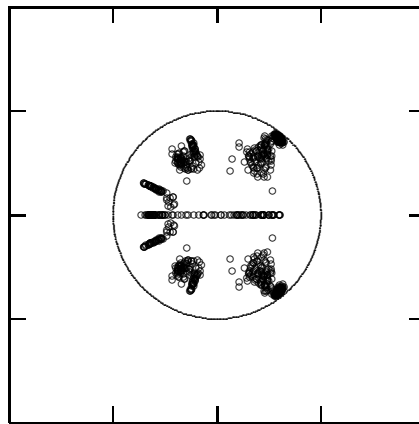


**Automatische Schätzung
von
Spektren**

Günter Edenharter



Die endliche Existenz vermag nur
Fragmentarisches zu erreichen,
bleibt immer unterwegs
und gelangt nie ans Ende.

Vorwort

Die vorliegende Arbeit ist eine geringfügig erweiterte und überarbeitete Fassung meiner gleichnamigen Dissertation, die bereits vorab auf Microfiche veröffentlicht wurde.

Ich danke Herrn Prof. Dr. Friedrich und Herrn Prof. Dr. Wolters für die zahlreichen Anregungen und fruchtbaren Diskussionen, sowie für die freundliche Betreuung während der Durchführung meiner Dissertation.

Frau Diplom-Volkswirtin Claudia Funke danke ich für ihre Hilfe bei der Literaturbeschaffung sowie der aufmerksamen Durchsicht von Teilen der Arbeit.

Zur Erstellung der schriftlichen Vorlage diente das Satzsystem $\text{T}_{\text{E}}\text{X}$ in der Implementierung `emtex 3.141` mit den Makropaketen $\text{AMSI}\text{T}_{\text{E}}\text{X}$ und $\text{PICI}\text{T}_{\text{E}}\text{X}$.

Berlin-Tiergarten, August 1995

Günter Edenharter

Inhaltsverzeichnis

1	Einleitung	1
2	Grundlagen stochastischer Prozesse	6
2.1	Stochastische Prozesse	6
2.2	Spektraldarstellung stochastischer Prozesse	11
2.3	Lineare Filter und stochastische Prozesse	16
2.4	Lineare Filter und stochastische Prozesse im Frequenzbereich	30
2.5	Schätzung von Momentfunktionen	35
2.6	Entropie, Information und stochastische Prozesse	44
3	Autoregressive Spektralschätzung	56
3.1	Autoregressive Prozesse und lineare Prädiktion	58
3.2	Levinson-Durbin-Rekursion	62
3.3	AR-Prozesse und das Prinzip der maximalen Entropie	66
3.4	Autoregressive Schätzverfahren	71
3.4.1	Yule-Walker-Schätzung	74
3.4.2	Burg-Schätzung	77
3.4.3	FBLP-Schätzung	80
3.4.4	FLP-Schätzung	83
3.4.5	Vergleich autoregressiver Schätzverfahren	84
4	Nichtparametrische Spektralschätzung	96
4.1	Das Periodogramm	96
4.2	Direkte Spektralschätzer	100
4.3	Indirekte Spektralschätzer	109

5	Diskrepanzmaße für Spektraldichtefunktionen	118
5.1	Evaluierung klassischer Diskrepanzmaße	119
5.2	Global korrigierte Diskrepanzmaße	131
6	Automatische Schätzung von Spektren	140
6.1	Modellselektionskriterien für autoregressive Spektralschätzungen	140
6.2	Modellselektionskriterien für nichtparametrische Spektralschätzungen .	151
6.3	Simulation	168
6.4	Anwendungsbeispiele	200
6.4.1	Ökonomie: Festgeldzinsen	200
6.4.2	Endokrinologie: Luteinisierendes Hormon (LH)	207
6.4.3	Astronomie: Sonnenfleckenindex	214
7	Zusammenfassung	221
A	Toeplitzmatrizen	224
	Literatur	233

Abbildungsverzeichnis

2.1	Sprungstellen im Phasendiagramm eines Filters	32
3.1	Spektrum und Wurzeln des AR(8)-Prozesses (3.70)	89
3.2	Auflösung, gemittelte Spektren, AR(8)-Prozeß (3.70), $N = 64$	90
3.3	Auflösung, Wurzelplots, AR(8)-Prozeß (3.70), $N = 64$	91
3.4	Auflösung, gemittelte Spektren, AR(8)-Prozeß (3.70), $N = 256$	92
3.5	Auflösung, Wurzelplots, AR(8)-Prozeß (3.70), $N = 256$	93
3.6	Spectral-Line-Splitting	94
3.7	Peak-Shifting	95
5.1	Diskrepanzvergleich: Daniellschätzung und Parzenschätzung	122
5.2	(a) - (b) Diskrepanzvergleich von theoretischen Spektraldichten	127
5.2	(c) - (d) (Fortsetzung)	128
6.1	(a) Spektrum des AR(4)-Prozesses (6.8), (b) Referenzdiskrepanz	162
6.2	(a) - (d) unkorrigierte/korrigierte Modellselektionskriterien	163
6.2	(e) - (h) (Fortsetzung)	164
6.3	EDF bei iterativer Hamming-Filterung	165
6.4	(a) Referenzdiskrepanz, (b) - (c) Kreuzvalidierungskriterien	166
6.5	(a) - (d) Nichtpara. Kriterien, Hamming, AR(4)-Prozeß (6.8)	167
6.6	(a) Theor. Spektrum AR(8)-Prozeß (6.27) (b) Wurzelplot	176
6.7	(a) - (c) Ref.-Diskr. für AR(8)-Prozeß (6.27)	177
6.8	(a) - (d) Para. Kriterien, AR(8)-Prozeß (6.27)	178
6.9	(a) - (d) Nichtpara. Kriterien, Hamming, AR(8)-Prozeß (6.27)	179
6.10	(a) - (d) Nichtpara. Kriterien, Parzen, AR(8)-Prozeß (6.27)	180
6.11	(a) - (c) Boxplots, AR(8)-Prozeß (6.27)	181
6.12	(a) Theor. Spektrum AR(2)-Prozeß (6.28) (b) Wurzelplot	182

6.13 (a) - (c) Ref.-Diskr. AR(2)-Prozeß (6.28)	183
6.14 (a) - (d) Para. Kriterien, AR(8)-Prozeß (6.28)	184
6.15 (a) - (d) Nichtpara. Kriterien, Hamming, AR(8)-Prozeß (6.28)	185
6.16 (a) - (d), Nichtpara. Kriterien, Parzen, AR(8)-Prozeß (6.28)	186
6.17 (a) - (c) Boxplots, AR(2)-Prozeß (6.28)	187
6.18 (a) Theor. Spektrum AR(8)-Prozeß (6.29) (b) Wurzelplot	188
6.19 (a) - (c) Ref.-Diskr. AR(8)-Prozeß (6.29)	189
6.20 (a) - (d) Param. Kriterien, AR(8)-Prozeß(6.29)	190
6.21 (a) - (d) Nichtpara. Kriterien, Hamming, AR(8)-Prozeß (6.29)	191
6.22 (a) - (d) Nichtpara. Kriterien, Parzen, AR(8)-Prozeß (6.29)	192
6.23 (a) - (c) Boxplots, AR(8)-Prozeß (6.29)	193
6.24	194
6.25 (a) - (c) Ref.-Diskr. für MA(1)-Prozeß (6.30)	195
6.26 (a) - (d) Para. Kriterien, MA(1)-Prozeß (6.30)	196
6.27 (a) - (d) Nichtpara. Kriterien, Hamming, MA(1)-Prozeß (6.30)	197
6.28 (a) - (d) Nichtpara. Kriterien, Parzen, MA(1)-Prozeß (6.30)	198
6.29 (a) - (c), Boxplots, MA(1)-Prozeß (6.30)	199
6.30 Festgeldzins, (a) Originalreihe, (b) Emp. Autokorrelationsfunktion	202
6.31 Festgeldzins, (a) AIC, (b) AICC, (c) SIC, (d) SICC	203
6.32 Festgeldzins, (a) AIC, AICC, SIC: Spektrum, (b) SICC: Spektrum	204
6.33 Festgeldzins, (a) FAIC, (b) FAICC, (c) FSIC, (d) FHQ	205
6.34 Festgeldzins, (a) - (d), FAIC, FAICC, FSIC, FHQ: Spektren	206
6.35 LH, (a) Originalreihe, (b) Emp. Autokorrelationsfunktion	209
6.36 LH, (a) AIC, (b) AICC, (c) SIC, (d) SICC	210
6.37 LH, (a) - (b) AIC, AICC: Spektren, (c) SICC, SIC: Spektrum	211
6.38 LH, (a) FAIC, (b) FAICC, (c) FSIC, (d) FHQ	212
6.39 LH, (a) - (d), FAIC, FAICC, FSIC, FHQ: Spektren	213
6.40 Sonnenfleckenindex, (a) Aggregierte Reihe, (b) Emp. Autokorrelations- funktion	216
6.41 Sonnenfleckenindex, (a) AIC, (b) AICC, (c) SIC, (d) SICC	217
6.42 Sonnenfleckenindex, (a) AIC, AICC, SIC, SICC: Spektrum	218
6.43 Sonnenfleckenindex, (a) FAIC, (b) FAICC, (c) FSIC, (d) FHQ	219
6.44 Sonnenfleckenindex, (a) - (d), FAIC, FAICC, FSIC, FHQ: Spektren	220

Tabellenverzeichnis

4.1	Lagfenstergeneratoren: Varianz- und Biaskomponente	116
4.2	Lagfenstergeneratoren: EBW und EDF	117
5.1	Vergleich Daniell- u. Parzenschätzer	123
5.2	Referenzrangfolge für theoretische Spektraldichten	126
5.3	Diskrepanzvergleich von AR(8)-Prozessen im Zeitbereich	129
5.4	Diskrepanzvergleich von AR(8)-Spektren im Frequenzbereich	130
5.5	Dissimilaritäten, AR(8)-Prozesse im Frequenzbereich	133
5.6	AR(8)-Prozesse, global korrigierter ISE	138
5.7	AR(8)-Prozesse, global korrigiertes KLIC	139
6.1	Vergleich globaler Minima	160
6.2	Medianvergleich der global korrigierten Diskrepanz	172

Symbolverzeichnis

$a(L)$	Filter als Lagoperatorpolynom
\mathcal{C}	Menge der komplexen Zahlen
c_τ	empirische Autokorrelationsfunktion
$\text{Cov}(X, Y)$	Kovarianz der Zufallsvariablen X und Y
$\text{Corr}(X, Y)$	Korrelation der Zufallsvariablen X und Y
δ_t	Einheitsimpuls
$E(X)$	Erwartungswert der ZV X
EBW	äquivalente Bandbreite
EDF	äquivalente Zahl von Freiheitsgraden, ν
\mathcal{F}	Filter
$f, f(x), f_X(x)$	Dichtefunktion
$F(x), F_X(x)$	Verteilungsfunktion
$f, f(\omega), f_X(\omega)$	Spektraldichtefunktion
$F_X(\omega), F(\omega)$	Spektralverteilungsfunktion
$\gamma_X(\tau), \gamma(\tau), \gamma_\tau$	Kovarianzfunktion
$h(\omega), h_X(\omega)$	normierte Spektraldichtefunktion
$H(\omega), H_X(\omega)$	normierte Spektralverteilungsfunktion
$I(\omega)$	Periodogramm
$\tilde{I}(\omega)$	empirisches Periodogramm
$\text{Inf}(X)$	Selbstinformation
L	Lagoperator
r_τ	empirische Autokorrelation
$\rho_X(\tau), \rho(\tau), \rho_\tau$	Autokorrelationsfunktion
σ^2	Varianz
\mathcal{T}	Indexmenge
$S(X)$	Entropie
μ_X, μ	Erwartungswert
\mathbf{M}	Matrix
\mathbf{M}'	transponierte Matrix
\mathbf{M}^H	transponiert-konjugiert-komplexe Matrix
\mathcal{N}	Menge der natürlichen Zahlen

N	Stichprobenumfang
N_ω	Anzahl von Frequenzen im Intervall $[0.0, 0.5]$ oder $[-0.5, 0.5]$
N_E	Anzahl der Einschwingwerte
P	Einschritt-Prognosefehler
\mathcal{R}	Menge der reellen Zahlen
$\text{Var}(X)$	Varianz der Zufallsvariablen X
ω	Frequenz
T	Stützpunkt
\mathcal{Z}	Menge der ganzen Zahlen
z	komplexe Zahl
z^*	konjugiert komplexe Zahl zu z
$ z $	Betrag von z

Schätzer werden durch ein $\hat{}$ gekennzeichnet: z.B. $\hat{\rho}_\tau, \hat{f}(\omega)$ usw.

