterschiede um das Jahr 1972 herum, die korrigierten Daten liegen fast zur Gänze unter den Originalwerten. Der robuste Filter wirkt stationarisierend auf die ersten Differenzen und damit linearisierend auf den Trend.

Die Seiten 26 bis 29 geben den Output des Programms "meth" wieder. Sicher wird sich die Idee der robusten Spektralschätzung in Hinkunft noch viele begeisterte Anhänger erobern.

III. NICHTPARAMETRISCHE VERFAHREN

Seien $y_1 \dots y_T$ Beobachtungen eines rein indeterministischen Prozesses mit quadratisch integrierbarer Spektraldichte f .

f kann dann (im quadratischen Mittel) durch ihre FOURIER-Reihe

$$\sum_{k \in \mathbb{Z}} c_k e_k, \quad c_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\lambda) e^{-ik\lambda} d\lambda$$

dargestellt werden.

Wir betrachten Schätzer \hat{f} für f von der Form

$$\hat{f} = \sum_{|k| \leq M} w_k \hat{c}_k e_k, \\ \hat{c}_k = \frac{1}{2\pi} \frac{1}{T-|k|} \sum_{t=1}^{T-|k|} y_t y_{t+|k|}, \quad w_k \in \mathbb{R}.$$

Offensichtlich sind die \hat{c}_k erwartungstreue Schätzer für die FOURIER-Koeffizienten c_k .

$(w_k)_{k \in \{-M, \dots, M\}}$ heißt Lagfenster, wenn $w_{-k} = w_k$ und $w_0 = 1$ ist.

Der mean integrated square error (MISE) eines solchen \hat{f} ist gegeben durch

$$\begin{aligned} & E \frac{1}{2\pi} \int_{-\pi}^{\pi} |f(\lambda) - \hat{f}(\lambda)|^2 d\lambda \\ &= E \frac{1}{2\pi} \int_{-\pi}^{\pi} \left| \sum_{k \in \mathbb{Z}} c_k e^{ik\lambda} - \sum_{|k| \leq M} w_k \hat{c}_k e^{ik\lambda} \right|^2 d\lambda \\ &= E \frac{1}{2\pi} \int_{-\pi}^{\pi} \left| \sum_{|k| \leq M} (c_k - w_k \hat{c}_k) e^{ik\lambda} + \sum_{|k| > M} c_k e^{ik\lambda} \right|^2 d\lambda \\ &= \sum_{|k| \leq M} E |c_k - w_k \hat{c}_k|^2 + \sum_{|k| > M} |c_k|^2. \end{aligned}$$

Der von w_k , $k \neq 0$, unabhängige Beitrag

$$\begin{aligned} & E |c_k - w_k \hat{c}_k|^2 - |c_k|^2 \\ &= \text{Var}(w_k \hat{c}_k) + |c_k - w_k c_k|^2 \\ &= w_k^2 (\text{Var}(\hat{c}_k) + |c_k|^2) - 2w_k |c_k|^2 \end{aligned}$$

des k -ten Terms $w_k \hat{c}_k e_k$ zum MISE ist genau dann negativ, wenn

$$0 < w_k < \frac{2|c_k|^2}{\text{Var}(\hat{c}_k) + |c_k|^2} \quad \text{gilt,}$$

und ist für

$$w_k = w_k^* = \frac{|c_k|^2}{\text{Var}(\hat{c}_k) + |c_k|^2}$$

minimal.