Es wurden nur die wichtigsten Symbole in das Symbolverzeichnis aufgenommen.

Kapitel 1

Einleitung

Historisch bezieht sich der Begriff der Spektralanalyse auf die Zerlegung von polychromatischem Licht mit Hilfe eines Prismas, so daß sich einzelne unterschiedliche Farbanteile getrennt in dem zugehörigen Spektrum wahrnehmen lassen. Moderne spektralanalytische Verfahren lassen sich allgemein als mathematisch-statistische Methoden definieren, die im wesentlichen dazu dienen, unscharfe periodische Strukturen innerhalb bzw. zwischen zeitlich und/oder räumlich geordneten Beobachtungen deutlich zu machen. Es existiert eine kaum noch überschaubare Vielfalt von Methoden und wissenschaftlichen Anwendungsgebieten. Eine kleine Auswahl von Anwendungsgebieten, wobei Überschneidungen möglich sind, umfaßt unter anderem:

Ökonomie. Seit Anfang der sechziger Jahre werden spektralanalytische Verfahren zur Analyse der Eigenschaften von ökonomischen Zeitreihen eingesetzt (vgl. u.a. GRANGER/HATANAKA[1964], KÖNIG/WOLTERS[1972], FRIEDRICH[1974, III.2.4]). Neuere Anwendungen in der Ökonometrie, bei denen spektralanalytische Methoden eingesetzt werden, umfassen z.B. die Messung von Persistenz (vgl. z.B. WOLTERS[1991, 1992]) oder die Schätzung von Heteroskedastie-Autokorrelations-konsistenten Kovarianzmatrizen (vgl. z.B. ANDREWS[1991], DEN HAAN/LEVIN[1994]).

Chemie. Eines der zur Zeit anspruchvollsten Anwendungsgebiete für *hochauflösende* Verfahren der Spektralschätzung ist die magnetische Kernresonanz-Spektroskopie zur Strukturaufklärung von Molekülen (vgl. z.B. BRETTHORST[1988], SCHMIDT[1989], BRILLINGER/KAISER[1992], DREHER[1992]). Eine enge Verbindung zur Biomedizin bildet die in-vivo Kernresonanz-Spektroskopie, z.B. zur vergleichenden quantitativen Analyse des Stoffwechsels in gesunden Zellen und in Krebszellen (vgl. z.B. SCHMIDT[1989, S. 69ff]).

Ingenieurwissenschaften. Beispiele finden sich u.a. bei der Analyse von Radarsignalen (vgl. z.B. HAYKIN/KESLER[1983]), Sonarsignalen (vgl. z.B. KAY[1988, S. 4ff]) oder in der Sprachsignalverarbeitung (vgl. z.B. ITAKURA/SAITO[1970], SHUMWAY[1988, S. 82ff]).

Astronomie. Die Analyse der Periodizität von Sonnenflecken gehört mit zu den klassischen Anwendungsbeispielen der Spektralanalyse (vgl. z.B. YULE[1927]).

Medizin. In der Neurologie werden z.B. EEG-Signale mit Hilfe von Spektralschätzungen analysiert (vgl. z.B. STEINBERG/GASSER/FRANKE[1985], ALAGÓN[1989]), endokrinologische Anwendungen der Spektralanalyse findet man z.B. in COATES/DIGGLE[1986], PRANK et al. [1990].

Geophysik. Beispiele sind u.a. die spektralanalytische Analyse von künstlich erzeugten oder natürlichen seismologischen Signalen wie Sprengungen zur Lagerstättenexploration, Atombombenexplosionen, bzw. Erdbeben (vgl. z.B. BÅTH[1974], DARGAHINOUBARY/LAYCOCK[1981]).

Klimatologie. Periodische Schwankungen von orbitalen Parametern der Erde (Exzentrizität, Obliquität, Präzession) und ihr Zusammenhang mit langfristigen Klimaschwankungen können mit Hilfe von spektralanalytischen Verfahren untersucht werden (vgl. z.B. MACDONALD[1990]).

Array Processing. Ein übergreifendes Spezialgebiet mit Anwendungen in der Astronomie, Ingenieurwissenschaften oder der Geophysik ist das sog. „Array Processing“, wo Empfangsdaten räumlich angeordneter Sensoren mit Hilfe spektralanalytischer Verfahren ausgewertet werden (vgl. z.B. JOHNSON/DUDGEON[1993]).

Einerseits verleihen die vielfältigen Anwendungsgebiete der Spektralanalyse als statistischer Methode einen besonderen Reiz. Andererseits hat die Tatsache, daß Wissenschaftler aus den unterschiedlichsten Bereichen seit langem zur Entwicklung des spektralanalytischen Instrumentariums beigetragen haben, dazu geführt, daß Arbeiten bezüglich des grundsätzlich gleichen Forschungsgegenstandes in zahlreichen verschiedenen Soziolekten formuliert werden. KAY[1988, S. xiii] charakterisiert treffend diese Entwicklung als:

„ ... a virtual ”tower of Babel” of nomenclature, terminology, and symbolism“.

Hierdurch wird die für schnelle und erfolgreiche Weiterentwicklungen notwendige interdisziplinäre Kommunikation erheblich behindert.

Schwerpunkt der vorliegenden Arbeit ist die Diskussion und Evaluierung von automatischen parametrischen und nichtparametrischen Verfahren der Spektralschätzung:

Die mathematisch-statistischen Methoden zur Spektralanalyse lassen sich grob in parametrische und nichtparametrische Verfahren unterteilen. Beide Klassen von Verfahren unterscheiden sich im wesentlichen durch die Art der getroffenen Modellannahmen. Bei parametrischen Modellen werden stärkere Annahmen über die Natur des zu modellierenden datenerzeugenden Prozesses getroffen werden als bei nichtparametrischen Modellen. Sowohl bei parametrischen Modellen als auch bei nichtparametrischen Modellen müssen vor Durchführung einer Schätzung weitere a priori Entscheidungen zur Modellspezifikation getroffen werden: Bei parametrischen Spektralschätzungen muß eine spezielle parametrische Modellklasse und die Anzahl der zu schätzenden Parameter vorgegeben werden. Bei nichtparametrischen Spektralschätzverfahren muß ein spezielles Glättungsverfahren ausgesucht und das Ausmaß der Glättung durch Vorgabe eines Glättungsparameters festgelegt werden. Insbesondere bei parametrischen Spektralschätzungen kann darüberhinaus für eine gegebene Modellklasse die Wahl eines geeigneten Schätzalgorithmus eine erhebliche Rolle spielen. Ziel von automatischen Verfahren der Spektralschätzung ist es, möglichst alle subjektiven Entscheidungen bei der Modellspezifikationen durch objektive, datengestützte Entscheidungsverfahren zu ersetzen. Der Schwerpunkt bei parametrischen Spektralschätzverfahren liegt hierbei auf der Bestimmung der Zahl der zu schätzenden Parameter, bei nichtparametrischen Spektralschätzverfahren auf der Bestimmung des Glättungsparameters. Grundsätzlich ist es allerdings möglich, weitere Entscheidungen der Modellspezifikation mit Hilfe von objektiven, datengestützten Entscheidungsverfahren durchzuführen. Die Ziele von automatischen Spektralschätzverfahren sind zum einen darin zu sehen, die Standardauswertung von Massendaten zu erleichtern. Zum anderen sollen automatische Spektralschätzverfahren es ermöglichen, weniger erfahrenen Anwendern bei der Modellspezifikation Hilfestellung zu geben, so daß methodisch anspruchsvolle Verfahren so weit wie möglich ohne Expertenunterstützung eingesetzt werden können.

Für parametrische Spektralschätzungen sind seit den grundlegenden Arbeiten von AKAIKE[1969a,b] zahlreiche Modellselektionskriterien entwickelt worden. WHABA/WOLD[1975b] entwickelten auf der Grundlage von Glättungsplines unter Verwendung von Kreuzvalidierung ein automatisches nichtparametrisches Spektralschätzverfahren.

Einen ersten allgemeinen Ansatz, sowohl parametrische als auch nichtparametrische automatische Spektralschätzungen mit Hilfe von Kreuzvalidierung in einem einheitlichen Rahmen zu behandeln macht HURVICH[1985]. Verschiedene Weiterentwicklungen folgten, ohne daß die fast ausschließlich auf Kreuzvalidierungstechniken beruhenden nichtparametrischen Modellselektionskriterien sich auch nur annähernd so durchsetzen konnten, wie parametrische Modellselektionskriterien.

Angesichts des weiten Einsatzgebietes von Spektralschätzverfahren kann von vornherein nicht erwartet werden, ein automatisches Spektralschätzverfahren zu entwickeln, das sich bei allen Anwendungen gleichermaßen gut bewährt. Vielmehr kann nur, aufbauend auf einem gründlichen Verständnis der mathematisch-statistischen Grundlagen sowie der Vielfalt von parametrischen und nichtparametrischen Schätzverfahren, ein allgemeiner methodischer Rahmen zur Entwicklung und Evaluierung von verschiedenen automatischen Schätzverfahren vorgegeben werden. Dieser Zielsetzung entsprechend, kann kurz das weitere Vorgehen in dieser Arbeit skizziert werden:

In Kapitel 2 werden die Grundlagen von stochastischen Prozessen vorgestellt sowie die grundlegende Terminologie eingeführt. Für die Entwicklung von automatischen Spektralschätzverfahren ist insbesondere der Abschnitt 2.6 und die Spektraldarstellung der Kullback-Leibler-Information von stochastischen Prozessen im Frequenzbereich von Bedeutung.

In Kapitel 3 wird die wichtigste Modellklasse zur parametrischen Spektralschätzung, das autoregressive Modell, behandelt. Das Konzept der maximalen Entropie wird kritisch diskutiert. Einen besonderen Stellenwert besitzt die Darstellung der unterschiedlichen Algorithmen zur Schätzung von autoregressiven Modellen sowie deren Relevanz bei unterschiedlichen Problemstellungen.

Direkte und indirekte Spektralschätzer, die beiden klassischen Verfahren der nichtparametrischen Spektralschätzung, werden in Kapitel 4 diskutiert.

In Kapitel 5 werden die methodischen Grundlagen entwickelt, die einen sinnvollen Vergleich von verschiedenen automatischen parametrischen und nichtparametrischen Spektralschätzverfahren ermöglichen. Unter der Bezeichnung „global korrigierte Diskrepanzen“ wird eine neue Klasse von Diskrepanzmaßen für Spektraldichtefunktionen eingeführt.

In Kapitel 6 werden Modellselektionskriterien sowohl für parametrische als auch für nichtparametrische Methoden der automatischen Spektralschätzung diskutiert. Aufbauend auf den Arbeiten von HANNAN/RISSANEN[1988], HURVICH/BELTRÃO[1990]

und HURVICH[1992] wird eine neue Klasse von „strafenden nichtparametrischen“ Modellselektionskriterien vorgeschlagen. Der größte Vorzug dieser Modellselektionskriterien besteht in ihrer einfachen Konstruktion und wenig aufwendigen Berechenbarkeit, die in völliger Analogie zu den gewohnten parametrischen Modellselektionskriterien erfolgen kann. Für ausgewählte parametrische und nichtparametrische Modellselektionskriterien erfolgt im Rahmen einer Simulationsstudie ein umfassender Vergleich. Eine abschließende Evaluierung von verschiedenen Modellselektionskriterien erfolgt anhand von jeweils einer Beispielzeitreihe aus Ökonomie, Biomedizin und Astronomie.

Ergänzende Bemerkungen:

In weiten Bereichen der Literatur zur Spektralschätzung ist fast durchgängig eine Darstellung mit Bezug auf komplexwertige stochastische Prozesse üblich. Ebenso werden, sofern möglich, alle Schätzalgorithmen sowie Programme für komplexwertige Beobachtungen formuliert bzw. komplexwertige Testdatensätze zur Programmvalidierung zur Verfügung gestellt. Dies wird nicht nur deshalb gemacht, weil Anwendungen existieren, bei denen es sinnvoll ist, zwei reellwertige Beobachtungsreihen als eine komplexwertige Reihe zu interpretieren, wie z.B. bei der Kernresonanz-Spektroskopie. Vielmehr basieren weitergehende Analysetechniken auf Datentransformationen, wie z.B. der komplexen De- und Remodulation, deren einzelne Schritte nur im Rahmen von komplexwertigen Prozessen sinnvoll interpretiert werden können (vgl. z.B. SCHLITTEGEN/STREITBERG[1984, S. 136ff], MARPLE[1987, xix]). Da es erheblich einfacher ist, von der komplexen Darstellung auf die reelle Darstellung überzugehen als umgekehrt, wird in dieser Arbeit meist von der komplexwertigen Darstellung Gebrauch gemacht. Ausnahmen sind insbesondere die Bereiche, wo für den komplexwertigen Fall keine mathematisch-statistischen Aussagen zur Verfügung stehen. Die Entscheidung für eine weitgehend komplexwertige Darstellung wurde dadurch erheblich erleichtert, daß die matrizenorientierte Programmiersprache GAUSSI 3.0 den komplexen Datentyp soweit wie möglich bei allen Datenmanipulationen unterstützt. Daher war es auch ohne erheblichen Programmiermehraufwand möglich, alle Schätzverfahren in Programmen zu implementieren, die komplexwertige Beobachtungsreihen als Input akzeptieren.

In das Literaturverzeichnis wurden mit wenigen Ausnahmen Arbeiten aufgenommen, die unmittelbar im vorliegenden Text zitiert werden. Ein ausführlicheres Literaturverzeichnis, das auch Gebiete der Spektralanalyse umfaßt, die im Rahmen dieser Arbeit nicht behandelt werden können, steht zur Verfügung.

Kapitel 2

Grundlagen stochastischer Prozesse

Die ersten Abschnitte dieses Kapitels dienen der allgemeinen Einführung von Grundbegriffen sowie der Notation. Allgemeine Literatur hierzu ist z.B. LAMPERTI[1977], PRIESTLEY[1981], LAMBERT/POSKITT[1983], BROCKWELL/DAVIS[1991], SCHLITTEGEN/STREITBERG[1984], NEWTON[1988]. Der letzte Abschnitt dieses Kapitels über die Zusammenhänge von Entropie, Information und stochastischen Prozessen bildet eine wichtige Grundlage zum Verständnis von autoregressiven Spektralschätzern sowie der Repräsentation von Informationskriterien im Frequenzbereich.

2.1 Stochastische Prozesse

Def. 2.1.1 (*Stochastischer Prozeß*)

Sei \mathcal{T} eine Teilmenge der reellen Zahlen \mathcal{R} . Dann heißt eine Familie von Zufallsvariablen auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathcal{P})$

$$(2.1) \quad X_t(\omega) : \Omega \mapsto \mathcal{R} \quad t \in \mathcal{T}, \omega \in \Omega$$

reellwertiger stochastischer Prozeß mit der Indexmenge \mathcal{T} .

■

Die Indexmenge \mathcal{T} wird hier als Zeit interpretiert. Ist die Indexmenge überabzählbar, so heißt der stochastische Prozeß kontinuierlich, ist sie abzählbar, wenn z. B. die Indexmenge \mathcal{T} gleich den ganzen Zahlen \mathcal{Z} ist, so heißt der stochastische Prozeß diskret. In dieser Arbeit werden nur stochastische Prozesse betrachtet, deren Indexmenge eine Teilmenge der ganzen Zahlen ist.

In der Betrachtungsweise von Definition 2.1.1 ist $X_t(\omega)$ für jedes fest vorgegebene t eine Zufallsvariable. Hält man dagegen ein $\omega \in \Omega$ fest und läßt den Zeitindex t variieren, so erhält man eine reellwertige Abbildung der Variablen t :

Def. 2.1.2 (*Realisationen stochastischer Prozesse*)

Die Abbildungen

$$(2.2) \quad X_t(\omega) : \mathcal{T} \mapsto \mathcal{R} \quad \omega \in \Omega, \omega \text{ fest}$$

heißen *Realisationen des stochastischen Prozesses* $\{X_t(\omega), t \in \mathcal{T}\}$.

■

Realisationen stochastischer Prozesse werden auch Stichprobenpfade oder Trajektorien genannt.

Die endlich-dimensionalen Verteilungen eines stochastischen Prozesses lassen sich angeben als:

Def. 2.1.3 (*Verteilungen eines stochastischen Prozesses*)

Die Familie $\{F_{t_1, \dots, t_n}\}$ der endlich-dimensionalen Verteilungsfunktionen eines stochastischen Prozesses $\{X_t(\omega), t \in \mathcal{T}\}$ sei definiert als:

$$(2.3) \quad F_{t_1, \dots, t_n}(x_1, \dots, x_n) = P(X_{t_1} < x_1, \dots, X_{t_n} < x_n)$$

$$\forall (t_1, \dots, t_n) \quad \text{mit} \quad t_1 < t_2 < \dots < t_n \quad \text{und} \quad n \in \mathcal{N}$$

■

Das Theorem von Kolmogorov sagt aus, daß sich stochastische Prozesse durch ihre endlich-dimensionalen Verteilungen eindeutig bestimmen lassen, so daß eine Beschreibung von stochastischen Prozessen durch ein System endlich-dimensionaler Verteilungen erfolgen kann. Eine allgemeine Darstellung des Theorems von Kolmogorov mit Beweis findet sich z.B. in LAMPERTI [1977, Anhang 1]. Die vereinfachte Darstellung hier, mit den geordneten Indizes (t_1, \dots, t_n) , erlaubt den Verzicht auf eine zusätzliche Permutationsbedingung (BROCKWELL/DAVIS [1991, S. 10/11]).

Satz 2.1.1 (*Theorem von Kolmogorov*)

Die Verteilungsfunktionen $\{F_{t_1, \dots, t_n}\}$ sind dann und nur dann Verteilungsfunktionen eines stochastischen Prozesses für $n \in \mathcal{N}$ und $1 \leq i \leq n$, wenn gilt:

$$(2.4) \quad \begin{aligned} & F_{t_1, \dots, t_{i-1}, t_{i+1}, \dots, t_n}(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n) = \\ & F_{t_1, \dots, t_{i-1}, t_i, t_{i+1}, \dots, t_n}(x_1, \dots, x_{i-1}, \infty, x_{i+1}, \dots, x_n) \\ & \forall (t_1, \dots, t_n) \quad \text{mit} \quad t_1 < t_2 < \dots < t_n \quad \text{und} \quad n \in \mathcal{N} \end{aligned}$$

Gleichung (2.4) fordert, daß sich die Verteilungen niedrigerer Dimension eines stochastischen Prozesses als Randverteilungen seiner höherdimensionalen Verteilungen ableiten lassen. ■

Def. 2.1.4 (*Strenge Stationarität*)

Ein stochastischer Prozeß heißt *streng stationär*, wenn gilt:

$$(2.5) \quad \begin{aligned} & F_{t_1, \dots, t_n}(x_1, \dots, x_n) = F_{t_1+h, \dots, t_n+h}(x_1, \dots, x_n) \\ & \text{mit} \quad t_i, t_i + h \in \mathcal{T}, \quad i = 1, \dots, n, n \in \mathcal{N}. \end{aligned}$$

Die gemeinsame Verteilung der Zufallsvariablen X_{t_1}, \dots, X_{t_n} eines streng stationären Prozesses ist also invariant gegenüber Verschiebungen in der Zeit. Sie hängt nur von der Entfernung zwischen den einzelnen Elementen ab und nicht von deren absoluter Lage auf der Zeitachse. Die Kenntnis der gesamten gemeinsamen Verteilung ist im allgemeinen eine zu hohe Anforderung. Für praktische Belange beschränkt man sich auf Aussagen über die ersten und zweiten Momente eines stochastischen Prozesses. Die Beschreibung der zeitlichen Abhängigkeiten erfolgt über die Autokovarianzen. ■

Def. 2.1.5 (*Autokovarianzfunktion*)

Sei $\{X_t, t \in \mathcal{T}\}$ ein stochastischer Prozeß mit $\text{Var}(X_t) < \infty \quad \forall t \in \mathcal{T}$. Dann ist die Autokovarianzfunktion $\gamma_X(r, s)$ mit $r, s \in \mathcal{T}$ definiert als:

$$(2.6) \quad \gamma_X(r, s) = \text{Cov}(X_r, X_s) = E[(X_r - \mu_{X_r})(X_s - \mu_{X_s})],$$

wobei $\mu_{X_r} = E(X_r)$. ■

Def. 2.1.6 (*Schwache Stationarität*)

Ein stochastischer Prozeß $\{X_t, t \in \mathcal{Z}\}$ heißt *schwach stationär*, wenn gilt:

- (i) $E|X_t|^2 < \infty \quad \forall t \in \mathcal{Z}$
- (ii) $E(X_t) = \mu \quad \forall t \in \mathcal{Z}$
(mittelwertstationär)
- (iii) $\gamma_X(r, s) = \gamma_X(s - r) \quad \forall r, s \in \mathcal{Z}$
(kovarianzstationär)

■

Als abkürzende Schreibweise für die Autokovarianzfunktion eines schwach stationären stochastischen Prozesses wird statt $\gamma_X(r - s)$ auch $\gamma_X(\tau)$ oder einfach γ_τ verwandt, wobei $\tau = r - s$ ist.

Durch Normierung mit $\gamma(0)$ erhält man aus der Autokovarianzfunktion γ_τ die Autokorrelationsfunktion $\rho(\tau)$, d.h.

$$\rho(\tau) = \frac{\gamma(\tau)}{\gamma(0)}.$$

Mittelwert- und Kovarianzstationarität wird zusammenfassend auch als schwache Stationarität bezeichnet. Aus der strengen Stationarität folgt die schwache Stationarität. Für Gauss-Prozesse, d.h. für Prozesse, deren Verteilungsfunktionen multivariate Normalverteilungen sind, folgt aus der schwachen Stationarität auch die strenge Stationarität.

Im weiteren soll unter Stationarität immer schwache Stationarität verstanden werden. Die Autokovarianzfunktion eines stationären stochastischen Prozesses kann wie folgt beschrieben werden (BROCKWELL/DAVIS [1991, S. 26]):

Satz 2.1.2 (*Eigenschaften der Autokovarianzfunktion*)

Sei $\gamma_X(\cdot)$ die Autokovarianzfunktion eines stationären stochastischen Prozesses. Dann gilt:

- (i) $\gamma_X(0) \geq 0$
- (ii) $|\gamma_X(\tau)| \leq \gamma_X(0)$
- (iii) $\gamma_X(\tau) = \gamma_X(-\tau), \quad \forall \tau \in \mathcal{Z}.$

■

Autokovarianzfunktionen besitzen eine weitere Eigenschaft, die besonders hervorgehoben werden soll. Zunächst ist noch eine weitere Definition erforderlich:

Def. 2.1.7

Eine Funktion $h : \mathcal{Z} \mapsto \mathcal{R}$ heißt *positiv-semidefinit*, wenn gilt:

$$\sum_{i=1}^n \sum_{j=1}^n a_i a_j h(t_i - t_j) \geq 0$$

$$\forall (a_1, \dots, a_n) \in \mathcal{R}^n, \quad \forall (t_1, \dots, t_n) \in \mathcal{Z}^n, \quad n \in \mathcal{N}$$

■

Der folgende Satz trifft nun eine sehr allgemeine Aussage über positiv-semidefinite Funktionen und Autokovarianzfunktionen, die für die Spektraldarstellung von stationären Prozessen von grundlegender Bedeutung ist (siehe z.B. BROCKWELL/DAVIS-[1991, S. 27]):

Satz 2.1.3 (*Positiv-Semidefinitheit/Autokovarianzfunktion*)

Eine reellwertige Funktion $\gamma(\tau)$, $\tau \in \mathcal{Z}$ mit $\gamma(\tau) = \gamma(-\tau)$ ist dann und nur dann positiv-semidefinit, wenn sie die Autokovarianzfunktion eines stationären Prozesses ist.

■

In praktischen Anwendungen treten häufig Zeitreihen auf, die nicht stationär sind. Mit Hilfe eines Beispiels sei kurz darauf hingewiesen, daß ein stochastischer Prozeß kovarianzstationär sein kann, ohne mittelwertstationär zu sein:

Sei $\{Z_t\}$ ein stationärer stochastischer Prozeß mit $E(Z_t) = \mu$ für alle $t \in \mathcal{T}$. Bildet man den neuen Prozeß $\{X_t\}$ als:

$$X_t = m_t + Z_t,$$

wobei m_t eine deterministische Funktion der Zeit t ist, so ergibt sich:

$$\begin{aligned} (i) \quad E(X_t) &= m_t + \mu \\ (ii) \quad \gamma_X(\tau) &= E[(X_{t+\tau} - m_{t+\tau} - \mu)(X_t - m_t - \mu)] \\ &= E[(Z_{t+\tau} - \mu)(Z_t - \mu)] \\ &= \gamma_Z(\tau). \end{aligned}$$

Der Prozeß $\{X_t\}$ ist also nicht mittelwertstationär, aber kovarianzstationär. Dies impliziert, daß auch Zeitreihen, die nicht mittelwertstationär sind, nach geeigneter Trendbereinigung mit dem für stationäre Prozesse entwickeltem Instrumentarium behandelt werden können.