Für Wahrscheinlichkeitsdichten f mit Träger in $[-\pi, \pi]$ ist obige Prozedur eine Verallgemeinerung der Verfahren in KRONMAL & TARTER (1968), TARTER & KRONMAL (1970). *) Da die w_k^* in den Anwendungen unbekannt sind und nicht abzusehen ist, inwieweit die Optimalitätseigenschaften verändert werden, wenn man sie durch geschätzte Werte \hat{w}_k^* ersetzt, geben wir im folgenden noch einige vom jeweiligen Prozeß unabhängige Lagfenster an, die in der Praxis häufig verwendet werden:

BARTLETT : $w_k = (1 - \frac{|k|}{M})$

BARTLETT modifiziert: $w_k = (1 - \frac{|k|}{T}) (1 - \frac{|k|}{M})$

Hanning: $w_k = (1 - \frac{|k|}{T}) (\frac{1}{2} + \frac{1}{2} \cos(\frac{k\pi}{M}))$

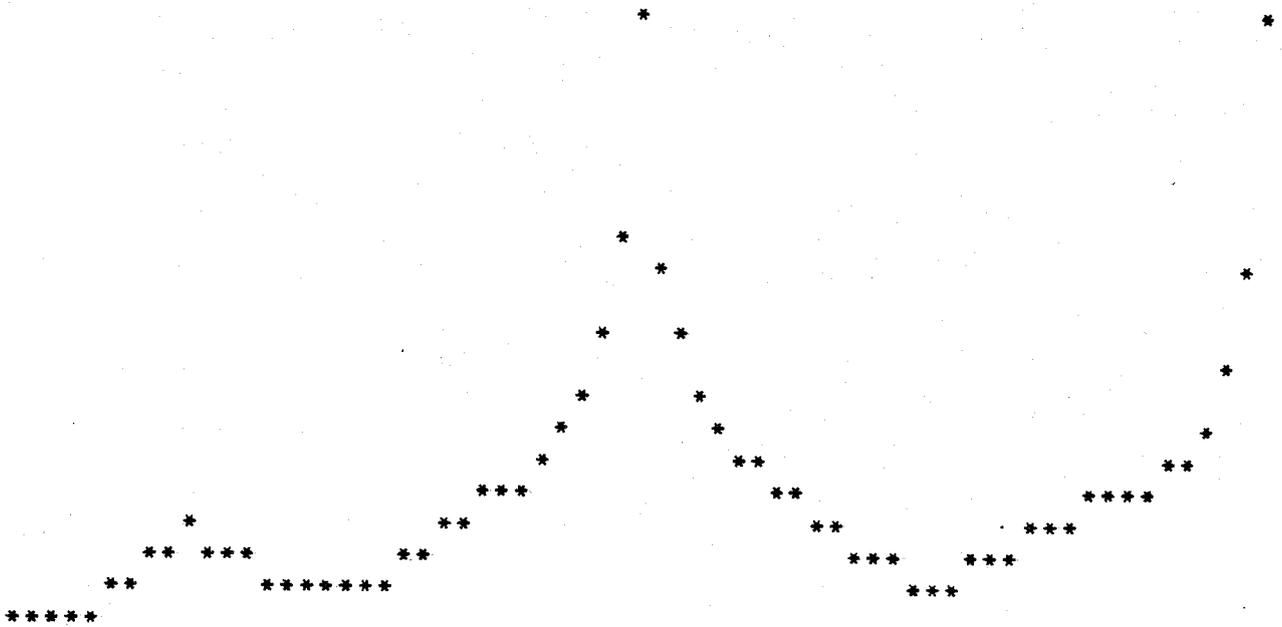
Hamming: $w_k = (1 - \frac{|k|}{T}) (0.54 + 0.46 \cos(\frac{k\pi}{M}))$