2.2 Spektraldarstellung stochastischer Prozesse

Bei der theoretischen Betrachtung der Spektraldarstellung stationärer stochastischer Prozesse ist es häufig einfacher, zunächst den allgemeineren Fall komplexwertiger Prozesse zu betrachten.

Sind U_t und V_t reellwertige stochastische Prozesse über dem gleichen Wahrscheinlichkeitsraum, dann läßt sich ein komplexwertiger stochastischer Prozeß X_t darstellen als:

$$X_t = U_t + iV_t \quad \text{mit} \quad i^2 = -1.$$

Für einen komplexwertigen stochastischen Prozeß wird die Definition (2.1.5) der Autokovarianz modifiziert zu:

Def. 2.2.1

Sei $\{X_t, t \in \mathcal{Z}\}$ ein komplexwertiger stochastischer Prozeß. Dann ist die Autokovarianzfunktion $\gamma_X(r, s)$ mit $r, s \in \mathcal{Z}$, definiert als:

$$(2.7) \quad \gamma_X(r, s) = \text{Cov}(X_r, X_s) = E[(X_r - \mu_{X_r})(X_s - \mu_{X_s})^*]$$

■

Für komplexwertige stationäre stochastische Prozesse muß die Eigenschaft (iii) der Autokovarianzfunktion aus Satz (2.1.2) verallgemeinert werden zu:

$$(2.8) \quad \text{(iii)} \quad \gamma(\tau) = \gamma^*(-\tau),$$

d.h. die Autokovarianzfunktion eines komplexwertigen stochastischen Prozesses ist hermitisch.

Analog kann Satz 2.1.3 mit einer geeigneten Definition für die Positiv-Semidefinitheit einer komplexwertigen Funktion reformuliert werden als:

Satz 2.2.1

Eine hermitesche Funktion $\gamma : \mathcal{Z} \mapsto \mathcal{C}$ ist dann und nur dann positiv-semidefinit, d.h. es gilt:

$$\sum_{i=1}^n \sum_{j=1}^n a_i a_j^* \gamma(t_i - t_j) \geq 0$$

$$\forall (a_1, \dots, a_n) \in \mathcal{C}^n, \quad \forall (t_1, \dots, t_n) \in \mathcal{Z}^n, \quad n \in \mathcal{N},$$

wenn sie die Autokovarianzfunktion eines stationären stochastischen Prozesses ist. ■

Die Grundidee der Spektraldarstellung von stochastischen Prozessen läßt sich an Hand eines einfachen Beispiels veranschaulichen:

Sei

$$(2.9) \quad X_t = \sum_{j=1}^n A_j e^{i2\pi t \omega_j}$$

eine Summe komplexwertiger Cosinoid-Prozesse mit $-0.5 < \omega_1 < \omega_2 < \dots < \omega_n \leq 0.5$ und den unkorrelierten komplexen Zufallskoeffizienten A_j , für die gilt:

$$(2.10) \quad E(A_j) = 0$$

$$(2.11) \quad E(A_j A_j^*) = \sigma_j^2.$$

Damit folgt:

$$(2.12) \quad E(X_t) = 0$$

$$(2.13) \quad E(X_{t+\tau} X_t^*) = \sum_{j=1}^n \sigma_j^2 e^{i2\pi \tau \omega_j}$$

$$(2.14) \quad = \gamma(\tau),$$

d.h. $\{X_t\}$ ist ein schwach stationärer Prozeß. Nimmt man als (nicht-normierte) Verteilungsfunktion

$$(2.15) \quad F(\omega) = \sum_{j: \omega_j \leq \omega} \sigma_j^2,$$

so läßt sich $\gamma(\tau)$ mit Hilfe eines Stieltjes-Integrales schreiben als:

$$(2.16) \quad \gamma(\tau) = \int_{-0.5}^{0.5} e^{i2\pi\tau\omega} dF(\omega) .$$

Die Darstellungen (2.9) und (2.16) lassen sich nun für beliebige schwach stationäre Prozesse mit Erwartungswert Null verallgemeinern:

Für die allgemeine Spektraldarstellung eines stochastischen Prozesses muß die Summe mit unkorrelierten, komplexen Zufallskoeffizienten durch ein stochastisches Integral bezüglich eines stochastischen Prozesses mit orthogonalen Zuwächsen ersetzt werden, so daß er sich darstellen läßt als (BROCKWELL/DAVIS[1991, S. 117, 133-144]):

$$(2.17) \quad X_t = \int_{-0.5}^{0.5} e^{i2\pi t\omega} dZ(\omega) .$$

Da dies ein zu aufwendiges mathematisches Instrumentarium erfordert, wird im folgenden für die Herleitung der Spektraldarstellung die Verallgemeinerung von (2.16) betrachtet, wobei man sich auf Stieltjes-Integrale beschränken kann.

Grundlage der Spektraldarstellung der Autokovarianzfunktion eines stationären stochastischen Prozesses mit ganzzahliger Indexmenge ist dann folgender Satz (z.B. BROCKWELL/DAVIS [1991, S. 116ff], LAMBERT/POSKITT [1983, S. 58/59]):

Satz 2.2.2 (*Theorem von Herglotz*)

Eine Funktion $\gamma : \mathcal{Z} \mapsto \mathcal{C}$ ist dann und nur dann positiv-semidefinit, wenn eine beschränkte, nicht-fallende Funktion $F : [-0.5, 0.5] \mapsto \mathcal{R}$ existiert, so daß gilt:

$$(2.18) \quad \gamma(\tau) = \int_{-0.5}^{0.5} e^{i2\pi\omega\tau} dF(\omega) \quad \forall \tau \in \mathcal{Z} .$$

■

Aufbauend auf Satz (2.2.2) läßt sich nun definieren:

Def. 2.2.2 (*Spektralverteilung/-dichte*)

Sei $\{X_t\}$ ein stationärer stochastischer Prozeß mit der Autokovarianzfunktion $\gamma(\tau)$, für die gilt:

$$\gamma(\tau) = \int_{-0.5}^{0.5} e^{i2\pi\omega\tau} dF(\omega) .$$

Dann heißt $F(\omega)$ die (nicht-normierte) Spektralverteilungsfunktion von $\{X_t\}$ und $\gamma(\tau)$. Ist $F(\omega)$ differenzierbar, so heißt die Funktion $f(\omega) = F'(\omega)$ die (nicht-normierte) Spektraldichtefunktion von $\{X_t\}$ und $\gamma(\tau)$. ■

Die normierte Spektralverteilungsfunktion $H(\omega)$ ergibt sich als:

$$H(\omega) = \frac{1}{\gamma(0)} F(\omega) ,$$

so daß sich die Autokorrelationsfunktion $\rho(\tau)$ eines stationären stochastischen Prozesses darstellen läßt als:

$$\rho(\tau) = \int_{-0.5}^{0.5} e^{i2\pi\omega\tau} dH(\omega) .$$

Wie in der Literatur üblich, soll bei der Verwendung der Begriffe Spektralverteilungsbzw. Spektraldichtefunktion nur dann explizit zwischen den normierten und den unnormierten Funktionen unterschieden werden, wenn dies für die aktuelle Fragestellung tatsächlich relevant ist.

Besitzt ein stationärer Prozeß eine absolut-summierbare Autokovarianzfunktion, d.h. wenn gilt:

$$\sum_{\tau=-\infty}^{\infty} |\gamma(\tau)| < \infty ,$$

so läßt sich der Zusammenhang zwischen Autokovarianzfunktion und Spektraldichtefunktion einfacher formulieren (z.B. BROCKWELL/DAVIS[1991, S. 118]):

Satz 2.2.3

Eine absolut-summierbare Funktion $\gamma : \mathcal{Z} \mapsto \mathcal{C}$ ist dann und nur dann Autokovarianzfunktion eines stationären stochastischen Prozesses $\{X_t\}$, wenn sich dessen Spektraldichtefunktion darstellen läßt als:

$$(2.19) \quad f(\omega) = \sum_{\tau=-\infty}^{\infty} e^{-i2\pi\omega\tau} \gamma(\tau) \quad \forall \omega \in [-0.5, 0.5].$$

Da sich bei Existenz von $f(\omega)$ die Autokovarianzfunktion darstellen läßt als:

$$(2.20) \quad \gamma(\tau) = \int_{-0.5}^{0.5} e^{i2\pi\omega\tau} f(\omega) d\omega ,$$

ergibt sich, daß Spektraldichtefunktion und Autokovarianzfunktion ein Paar von Fouriertransformierten bilden. ■

Der Zerlegungssatz von Lebesgue für Verteilungsfunktionen der Wahrscheinlichkeitstheorie läßt sich auch auf die normierten Spektralverteilungsfunktionen übertragen. Für alle praktisch relevanten Fälle läßt sich der Zerlegungssatz vereinfacht formulieren als (PRIESTLEY[1981, Vol. 1, S. 226ff]):

Satz 2.2.4 (Zerlegungssatz)

Jede normierte Spektralverteilungsfunktion läßt sich darstellen als:

$$(2.21) \quad H(\omega) = a_1 H_1(\omega) + a_2 H_2(\omega)$$

mit:

1. $a_1, a_2 \geq 0, \quad a_1 + a_2 = 1$

2. Für die Funktionen $H_1(\omega)$ und $H_2(\omega)$ gilt:

- (a) $H_1(\omega)$ ist eine absolut stetige Funktion mit $H_1' = h_1(\omega)$, so daß gilt $H_1(\omega) = \int_{-0.5}^{\omega} h_1(u) du$.

- (b) $H_2(\omega)$ ist eine Treppenfunktion mit Sprüngen der Höhe $\{h_r\}$ an den Stellen $\{\omega_r\}$, $r = 1, 2, 3, \dots$ und $\sum_r h_r = 1$.

■

Mit (2.21) läßt sich die Autokorrelationsfunktion $\rho(\tau)$ in Abhängigkeit von einer stetigen und diskreten Komponente darstellen als:

$$(2.22) \quad \rho(\tau) = a_1 \rho^{(1)}(\tau) + a_2 \rho^{(2)}(\tau)$$

mit:

$$\rho^{(1)}(\tau) = \int_{-0.5}^{0.5} e^{i2\pi\omega\tau} h_1(\omega) d\omega$$

und

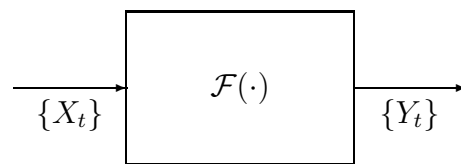
$$(2.23) \quad \rho^{(2)}(\tau) = \sum_{r=1}^{\infty} e^{i2\pi\omega_r\tau} h_r.$$

Ebenso wie Autokovarianzfunktion und nicht-normierte Spektraldichtefunktion sind $\rho^{(1)}(\tau)$ und $h_1(\omega)$ ein Paar von Fouriertransformierten und es gilt $\rho^{(1)}(\tau) \rightarrow 0$ für $\tau \rightarrow \pm\infty$ (PRIESTLEY[1981, S. 228]). Im Gegensatz dazu ergibt sich $\rho^{(2)}(\tau)$ als Summe von periodischen Cosinus- und Sinustermen, so daß $\rho^{(2)}(\tau)$ unabhängig davon, wie groß τ wird, nicht gegen Null konvergiert.

Die Frage, aus welchen Komponenten sich eine Spektraldichtefunktion zusammensetzt, ist insbesondere für die Schätzung von Spektraldichtefunktionen relevant, da dies für die Auswahl eines geeigneten Schätzverfahrens von Bedeutung ist. In vielen Fällen kann man sich auf die Betrachtung von Spektraldichtefunktionen mit absolut stetiger Komponente beschränken. Eine Entscheidung darüber kann aber nur in Hinblick auf ein spezielles Anwendungsgebiet getroffen werden.