PARZEN: $w_k = (1 - \frac{|k|}{T}) (1 - (\frac{k}{M})^2)$

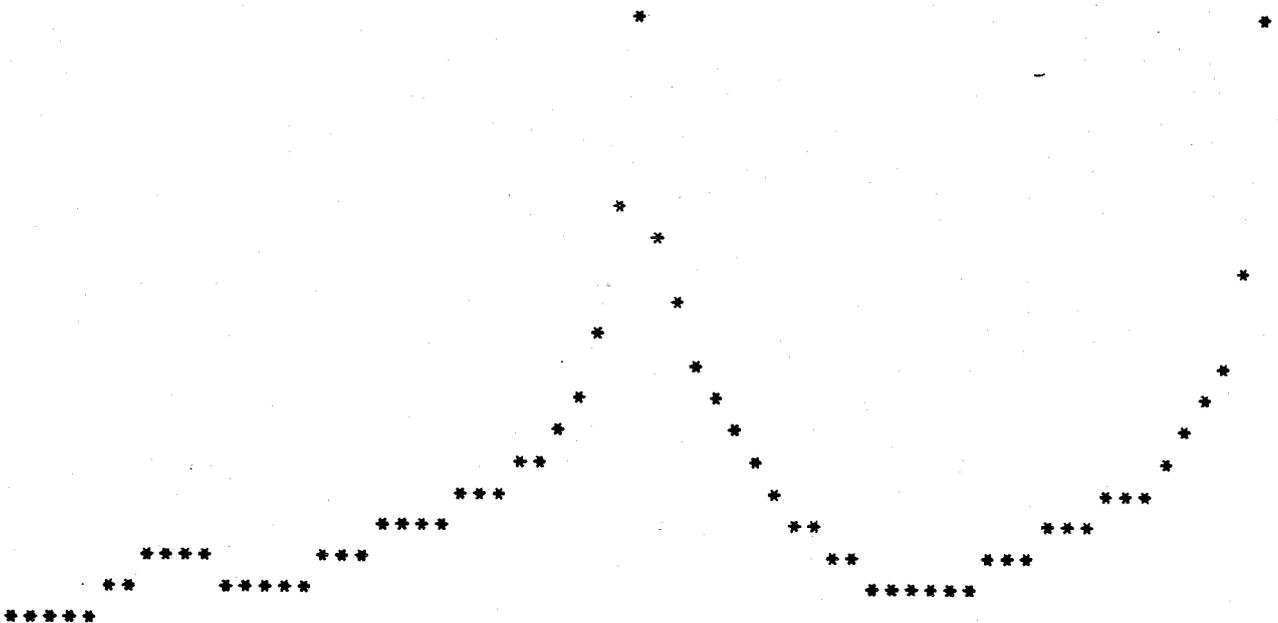
*) Siehe dazu auch TARTER & KRONMAL (1976) !

Anhang A

An dieser Stelle erscheinen exemplarische Ausdrücke des logarithmierten Spektrums der ersten Differenzen des BIP . Zunächst sehen wir die in II.2 angekündigten Resultate des Programmes "meth".