2.3 Lineare Filter und stochastische Prozesse

Für die Analyse stochastischer Prozesse ist es nützlich, ein einfaches Modell zu betrachten, bei dem ein Inputprozeß $\{X_t, t \in \mathcal{Z}\}$ durch einen Filter $\mathcal{F}(\cdot)$ in einen Outputprozeß $\{Y_t, t \in \mathcal{Z}\}$ transformiert wird:



Eine wichtige Kenngröße eines Filters ist seine Impulsantwortfolge:

Def. 2.3.1 (*Impulsantwort*)

Verwendet man als Input eines Filters die Einheitsimpulsfolge $\{\delta_t\}$ mit:

$$\delta_t = \begin{cases} 1 & t = 0 \\ 0 & t \neq 0, \end{cases}$$

so erhält man als Output die Folge $\{\psi_j\}$, die als Impulsantwortfolge bzw. Impulsantwort bezeichnet wird.

■

Für die Beschreibung von Filtern sind die folgenden Eigenschaften relevant: Zeitinvarianz, Linearität, Stabilität, Kausalität, Invertierbarkeit.

Für zeitinvariante Filter gilt: Wird dem Inputprozeß $\{X_t\}$ der Outputprozeß $\{Y_t\}$ zugeordnet, so wird für alle möglichen Verzögerungen $\tau \in \mathcal{Z}$ der verzögerte Inputprozeß $\{X_{t-\tau}\}$ in den verzögerten Outputprozeß $\{Y_{t-\tau}\}$ transformiert.

Lineare Filter sind durch das Superpositionsprinzip definiert:

$$\mathcal{F}(\alpha\{X_t^{(1)}\} + \beta\{X_t^{(2)}\}) = \alpha\mathcal{F}(\{X_t^{(1)}\}) + \beta\mathcal{F}(\{X_t^{(2)}\}),$$

d.h. eine Transformation einer Linearkombination zweier Inputprozesse ist gleich der Linearkombination der transformierten Outputprozesse.

Zeitinvariante lineare Filter für stochastische Prozesse lassen sich durch die folgende Transformation definieren:

Def. 2.3.2 (*Zeitinvariante lineare Filter*)

Die Transformation eines stochastischen Prozesses $\{X_t, t \in \mathcal{Z}\}$ in einen stochastischen Prozeß $\{Y_t, t \in \mathcal{Z}\}$ mit der Impulsantwortfolge $\{\psi_j, j \in \mathcal{M} \subseteq \mathcal{Z}\}$ gemäß

$$(2.24) \quad Y_t = \sum_{j \in \mathcal{M}} \psi_j X_{t-j}$$

wird als zeitinvarianter linearer Filter bezeichnet. ■

Der Outputprozeß $\{Y_t\}$ ergibt sich also als Faltung der Impulsantwortfolge $\{\psi_j\}$ mit dem stochastischen Prozeß $\{X_t\}$.

Mit den Annahmen

$$\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$$

und

$$E(|X_t|^2) < \infty \quad \forall t \in \mathcal{Z}$$

kann das Gleichheitszeichen in (2.24) als stochastische Konvergenz im quadratischen Mittel interpretiert werden, d.h. es gilt (vgl. z.B. SCHLITZGEN/STREITBERG [1984, S. 86]):

$$\lim_{n \rightarrow \infty} E[|Y_t - \sum_{j=-n}^n \psi_j X_{t-j}|^2] = 0.$$

Ein zeitinvarianter linearer Filter ist durch seine Impulsantwortfolge $\{\psi_j\}$ vollständig beschrieben.

Ein Filter heißt stabil, wenn zu einer beschränkten Inputfolge eine beschränkte Outputfolge erzeugt wird (BIBO-Stabilität). Hinreichende und notwendige Bedingung für die Stabilität eines linearen zeitinvarianten Filters ist die absolute Summierbarkeit der Impulsantwortfunktion (vgl. z.B. OPPENHEIM/SCHAFFER[1992, S. 31]).

Für die Diskussion von weiteren Filtereigenschaften ist es nützlich, sich noch anderer Darstellungsformen von Filtern zu bedienen:

Unter Verwendung des Lagoperators L , der definiert ist durch:

$$L^j X_t = X_{t-j}, \quad j \in \mathcal{Z},$$

kann ein linearer Filter dargestellt werden als:

$$Y_t = \sum_j a_j L^j X_t,$$

bzw. in kompakterer Form mit Hilfe des Lagpolynoms $a(L) = \sum_j a_j L^j$:

$$Y_t = a(L)X_t.$$

Das charakteristische Polynom eines Filters der Form

$$a(L) = 1 + a_1 L + \dots + a_p L^p$$

ist als z -Transformation seiner Filterkoeffizienten definiert:

$$a(z) = 1 + a_1 z^{-1} + \dots + a_p z^{-p} \quad z \in \mathcal{C}.$$

Es ist nützlich, zeitinvariante lineare Filter mit Hilfe der folgenden Eigenschaften zu klassifizieren:

- Kausalität bzw. Antikausalität,
- Invertierbarkeit bzw. Nicht-Invertierbarkeit.

Def. 2.3.3 (*Kausale und antikausale Filter*)

Ein Filter $a(L)$, der sich darstellen läßt als:

$$a(L) = \sum_{j=0}^{\infty} a_j L^j,$$

d.h. für dessen Koeffizienten

$$a_j = 0 \quad \forall j < 0$$

gilt, heißt *kausaler Filter*. Für einen *antikausalen Filter*

$$a(L) = \sum_{j=-\infty}^0 a_j L^j$$

gilt entsprechend:

$$a_j = 0 \quad \forall j > 0.$$

Ansonsten heißt ein Filter *akausal*, d.h.

$$a(L) = \sum_{j=-\infty}^{\infty} a_j L^j.$$

■

Bei kausalen Filtern wird der Output Y_t ausschließlich durch X_s mit $s \leq t$ beeinflusst, so daß zukünftige X_s auf gegenwärtige Y_t keine Wirkung haben.

Da bestimmte Eigenschaften eines zeitinvarianten linearen Systems äquivalent durch kausale oder antikausale Filter dargestellt werden können, kann die Dualität zwischen kausaler und antikausaler Systembeschreibung zusätzlich wertvolle Einsichten vermitteln, was z.B. bei der Entwicklung moderner parametrischer Spektraldichteschätzer von fundamentaler Bedeutung ist¹.

Zur Fragestellung der Existenz eines inversen Filters gelangt man, wenn ein Outputprozeß $\{Y_t\}$ vorliegt, der mit einem bekannten Filter $a(L)$ aus einem unbekanntem Inputprozeß $\{X_t\}$ berechnet wurde: Unter welchen Bedingungen gibt es einen zu $a(L)$ inversen Filter $a^{-1}(L)$, so daß der Inputprozeß $\{X_t\}$ rekonstruiert werden kann?

Für kausale inverse Filter lassen sich die erforderlichen Bedingungen mit Hilfe des charakteristischen Polynoms festlegen:

¹vgl. z.B. Kapitel 3

Satz 2.3.1 (*Existenz kausaler inverser Filter*)

Sei $a(L) = 1 + a_1L + \dots + a_pL^p$ ein linearer, zeitinvarianter Filter. Liegen alle Nullstellen des charakteristischen Polynoms

$$a(z) = 1 + a_1z^{-1} + \dots + a_pz^{-p}$$

innerhalb des Einheitskreises, dann existiert ein kausaler inverser Filter $a^{-1}(L)$ mit absolut summierbarer Koeffizientenfolge. ■

Zu einem Filter $a(L)$ existiert ein akausaler inverser Filter $a^{-1}(L)$ der Form:

$$a^{-1}(L) = \sum_{j=-\infty}^{\infty} \psi_j L^j$$

dann, wenn keine Wurzel seines charakteristischen Polynoms $a(z)$ auf dem Einheitskreis liegt. Wenn im weiteren von einem invertierbaren Filter gesprochen wird, dann ist, wenn nicht ausdrücklich anders angegeben, ein kausal invertierbarer Filter gemeint.

Elementarer Grundbaustein für stochastische Prozesse, die mit Hilfe von zeitinvarianten linearen Filtern dargestellt werden können, ist der White-Noise-Prozeß:

Def. 2.3.4 (*White-Noise-Prozeß*)

Eine Folge $\{\epsilon_t, t \in \mathcal{Z}\}$ von identisch verteilten Zufallsvariablen für die gilt:

$$E(\epsilon_t) = 0 \quad \forall t \in \mathcal{Z}$$

$$\gamma(t, s) = \begin{cases} \sigma_\epsilon^2 & t = s \\ 0 & t \neq s \end{cases}$$

heißt White-Noise-Prozeß mit Erwartungswert 0 und Varianz σ^2 , in Kurznotation:

$$\{\epsilon_t\} \sim \text{WN}(0, \sigma^2).$$

■

Sind die Zufallsvariablen $\{\epsilon_t, t \in \mathcal{Z}\}$ zudem auch noch unabhängig, so wird dies in der üblichen Notation angegeben als:

$$\{\epsilon_t\} \sim \text{IID}(0, \sigma^2).$$

Mit Hilfe eines White-Noise-Prozesses und eines zweiseitig unendlichen linearen Filters läßt sich der allgemeine lineare Prozeß darstellen als:

Def. 2.3.5 (*Allgemeiner linearer Prozeß*)

Sei $\{\epsilon_t\} \sim \text{WN}(0, \sigma^2)$ und $\psi(L) = \sum_{j=-\infty}^{\infty} \psi_j L^j$ ein absolut summierbarer linearer Filter, d.h.

$$\sum_{j=-\infty}^{\infty} |\psi_j| < \infty,$$

dann heißt

$$\psi(L)\epsilon_t = \sum_{j=-\infty}^{\infty} \psi_j L^j \epsilon_t = \sum_{j=-\infty}^{\infty} \psi_j \epsilon_{t-j}$$

allgemeiner linearer Prozeß. ■

Es ist nützlich, kovarianzerzeugende Funktionen als z-Transformation der Autokovarianzfunktion einzuführen:

Def. 2.3.6 (*Kovarianzerzeugende Funktion*)

Die kovarianzerzeugende Funktion $G_X(z)$ eines stationären Prozesses $\{X_t\}$ mit der Autokovarianzfunktion γ_τ ist die komplexwertige Funktion

$$G_X(z) = \sum_{\tau=-\infty}^{\infty} \gamma_\tau z^{-\tau},$$

vorausgesetzt, die Reihe konvergiert in einem ringförmigen Bereich $\{r^{-1} < |z| < r\}$ mit $r > 1$. ■

Durch die Angabe des Konvergenzbereiches $\{r^{-1} < |z| < r\}$ mit $r > 1$ ist gewährleistet, daß der Einheitskreis im Konvergenzbereich enthalten ist, so daß als Spezialfall die Fouriertransformation mit $z = e^{i2\pi\omega}$ existiert, so daß gilt:

$$(2.25) \quad f_X(\omega) = G_X(e^{i2\pi\omega}).$$

Der folgende Satz beschreibt nun allgemein, wie bei bekannter Autokovarianzfunktion eines Inputprozesses und bekanntem linearen Filter die Autokovarianzfunktion des Outputprozesses abgeleitet werden kann (vgl. z.B. NEWTON[1988, S. 79ff], SCHLITTEGEN/STREITBERG [1984, S. 87ff]):

Satz 2.3.2 (*Filtertheorem*)

Sei $\{X_t\}$ ein schwach stationärer stochastischer Prozeß mit Erwartungswert μ_X und der Autokovarianzfunktion $\gamma_X(\tau)$. Der stochastische Prozeß $\{Y_t\}$ ergebe sich als:

$$Y_t = \sum_{j=-\infty}^{\infty} \psi_j X_{t-j}$$

mit

$$\sum_{j=-\infty}^{\infty} |\psi_j| < \infty.$$

Dann gilt: $\{Y_t\}$ ist schwach stationär mit dem Erwartungswert

$$(2.26) \quad E(Y_t) = \mu_X \sum_{j=-\infty}^{\infty} \psi_j$$

und der Autokovarianzfunktion

$$(2.27) \quad \gamma_Y(\tau) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \psi_j \psi_k^* \gamma_X(\tau - j + k)$$

sowie der kovarianzerzeugenden Funktion:

$$(2.28) \quad G_Y(z) = G_X(z) G_\psi(z) G_\psi^*(1/z^*)$$

mit

$$G_\psi(z) = \sum_{j=-\infty}^{\infty} \psi_j z^{-j}.$$

■

Die Funktion $G_\psi(z) = \sum_{j=-\infty}^{\infty} \psi_j z^{-j}$ ist die z-Transformierte der Impulsantwortfunktion des Filters und wird auch als Transferfunktion des Filters bezeichnet.