AR(12)-Modell für die Originaldaten



AR(12)-Modell für die korrigierten Daten

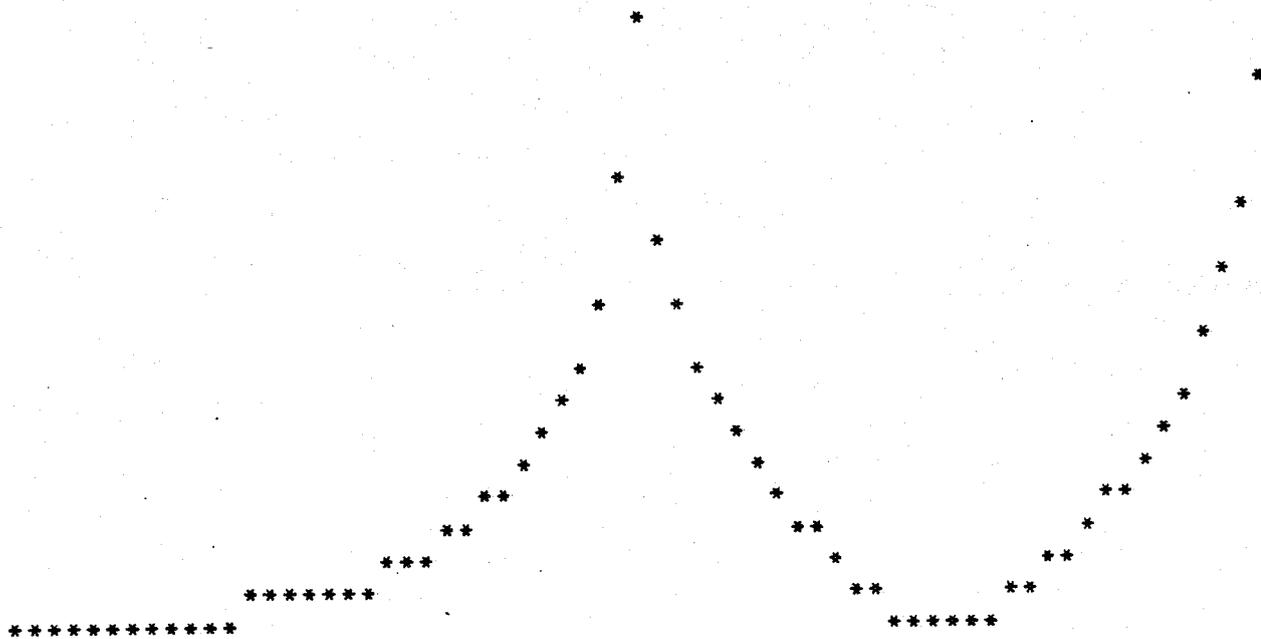


AR(24)-Modell für die Originaldaten

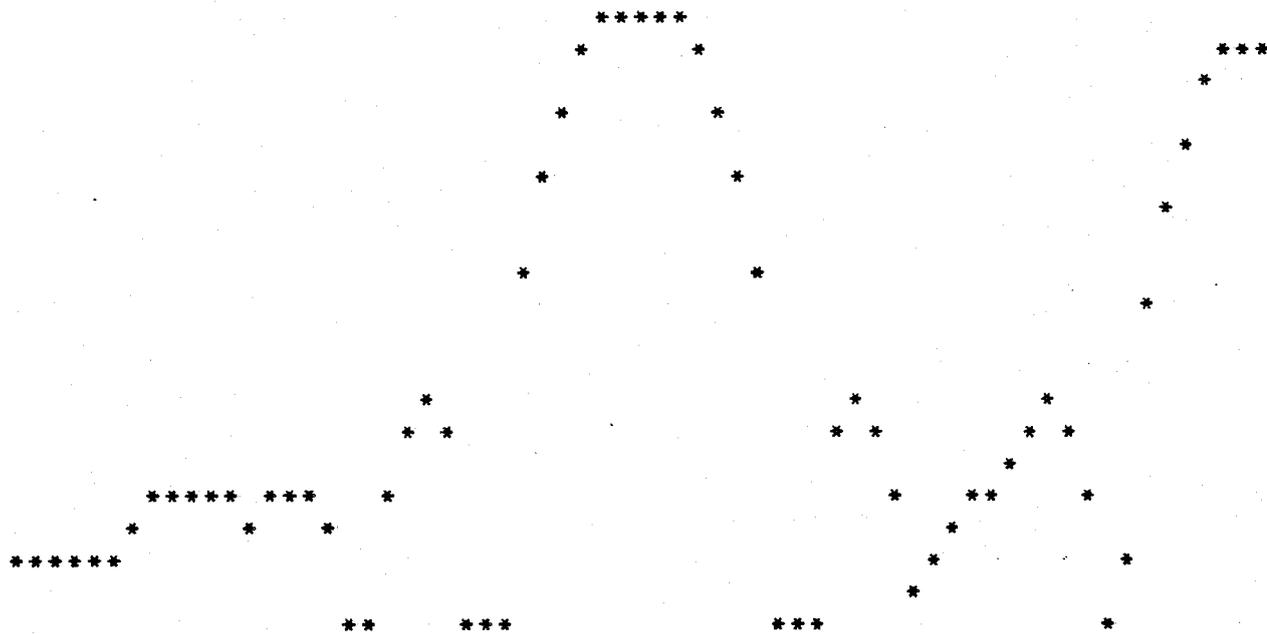


AR(24)-Modell für die korrigierten Daten

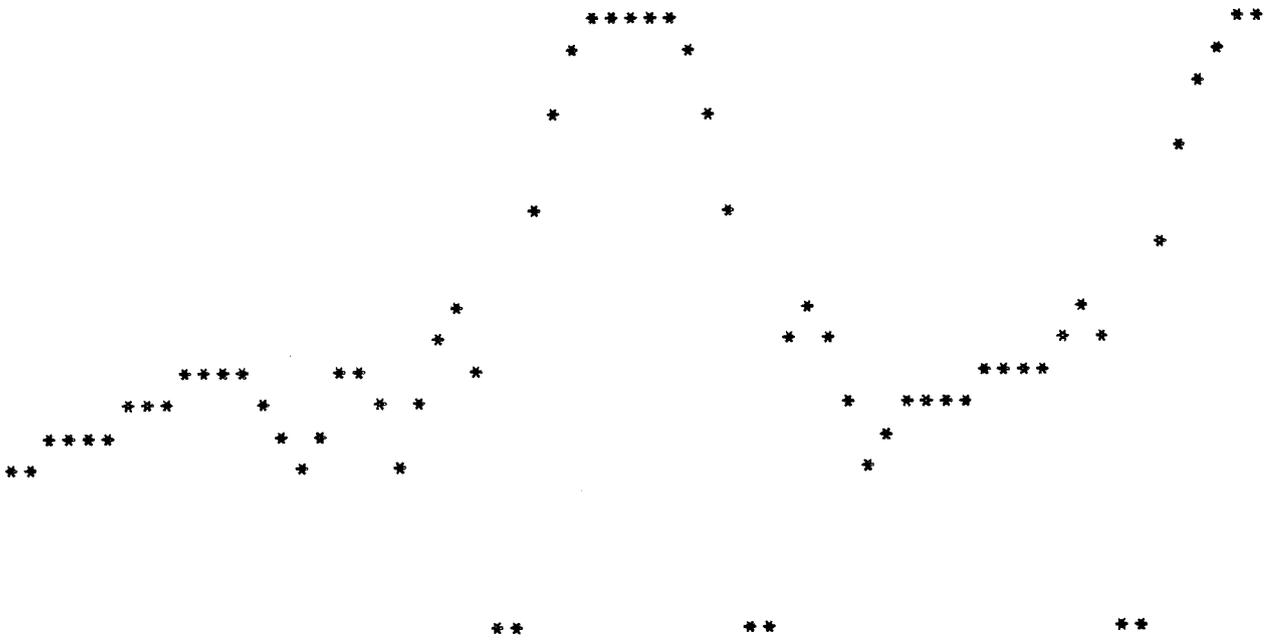
Zum Vergleich bilden wir jetzt mit anderen Methoden
geschätzte Spektren ab :



ARMA(6,6)-Model1



Hanning-Fenster der Länge M = 20



Hanning-Fenster der Länge $M = 24$



Hanning-Fenster der Länge $M = 28$

Die BIP-Daten entstammen der WIFO-Datenbank.

Anhang B

Im folgenden sei mit $\langle x, y \rangle := E x \bar{y}$ das innere Produkt in \mathcal{X} bezeichnet.