Wesentliche Aussage des Filtertheorems ist, daß ein stationärer Prozeß durch einen stabilen Filter mit absolut summierbarer Impulsantwort wieder in einen stationären Prozeß transformiert wird.

Ist der Inputprozeß des linearen Filters ein White-Noise-Prozeß $\{\epsilon_t\} \sim \text{WN}(0, \sigma_\epsilon^2)$, so vereinfacht sich die Darstellung von Erwartungswert und Autokovarianzfunktion des Outputprozesses $\{Y_t\}$ zu:

$$(2.29) \quad E(Y_t) = \sum_{j=-\infty}^{\infty} \psi_j E(\epsilon_{t-j}) = 0$$

$$(2.30) \quad \gamma_Y(\tau) = \sigma_\epsilon^2 \sum_{j=-\infty}^{\infty} \psi_j \psi_{j-\tau}^*$$

und die kovarianzerzeugende Funktion ergibt sich in diesem Fall als:

$$(2.31) \quad G_Y(z) = \sigma_\epsilon^2 G_\psi(z) G_\psi^*(1/z^*).$$

Eine wichtige Klasse von stochastischen Prozessen, die unmittelbar aus einem White-Noise-Prozeß abgeleitet werden, sind Moving-Average-Prozesse der Ordnung q :

Def. 2.3.7 (*Moving-Average-Prozeß*)

Ein stochastischer Prozeß $\{X_t, t \in \mathcal{Z}\}$ heißt *Moving-Average-Prozeß der Ordnung q* bzw. *MA(q)-Prozeß* mit den Koeffizienten $\boldsymbol{\theta} = (\theta_0, \dots, \theta_q)'$ und der Innovationsprozeßvarianz σ_ϵ^2 , wenn er sich darstellen läßt als:

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

bzw.

$$X_t = \sum_{j=0}^q \theta_j \epsilon_{t-j} = \theta(L) \epsilon_t$$

mit

$$\theta_0 = 1, \theta_q \neq 0 \quad \text{und} \quad \{\epsilon_t\} \sim \text{WN}(0, \sigma_\epsilon^2).$$

■

Folgt ein stochastischer Prozeß $\{X_t\}$ einem Moving-Average-Prozeß, so wird dies angegeben als $\{X_t\} \sim \text{MA}(q, \boldsymbol{\theta}, \sigma^2)$ oder einfach als $\{X_t\} \sim \text{MA}(q)$.

Ist $\{X_t\}$ ein MA(q)-Prozeß gemäß Def. 2.3.7, so gilt für Erwartungswert und Autokovarianzfunktion:

$$E(X_t) = \sum_{j=0}^q \theta_j E \epsilon_{t-j} = 0 \quad \forall t \in \mathcal{Z}$$

$$\text{Cov}(X_t, X_{t+\tau}) = \gamma_X(\tau) = \begin{cases} 0 & \tau > q \\ \sigma_\epsilon^2 \sum_{j=\tau}^q \theta_j \theta_{j-\tau}^* & 0 \leq \tau \leq q \\ \gamma_X^*(-\tau) & \tau < 0. \end{cases}$$

Die Stationarität von MA(q)-Prozessen ergibt sich also unmittelbar aus ihrer Definition.

Ist eine Autokovarianzfunktion gegeben, der man einen MA($q, \boldsymbol{\theta}, \sigma^2$)-Prozeß zuordnen will, so ergeben sich allerdings Identifikationsprobleme: Im allgemeinen ist die Zuordnung zwischen einer Autokovarianzfunktion und einem MA($q, \boldsymbol{\theta}, \sigma^2$)-Prozeß nicht eindeutig. So besitzen zum Beispiel die beiden MA(1)-Prozesse

$$X_t \sim \text{MA}(1, \theta_1, \sigma_\epsilon^2) \quad \text{und} \quad Y_t \sim \text{MA}(1, 1/\theta_1, \sigma_\epsilon^2 \theta_1 \theta_1^*)$$

die gleiche Autokovarianzfunktion

$$\gamma_X(\tau) = \gamma_Y(\tau) = \begin{cases} \sigma_\epsilon^2(1 + \theta_1 \theta_1^*) & \tau = 0 \\ \sigma_\epsilon^2 \theta_1 & \tau = 1 \\ \sigma_\epsilon^2 \theta_1^* & \tau = -1 \\ 0 & \text{sonst.} \end{cases}$$

Eindeutigkeit kann erreicht werden, indem man sich z.B. auf die Klasse der kausal invertierbaren MA(q)-Prozesse beschränkt (vgl. Satz 2.3.1), d.h. man betrachtet nur diejenigen MA(q)-Prozesse, bei denen alle Wurzeln des charakteristischen Polynoms

$$\theta(z) = 1 + \theta_1 z^{-1} + \dots + \theta_q z^{-q}$$

innerhalb des Einheitskreises liegen.

Die vereinfachende Annahme der Invertierbarkeit kann allerdings nicht immer aufrechterhalten werden: ÅSTRÖM et al. [1984] stellen fest, daß bei der diskreten Abtastung von zeitstetigen Prozessen nicht invertierbare MA-Polynome auftreten können. ROBINSON[1983, S. 148] sowie MATSUOKA/ULRYCH[1984] verweisen vor dem Hintergrund

seismographischer Anwendungen darauf, daß die Annahme der Invertierbarkeit von MA-Prozessen nicht a priori gerechtfertigt werden kann.

Eine für die Spektralschätzung besonders wichtige Klasse von stochastischen Prozessen ist die Klasse der autoregressiven Prozesse:

Def. 2.3.8 (*Autoregressiver Prozeß*)

Ein stochastischer Prozeß $\{X_t, t \in \mathcal{Z}\}$ heißt autoregressiver Prozeß der Ordnung p bzw. AR(p)-Prozeß mit den Koeffizienten $\boldsymbol{\phi} = (\phi_0, \dots, \phi_p)'$ und der Innovationsprozeßvarianz σ_ϵ^2 , wenn er sich darstellen läßt als:

$$X_t + \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} = \epsilon_t \quad \forall t \in \mathcal{Z}$$

bzw.

$$\phi(L)X_t = \sum_{j=0}^p \phi_j X_{t-j} = \epsilon_t$$

mit

$$\phi_0 = 1, \phi_p \neq 0 \quad \text{und} \quad \{\epsilon_t\} \sim \text{WN}(0, \sigma_\epsilon^2).$$

■

Analog wie bei MA(q)-Prozessen werden die Kurzschreibweisen $\{X_t\} \sim \text{AR}(p, \boldsymbol{\phi}, \sigma^2)$ bzw. $\{X_t\} \sim \text{AR}(p)$ verwendet.

Ein AR(p)-Prozeß $\{X_t\}$ ist dann stationär, wenn er sich darstellen läßt als:

$$(2.32) \quad \begin{aligned} X_t &= \phi^{-1}(L)\epsilon_t \\ &= \sum_{j=-\infty}^{\infty} \psi_j \epsilon_{t-j}, \end{aligned}$$

d.h. die Stationarität eines AR(p)-Prozesses kann anhand der Existenz eines zu $\phi(L)$ inversen Filters definiert werden. Im Falle von kausalen Filtern ergibt sich damit als Stationaritätsbedingung für AR(p)-Prozesse, daß alle Wurzeln des charakteristischen Polynoms

$$1 + \phi_1 z^{-1} + \dots + \phi_p z^{-p}$$

innerhalb des Einheitskreises liegen müssen. In diesem Fall existiert eine kausale MA(∞)-Darstellung eines AR(p)-Prozesses:

$$(2.33) \quad X_t = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j}.$$

Die MA(∞)-Darstellung (2.33) eines kausalen AR(p)-Prozesses ist ein Spezialfall von (2.29) bzw. (2.30), so daß Erwartungswert und Autokovarianzfunktion sich angeben lassen als:

$$E(X_t) = \sum_{j=0}^{\infty} \psi_j E(\epsilon_{t-j}) = 0$$

und

$$\text{Cov}(X_t, X_{t+\tau}) = \gamma_X(\tau) = \sigma_\epsilon^2 \sum_{j=\tau}^{\infty} \psi_j \psi_{j-\tau}^*.$$

Die Darstellung der Autokovarianzfunktion eines AR(p)-Prozesses über seine MA(∞)-Darstellung ist allerdings nicht sehr praktikabel. Es ist im allgemeinen einfacher, die Autokovarianzfunktion implizit über eine Differenzgleichung zu spezifizieren. Hierzu geht man von der Definition eines AR(p)-Prozesses aus:

$$X_t + \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} = \epsilon_t,$$

multipliziert beide Seiten mit $X_{t-\tau}^*$, $\tau > 0$ und bildet die Erwartungswerte:

$$(2.34) \quad E(X_t X_{t-\tau}^*) + \phi_1 E(X_{t-1} X_{t-\tau}^*) + \dots + \phi_p E(X_{t-p} X_{t-\tau}^*) = E(\epsilon_t X_{t-\tau}^*),$$

so daß sich ergibt:

$$(2.35) \quad \gamma_\tau + \phi_1 \gamma_{\tau-1} + \dots + \phi_p \gamma_{\tau-p} = 0.$$

Die Autokovarianzfunktion eines AR(p)-Prozesses läßt sich also durch eine deterministische lineare homogene Differenzgleichung beschreiben, deren allgemeine Lösung lautet:

$$(2.36) \quad \gamma_X(\tau) = c_1 \lambda_1^\tau + c_2 \lambda_2^\tau + \dots + c_p \lambda_p^\tau,$$

wobei $\lambda_1, \dots, \lambda_p$ die Wurzeln des charakteristischen Polynoms

$$1 + \phi_1 z^{-1} + \dots + \phi_p z^{-p}$$

des zugrundeliegenden AR(p)-Prozesses sind. Aus Vereinfachungsgründen ist hierbei unterstellt, daß gilt: $\lambda_i \neq \lambda_j$.

Soll die Autokorrelationsfunktion eines AR(p)-Prozesses in der expliziten Form (2.36) dargestellt werden, dann sind die Konstanten c_1, \dots, c_p so zu wählen, daß die durch die folgenden Gleichungen vorgegebenen p Anfangsbedingungen

$$(2.37) \quad \gamma_\tau + \phi_1 \gamma_{\tau-1} + \dots + \phi_p \gamma_{\tau-p} = 0 \quad \tau = 1, \dots, p$$

bzw. in Matrixform und mit $\gamma_{-\tau} = \gamma_\tau^*$

$$(2.38) \quad \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_p \end{bmatrix} = - \begin{bmatrix} \gamma_0 & \gamma_1^* & \cdots & \gamma_{p-1}^* \\ \gamma_1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \gamma_1^* \\ \gamma_{p-1} & \cdots & \gamma_1 & \gamma_0 \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{bmatrix}$$

erfüllt sind.

Die Gleichungen (2.37) bzw. (2.38) heißen Yule-Walker-Gleichungen und sind für die Schätzung von AR(p)-Prozessen von besonderer Bedeutung. Die erweiterten Yule-Walker-Gleichungen erhält man, indem die Gleichungen (2.37) um die Gleichung für $\tau = 0$:

$$\gamma_0 + \phi_1 \gamma_{-1} + \dots + \phi_p \gamma_{-p} = \sigma_\epsilon^2.$$

ergänzt wird. In Matrixform lassen sich die erweiterten Yule-Walker-Gleichungen darstellen als:

$$(2.39) \quad \begin{bmatrix} \gamma_0 & \gamma_1^* & \cdots & \gamma_p^* \\ \gamma_1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \gamma_1^* \\ \gamma_p & \cdots & \gamma_1 & \gamma_0 \end{bmatrix} \begin{bmatrix} 1 \\ \phi_1 \\ \vdots \\ \phi_p \end{bmatrix} = \begin{bmatrix} \sigma_\epsilon^2 \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

AR(p)- und MA(q)-Prozesse können als Spezialfall von ARMA(p, q)-Prozessen betrachtet werden:

Def. 2.3.9 (*Autoregressiver-Moving-Average-Prozeß*)

Ein stochastischer Prozeß $\{X_t, t \in \mathcal{Z}\}$ heißt Autoregressiver-Moving-Average-Prozeß der Ordnung (p, q) bzw. ARMA(p, q)-Prozeß mit den Koeffizienten $\boldsymbol{\phi} = (\phi_0, \dots, \phi_p)'$, $\boldsymbol{\theta} = (\theta_0, \dots, \theta_q)'$ und der Innovationsprozeßvarianz σ^2 , wenn er sich darstellen läßt als:

$$X_t + \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} = \epsilon_t + \theta_1 \epsilon_{t-1} + \dots + \theta_q \epsilon_{t-q} \quad \forall t \in \mathcal{Z}$$

bzw.

$$\phi(L)X_t = \theta(L)\epsilon_t$$

mit

$$\phi_0 = 1, \phi_p \neq 0, \theta_0 = 1, \theta_q \neq 0 \quad \text{und} \quad \{\epsilon_t\} \sim \text{WN}(0, \sigma^2).$$

■

Folgt ein Prozeß einem ARMA(p, q)-Prozeß so wird dies in Kurzschreibweise angegeben als $\{X_t\} \sim \text{ARMA}(p, q, \boldsymbol{\phi}, \boldsymbol{\theta}, \sigma^2)$ bzw. einfach als $\{X_t\} \sim \text{ARMA}(p, q)$.