Für $m > 0$ seien

(i) P_t^m bzw. P_t^{-m} die Orthoprojektoren auf die von $\{y_{t-1}, \dots, y_{t-m}\}$ bzw. $\{y_{t+1}, \dots, y_{t+m}\}$ aufgespannten Teilräume des Zeitbereichs eines indeterministischen stationären Prozesses $\underline{y} = (y_t)_{t \in \mathbb{Z}}$,

(ii) $u_t^m := y_t - P_t^m y_t$ bzw. $u_t^{-m} := y_t - P_t^{-m} y_t$ die (wegen der Indeterminiertheit von \underline{y}) nicht verschwindenden Prognosefehler bei Vorwärtsregression (= Rückwärtsprojektion) bzw. Rückwärtsregression (= Vorwärtsprojektion),

(iii) $\sigma_m^2 := \|u_t^m\|^2$ bzw. $\sigma_{-m}^2 := \|u_t^{-m}\|^2$ die (wegen der Stationarität \underline{y} s) von t unabhängigen Varianzen der Prognosefehler und

(iv) $\rho_m := \frac{\langle y_t, u_{t-m}^{-(m-1)} \rangle}{\sigma_{-(m-1)}^2}$ bzw.

$\rho_{-m} := \frac{\langle y_t, u_{t+m}^{m-1} \rangle}{\sigma_{m-1}^2}$ die (wegen der

Stationarität von \underline{y}) von t unabhängigen partiellen Korrelationen.

Für $m = 0$ setzen wir $u_t^0 := y_t$, $\sigma_0^2 := K(0) = \|y_t\|^2$.

Berücksichtigt man, daß $\{u_{t-1}^0, u_{t-2}^{-1}, \dots, u_{t-m}^{-(m-1)}\}$

bzw. $\{u_{t+1}^0, u_{t+2}^1, \dots, u_{t+m}^{m-1}\}$ die GRAM-SCHMIDT-Orthogonalisierungen von $\{y_{t-1}, \dots, y_{t-m}\}$ bzw.

$\{y_{t+1}, \dots, y_{t+m}\}$ sind, so erhält man unmittelbar die Darstellungen

$$P_t^m y_t = \sum_{k=1}^m \rho_k u_{t-k}^{-(k-1)} \quad \text{und}$$

$$P_t^{-m} y_t = \sum_{k=1}^m \rho_{-k} u_{t+k}^{k-1}$$

Die folglich für die Prognosefehler bestehenden rekursiven Beziehungen

$$u_t^{m+1} = u_t^m - \rho_{m+1} u_{t-(m+1)}^{-m} \quad \text{und}$$

$$u_t^{-(m+1)} = u_t^{-m} - \rho_{-(m+1)} u_{t+m+1}^m$$

lassen sich wegen

$$\langle u_t^m, u_{t-(m+1)}^{-m} \rangle = \langle y_t, u_{t-(m+1)}^{-m} \rangle = \rho_{m+1} \sigma_{-m}^2$$

und

$$\langle u_t^{-m}, u_{t+(m+1)}^m \rangle = \langle y_t, u_{t+m+1}^m \rangle = \rho_{-(m+1)} \sigma_m^2$$

auf deren Varianzen übertragen:

$$\sigma_{m+1}^2 = \sigma_m^2 - |\rho_{m+1}|^2 \sigma_{-m}^2 \quad \text{und}$$

$$\sigma_{-(m+1)}^2 = \sigma_{-m}^2 - |\rho_{-(m+1)}|^2 \sigma_m^2.$$

Ist nun

$$P_t^m y_t = \sum_{k=1}^m \alpha_k^m y_{t-k} \quad \text{und}$$

$$P_t^{-m} y_t = \sum_{k=1}^m \beta_k^m y_{t+k} ,$$

so gilt

$$\begin{aligned} \rho_{m+1} &= \frac{1}{\sigma_{-m}^2} \langle y_t, y_{t-(m+1)} - \sum_{k=1}^m \beta_k^m y_{t-(m+1)+k} \rangle \\ &= \frac{K(-m-1) - \sum_{k=1}^m \beta_k^m K(-m-1+k)}{\sigma_{-m}^2} \end{aligned}$$

und analog

$$\rho_{-(m+1)} = \frac{K(m+1) - \sum_{k=1}^m \alpha_k^m K(m+1-k)}{\sigma_m^2} .$$