Ein ARMA(p, q)-Prozeß heißt stationär und kausal, wenn er sich darstellen läßt als

$$(2.40) \quad X_t = \frac{\theta(L)}{\phi(L)} \epsilon_t = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j}$$

und invertierbar, wenn die Darstellung

$$(2.41) \quad \epsilon_t = \frac{\phi(L)}{\theta(L)} X_t = \sum_{j=0}^{\infty} \xi_j X_{t-j},$$

existiert.

Für einen kausalen und invertierbaren ARMA(p, q)-Prozeß existiert also sowohl eine MA(∞)- als auch eine AR(∞)-Darstellung.

Aus $\{\epsilon_t\} \sim \text{WN}(0, \sigma^2)$ ergibt sich unmittelbar mit der Darstellung (2.40), daß der Erwartungswert eines stationären ARMA(p, q)-Prozesses gleich Null ist.

Zur Ableitung der Autokovarianz- bzw. Autokorrelationsfunktion kann man analog wie bei einem AR(p)-Prozeß vorgehen:

Multipliziert man beide Seiten eines ARMA(p, q)-Prozesses

$$\phi(L)X_t = \theta(L)\epsilon_t$$

mit $X_{t-\tau}^*$ und bildet die Erwartungswerte, so erhält man:

$$(2.42) \quad \begin{aligned} \gamma_X(\tau) + \phi_1 \gamma_X(\tau - 1) + \dots + \phi_p \gamma_X(\tau - p) \\ = \gamma_{X\epsilon}(\tau) + \theta_1 \gamma_{X\epsilon}(\tau - 1) + \dots + \theta_q \gamma_{X\epsilon}(\tau - q) \end{aligned}$$

bzw.

$$(2.43) \quad \gamma_X(\tau) = - \sum_{j=1}^p \phi_j \gamma_X(\tau - j) + \sum_{j=1}^q \theta_j \gamma_{X\epsilon}(\tau - j)$$

mit den Kreuzkovarianzen:

$$\gamma_{X\epsilon}(\tau) = E(\epsilon_t X_{t-\tau}^*).$$

Da sich die Kreuzkovarianzen $\gamma_{X\epsilon}(\tau)$ mit (2.40) darstellen lassen als:

$$\gamma_{X\epsilon}(\tau) = E(\epsilon_t \epsilon_{t-\tau}^* + \sum_{j=1}^{\infty} \psi_j^* \epsilon_t \epsilon_{t-\tau-j}^*) = \begin{cases} 0 & \tau > q \\ \sigma_\epsilon^2 & \tau = 0 \\ \sigma_\epsilon^2 \sum_{j=\tau}^{\infty} \psi_{\tau-j}^* & 0 < \tau \leq q, \end{cases}$$

ergibt sich die Autokovarianzfunktion eines ARMA(p, q)-Prozesses als:

$$(2.44) \quad \gamma_X(\tau) = \begin{cases} - \sum_{j=1}^p \phi_j \gamma_X(\tau - j) & \tau > q \\ - \sum_{j=1}^p \phi_j \gamma_X(\tau - j) + \sigma_\epsilon^2 \sum_{j=\tau}^q \theta_j \psi_{j-\tau}^* & 0 \leq \tau \leq q \\ \gamma^*(-\tau) & \tau < 0. \end{cases}$$

Eine Beschreibung von Algorithmen zur effizienten numerischen Berechnung von Autokovarianzfunktionen für (reelle) ARMA-Prozesse findet sich in NEWTON[1988, Abschnitt 2.6].

Aus Vereinfachungsgründen sind bislang nur stochastische Prozesse mit dem Erwartungswert Null betrachtet worden. Von Null verschiedene Erwartungswerte können,

z.B. bei ARMA-Prozessen, durch Betrachtung eines (mittelwertstationären) Prozesses $\{X_t\}$ der Form

$$(X_t - \mu) + \phi_1(X_{t-1} - \mu) + \dots + \phi_p(X_{t-p} - \mu) = \epsilon_t + \theta_1\epsilon_{t-1} + \dots + \theta_q\epsilon_{t-q} \quad \forall t \in \mathcal{Z}$$

mit $E(X_t) = \mu$ berücksichtigt werden.

2.4 Lineare Filter und stochastische Prozesse im Frequenzbereich

Die Analyse von zeitinvarianten linearen Filtern im Frequenzbereich ist häufig einfacher und anschaulicher als die Betrachtung im Zeitbereich.

Zunächst sei als Erweiterung des Filtertheorems (vgl. Satz 2.3.2) der folgende Satz formuliert, der eine einfache Beziehung zwischen den Spektraldichten eines Inputprozesses $\{X_t\}$ und eines Outputprozesses $\{Y_t\}$ ausdrückt:

Satz 2.4.1 (Filter im Frequenzbereich)

Sei $\{X_t\}$ ein schwach stationärer stochastischer Prozeß mit der Spektraldichte $f_X(\omega)$.

Der stochastische Prozeß $\{Y_t\}$ ergebe sich als:

$$Y_t = \sum_{j=-\infty}^{\infty} \psi_j X_{t-j}$$

mit

$$\sum_{j=-\infty}^{\infty} |\psi_j| < \infty.$$

Dann ergibt sich die Spektraldichte $f_Y(\omega)$ mit $-0.5 \leq \omega \leq 0.5$ als:

$$(2.45) \quad f_Y(\omega) = |A_\psi(\omega)|^2 f_X(\omega),$$

wobei:

$$A_\psi(\omega) = \sum_{j=-\infty}^{\infty} \psi_j e^{-i2\pi\omega j}$$

■

Die komplexwertige Funktion $A_\psi(\omega)$ ergibt sich als Fouriertransformierte der Impulsantwortfolge $\{\psi_j\}$ und heißt Frequenzantwortfunktion.

Wenn $\Im(\cdot)$ und $\Re(\cdot)$ den Imaginär- bzw. Realteil eines komplexwertigen Arguments bezeichnen, dann kann mit Hilfe der Gainfunktion $\Gamma_\psi(\omega)$ und der Phasenfunktion $\Upsilon_\psi(\omega)$ die Frequenzantwortfunktion dargestellt werden als:

$$(2.46) \quad A_\psi(\omega) = \Gamma_\psi(\omega) e^{i2\pi\Upsilon_\psi(\omega)},$$

wobei:

$$\Gamma_\psi(\omega) = |A_\psi(\omega)| = \sqrt{\Re(A_\psi(\omega))^2 + \Im(A_\psi(\omega))^2}$$

und

$$\Upsilon_\psi(\omega) = \arctan2[\Im(A_\psi(\omega)), \Re(A_\psi(\omega))] \quad -\pi < \Upsilon_\psi(\omega) \leq \pi,$$

wobei die Funktion $\arctan2(z)$ einer komplexen Zahl $z = a + bi$ definiert ist als:

$$\arctan2(b, a) = \begin{cases} \arctan(b/a) & a > 0 \\ \arctan(b/a) + \operatorname{sgn}(b)\pi & a < 0 \\ \operatorname{sgn}(b)\pi/2 & a = 0. \end{cases}$$

Als Kurzbezeichnung werden die Ausdrücke Gain und Phase verwandt.

Die graphischen Darstellungen von Gain bzw. Phase gegenüber der Frequenz ω heißen Gain- bzw. Phasendiagramm.

Da gilt

$$e^{i2\pi(\Upsilon_\psi(\omega)+k)} = e^{i2\pi\Upsilon_\psi(\omega)} e^{i2\pi k} = e^{i2\pi\Upsilon_\psi(\omega)}, \quad \forall k \in \mathbb{Z}$$

ist die Phase $\Upsilon_\psi(\omega)$ nur bis auf die Addition ganzer Zahlen k definiert.

Während die Interpretation des Gaindiagramms unproblematisch ist, ergeben sich beim Phasendiagramm Schwierigkeiten. Da $\Upsilon(\cdot)$ nicht eindeutig bestimmt, sondern auf die Hauptwerte im Bereich $(-\pi, \pi]$ beschränkt ist, können im Phasendiagramm Unstetigkeiten mit Sprüngen in der Größe von 2π auftreten. Ferner können Wurzeln des charakteristischen Polynoms eines Filters, die auf dem Einheitskreis liegen, zusätzliche Unstetigkeiten verursachen. Als Beispiel ist in Abbildung 2.1 das Phasendiagramm eines Filters mit der Transferfunktion

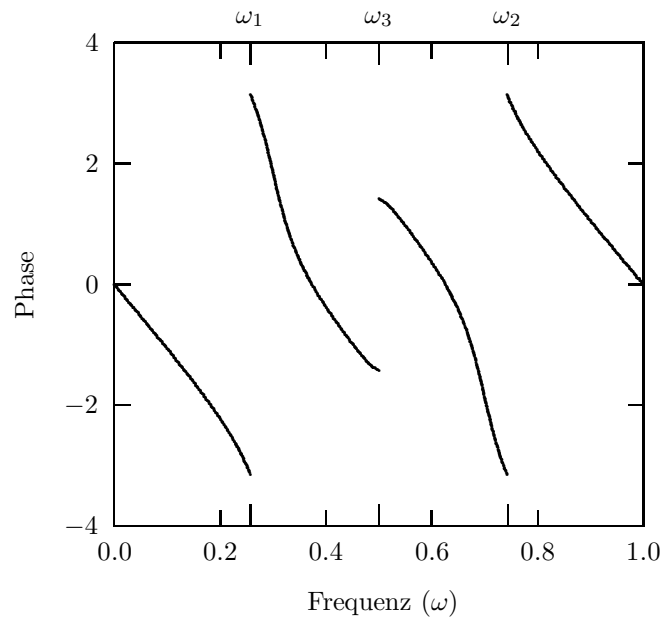


Abb. 2.1: Phasendiagramm eines Filters mit der Transferfunktion: $G(z) = (1 - 1.25e^{i0.6\pi}z^{-1})(1 - 1.25e^{-i0.6\pi}z^{-1})(1 + z^{-1})$, Sprungstellen durch Hauptwertbildung: $\omega_1 \approx 0.257$, $\omega_2 \approx 0.743$, Sprungstelle durch Wurzel auf dem Einheitskreis: $\omega_3 = 0.5$.

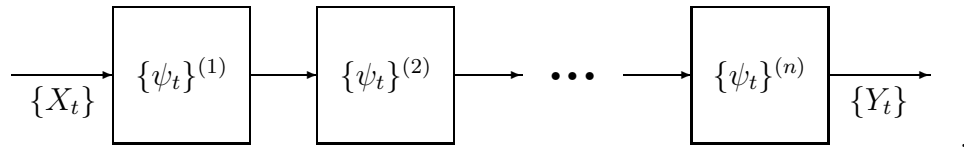
$$\begin{aligned}
 (2.47) \quad G(z) &= (1 - 1.25e^{i0.6\pi}z^{-1})(1 - 1.25e^{-i0.6\pi}z^{-1})(1 + z^{-1}) \\
 &= 1 + 1.7725z^{-1} + 2.3450z^{-2} + 1.5625z^{-3}
 \end{aligned}$$

angegeben.

Für Filter, deren charakteristisches Polynom keine Wurzel auf dem Einheitskreis besitzt, ist es möglich aus einer diskontinuierlichen Phasenfunktion eine eindeutig bestimmte kontinuierliche Phasenfunktion zu konstruieren („Entpacken einer Phase“). Numerische Vorgehensweisen hierzu sowie relevante Anwendungsgebiete sind in OPPENHEIM/SCHAFFER[1992, Kap. 12] sowie LIM[1990, S. 122ff und 293ff] beschrieben.