Mit

$$\alpha_k^{m+1} = \alpha_k^m - \rho_{m+1} \beta_{m-k+1}^m , \quad k = 1, \dots, m ,$$

$$\alpha_{m+1}^{m+1} = \rho_{m+1} ,$$

$$\beta_k^{m+1} = \beta_k^m - \rho_{-(m+1)} \alpha_{m-k+1}^m , \quad k=1, \dots, m ,$$

$$\beta_{m+1}^{m+1} = \rho_{-(m+1)}$$

gilt weiters

$$\begin{aligned} & \sum_{k=1}^{m+1} \alpha_k^{m+1} y_{t-k} = \\ = & \sum_{k=1}^m \alpha_k^m y_{t-k} - \rho_{m+1} \sum_{k=1}^m \beta_{m-k+1}^m y_{t-k} + \rho_{m+1} y_{t-(m+1)} = \\ = & p_t^m y_t + \rho_{m+1} (y_{t-(m+1)} - \sum_{j=1}^m \beta_j^m y_{t-(m+1)+j}) = \\ = & p_t^m y_t + \rho_{m+1} u_{t-(m+1)}^{-m} = \sum_{k=1}^{m+1} \rho_k u_{t-k}^{-(k-1)}, \text{ also} \\ p_t^{m+1} y_t & = \sum_{k=1}^{m+1} \alpha_k^{m+1} y_{t-k} \quad \text{und analog} \\ p_t^{-(m+1)} y_t & = \sum_{k=1}^{m+1} \beta_k^{m+1} y_{t+k} \end{aligned}$$

LITERATUR

=====

- DENBY, L. & R.D. MARTIN (1979) : Robust estimation of the
First-Order Autoregressive Parameter
Journal of the American Statistical Association 74, 140-146
- DURBIN, J. (1960) : The fitting of Time Series Models
Revue de l'institut international de statistique 28,
233-244
- FINDLEY, D.F. (1981) : Geometrical and Lattice Versions of
Levinson's General Algorithm, in: D.F. Findley (ed.),
Applied Time Series Analysis II, Academic Press
- FOX, A.J. (1972) : Outliers in Time Series
Journal of the Royal Statistical Society B 34, 350-363
- HANNAN, E.J. & M. KANTER (1977) : Autoregressive Processes
with Infinite Variance
Journal of Applied Probability 14, 411-415
- HANNAN, E.J. & B.G. QUINN (1979) : The Determination of the
Order of an Autoregression
Journal of the Royal Statistical Society B 41, 190-195
- HANNAN, E.J. & J. RISSANEN (1982) : Recursive Estimation of
Mixed Autoregressive-Moving Average Order,
Biometrika 67, 81-94
- (1983) : Amendments and Corrections
Biometrika 70, 303

- KLEINER, B., R.D. MARTIN & D.J. THOMSON (1979) :
Robust Estimation of Power Spectra
Journal of the Royal Statistical Society B 41, 313-351
- KRONMAL, R.A. & M.E. TARTER (1968) : The Estimation of
Probability Densities and Cumulatives by
Fourier Series Methods
Journal of the American Statistical Association 63, 925-952
- LEVINSON, N (1947) : The Wiener RMS (Root-Mean-Square) Error
Criterion in Filter Design and Prediction
Journal of Mathematical Physics 25, 261-278
- MARTIN, R.D., A. SAMAROV & W. VANDAELE (1981) :
Robust Methods for ARIMA Models
Massachusetts Institute of Technology,
Technical Report No. 29
- TARTER, M.E. & R.A. KRONMAL (1970) : On Multivariate
Density Estimates Based on Orthogonal Expansions
The Annals of Mathematical Statistics 41, 718-722
- (1976) : An Introduction to the Implementation and
Theory of Nonparametric Density Estimation
The American Statistician 30, 105-111
- WHITTLE, P. (1963) : On the Fitting of Multivariate
Autoregressions, and the Approximate Canonical
Factorisation of a Spectral Density Matrix
Biometrika 50, 129-134