In Satz 2.4.1 wurde angegeben, wie Spektraldichte, Gain und Phase eines Outputprozesses bei Anwendung eines Filters auf einen Inputprozeß berechnet werden können. Im folgenden wird kurz die Erweiterung bei mehrfacher Filterung eines Prozesses angegeben.

Durchläuft ein Inputprozeß $\{X_t\}$ eine Kaskade mehrerer Filter $\{\psi_j\}^{(1)}, \{\psi_j\}^{(2)}, \dots, \{\psi_j\}^{(n)}$:



dann ergibt sich die Spektraldichtefunktion des Outprozesses $\{Y_t\}$ als:

$$(2.48) \quad f_Y(\omega) = |A_{\psi^{(n)}}(\omega)|^2 |A_{\psi^{(n-1)}}(\omega)|^2 \dots |A_{\psi^{(1)}}(\omega)|^2 f_X(\omega)$$

mit

$$(2.49) \quad A_{\psi^{(k)}}(\omega) = \sum_{j=-\infty}^{\infty} \psi_j^{(k)} e^{-i2\pi\omega j} \quad k = 1, \dots, n$$

und der Gainfunktion:

$$(2.50) \quad \Gamma(\omega) = \Gamma_{\psi^{(n)}}(\omega) \Gamma_{\psi^{(n-1)}}(\omega) \dots \Gamma_{\psi^{(1)}}(\omega)$$

sowie der Phasenverschiebung:

$$(2.51) \quad \Upsilon(\omega) = \Upsilon_{\psi^{(n)}}(\omega) + \Upsilon_{\psi^{(n-1)}}(\omega) + \dots + \Upsilon_{\psi^{(1)}}(\omega) + \kappa(\omega)2\pi.$$

Der Korrekturterm $\kappa(\omega)2\pi$ mit $\kappa : [0, 1] \mapsto \mathcal{Z}$ in (2.51) ist erforderlich, da die Funktion $\arctan2(\cdot)$ definitionsgemäß auf die Hauptwerte im Bereich $(-\pi, \pi]$ beschränkt ist und die Summe der einzelnen Hauptwerte der Phasen im allgemeinen ungleich dem Hauptwert der gesamten Phase ist (vgl. OPPENHEIM/SCHAFFER[1992, S. 246]).

Aus der Autokovarianzfunktion eines White-Noise-Prozesses $\{\epsilon_t\}$,

$$\gamma_\epsilon(\tau) = \begin{cases} \sigma_\epsilon^2 & \tau = 0 \\ 0 & \text{sonst,} \end{cases}$$

ergibt sich unmittelbar mit Satz 2.2.3 die Spektraldichte eines White-Noise-Prozesses als:

$$(2.52) \quad f_\epsilon(\omega) = \sigma_\epsilon^2 \quad \omega \in [-0.5, 0.5]$$

Da sich ein ARMA-Prozeß als gefilterter White-Noise-Prozeß darstellen läßt,

$$(2.53) \quad X_t = \frac{\theta(L)}{\phi(L)}\epsilon_t = \sum_{j=-\infty}^{\infty} \psi_j \epsilon_{t-j} \quad \{\epsilon_t\} \sim \text{WN}(0, \sigma_{\epsilon_t}^2)$$

bildet Satz (2.4.1) zusammen mit der Spektraldichte des White-Noise-Prozesses (2.52) Grundlage für die Ableitung der Spektraldichte eines ARMA-Prozesses. Die Spektren für MA- und AR-Prozesse ergeben sich als Spezialfall.

Satz 2.4.2 (*Spektraldichten von ARMA-Prozessen*)

Sei $\{X_t\}$ ein ARMA-Prozeß mit der Darstellung

$$\phi(L)X_t = \theta(L)\epsilon_t$$

wobei die charakteristischen Polynome $\phi(z) = 1 + \phi_1 z^{-1} + \dots + \phi_p z^{-p}$ und $\theta(z) = 1 + \theta_1 z^{-1} + \dots + \theta_q z^{-q}$ keine gemeinsamen Nullstellen besitzen und $\phi(z)$ keine Nullstellen auf dem Einheitskreis besitzt.

Dann läßt sich die Spektraldichte von $\{X_t\}$ darstellen als:

$$(2.54) \quad f_X(\omega) = \sigma_\epsilon^2 \frac{|\sum_{j=0}^q \theta_j e^{-i2\pi\omega j}|^2}{|\sum_{j=0}^p \phi_j e^{-i2\pi\omega j}|^2} \quad \omega \in [-0.5, 0.5].$$

■

Für die Darstellung der Spektraldichtefunktionen von ARMA-Prozessen ist es nicht notwendig, Kausalität oder Invertierbarkeit zu unterstellen (vgl. BROCKWELL/DAVIS[1990, S. 123]).

Aus (2.54) ergibt sich für $\phi = (1, 0, \dots, 0)'$ die Spektraldichte eines MA(q)-Prozeß $\{X_t\}$ als:

$$(2.55) \quad f_X(\omega) = \sigma_\epsilon^2 \left| \sum_{j=0}^q \theta_j e^{-i2\pi\omega j} \right|^2 \quad \omega \in [-0.5, 0.5].$$

Da für MA(q)-Prozesse $\{X_t\}$ gilt:

$$\gamma_X(\tau) = 0 \quad \text{für} \quad \tau > q$$

bietet es sich allerdings auch an, die Spektraldarstellung eines MA(q)-Prozesses unmittelbar aus der Fouriertransformation der von Null verschiedenen Autokovarianzen abzuleiten.

Mit $\boldsymbol{\theta} = (1, 0, \dots, 0)'$ und (2.54) hat die Spektraldichte eines AR(p)-Prozesses die Form:

$$(2.56) \quad f_X(\omega) = \frac{\sigma_\epsilon^2}{|\sum_{j=0}^p \phi_j e^{-i2\pi\omega j}|^2} \quad \omega \in [-0.5, 0.5].$$

Die Beziehungen (2.54), (2.55), (2.56) bilden die theoretische Basis für parametrische Ansätze der Schätzung von Spektraldichtefunktionen.

Da die Spektraldichten von AR(p)-Prozessen besonders geeignet sind, Spektren mit ausgeprägten Peaks darzustellen und für die Schätzung verschiedene effiziente und stabile Algorithmen zur Verfügung stehen, spielen AR(p)-Prozesse eine besondere Rolle bei der parametrischen Schätzung von Spektraldichten.

2.5 Schätzung von Momentenfunktionen

Die Schätzung des Erwartungswertes μ , der Autokovarianzfunktion $\{\gamma_\tau\}$ bzw. der Autokorrelationsfunktion $\{\rho_\tau\}$ eines stochastischen Prozesses $\{X_t, t \in \mathcal{Z}\}$ bilden die Grundlage von nichtparametrischen und parametrischen Verfahren für die Schätzung von Spektraldichten.

Da die Ableitung der Eigenschaften von Schätzfunktionen für die genannten Momente teilweise sehr langwierig und aufwendig ist, beschränkt sich die Darstellung hier auf eine kurze Zusammenfassung der wichtigsten Definitionen und Aussagen für *reelle* stochastische Prozesse. Detailliertere Ableitungen sowie weitere Literaturhinweise finden sich u.a. in BROCKWELL/DAVIS[1991, Kap. 7], PRIESTLEY[1981, Kap. 5], NEWTON[1988, Kap. 3] sowie SCHLITTGEN/STREITBERG [1984, Kap. 4].

Für die Schätzung der Momente eines stochastischen Prozesses steht häufig nur eine einzige Realisation des stochastischen Prozesses zur Verfügung. Die Notwendigkeit,

Konsistenzaussagen für Momentenschätzer von stochastischen Prozessen zu formulieren, wobei Längsschnittdaten statt Querschnittdaten zur Schätzung der Momente verwandt werden, führte zum Begriff der Ergodizität (vgl. z.B. SCHLITZGEN/STREITBERG[1984, S. 147]):

Def. 2.5.1 (*Ergodizität*)

Sei $\{X_t, t \in \mathcal{Z}\}$ ein schwach stationärer Prozeß. Dann heißt $\{X_t, t \in \mathcal{Z}\}$ mittelwertergodisch, wenn gilt:

$$\lim_{N \rightarrow \infty} E\left[\left(\frac{1}{N} \sum_{t=1}^N X_t - \mu\right)^2\right] = 0$$

und kovarianzergodisch wenn gilt:

$$\lim_{N \rightarrow \infty} E\left[\left(\frac{1}{N} \sum_{t=1}^N (X_t - \mu)(X_{t+\tau} - \mu) - \gamma_\tau\right)^2\right] = 0.$$

■

Hinreichende Bedingungen für Mittelwert- und Kovarianzergodizität werden im folgenden Satz formuliert (vgl. z.B. SCHLITZGEN/STREITBERG [1984, S. 148]):

Satz 2.5.1

Sei $\{X_t, t \in \mathcal{Z}\}$ ein schwach stationärer Prozeß.

Dann gilt:

- (i) Ist die Autokovarianzfunktion $\{\gamma_\tau\}$ absolut summierbar, dann ist $\{X_t\}$ mittelwertergodisch.
- (ii) Ist $\{X_t\}$ ein Normalprozeß und ist die Autokovarianzfunktion $\{\gamma_\tau\}$ absolut summierbar, dann ist $\{X_t\}$ kovarianzergodisch.

■

Die Möglichkeit, bei der Schätzung von Momenten stationärer Prozesse Ensembledaten sinnvoll durch zeitliche Mittel ersetzen zu können, beruht also im wesentlichen auf der absoluten Summierbarkeit der Autokovarianzfunktion. Dies impliziert, daß gelten muß:

$$\lim_{\tau \rightarrow \infty} \gamma_\tau = 0.$$

Anschaulich bedeutet dies nichts anderes, als daß die Zufallsvariablen (X_{t_l}, X_{t_k}) eines stochastischen Prozesses, wenn sie schon nicht unabhängig sind, zumindest für „weit“ auseinanderliegende Indizes t_l und t_k unkorreliert sein müssen, um konsistente Schätzungen der ersten beiden Momente zu ermöglichen. Es sei daran erinnert, daß stochastische Prozesse mit absolut summierbarer Autokovarianzfunktion eine stetige Spektraldichtefunktion besitzen (vgl. Satz (2.2.3)).

Liegt eine Realisation $\mathbf{x} = (x_1, \dots, x_N)'$ eines stochastischen Prozesses mit N Beobachtungen vor, so wird der empirische Mittelwert berechnet als:

$$(2.57) \quad \bar{x} = \frac{1}{N} \sum_{t=1}^N x_t.$$

Der folgende Satz sagt aus, daß die dem empirischen Schätzer (2.57) entsprechende Schätzfunktion konsistent ist (vgl. SCHLITGEN/STREITBERG[1984, S. 150]):

Satz 2.5.2

Sei $\{X_t, t \in \mathcal{Z}\}$ ein schwach stationärer stochastischer Prozeß mit absolut-summierbarer Autokovarianzfunktion $\{\gamma_\tau\}$ und sei

$$\bar{X} = \frac{1}{N} \sum_{t=1}^N X_t$$

eine Schätzfunktion für den Erwartungswert μ . Dann gilt:

$$E[\bar{X}] = E\left[\frac{1}{N} \sum_{t=1}^N X_t\right] = \mu$$

und

$$\lim_{N \rightarrow \infty} \text{Var}(\bar{X}) = \lim_{N \rightarrow \infty} \frac{1}{N^2} \sum_{r=1}^N \sum_{s=1}^N \gamma_{s-r} \lim_{N \rightarrow \infty} = \frac{1}{N} [\gamma_0 + 2 \sum_{\tau=1}^{N-1} \frac{N-\tau}{N} \gamma_\tau] = 0.$$

■

Die Schätzung der Autokovarianz- bzw. der Autokorrelationsfunktion eines stochastischen Prozesses ist ein Grundproblem der Spektraldichteschätzung, das insbesondere für kleine Stichproben nicht befriedigend gelöst ist.

Als Schätzer der Autokovarianzfunktion $\{\gamma_\tau, \tau \in \mathcal{Z}\}$ werden in der theoretisch orientierten Diskussion hauptsächlich zwei Alternativen einander gegenübergestellt:

Die „unverzerrte“ Schätzfunktion $\{\hat{\gamma}_\tau^u, \tau = 0, \pm 1, \dots, \pm(N-1)\}$ mit:

$$(2.58) \quad \hat{\gamma}_\tau^u = \frac{1}{N - |\tau|} \sum_{t=1}^{N-|\tau|} (X_t - \bar{X})(X_{t+|\tau|} - \bar{X})$$

und die „verzerrte“ Schätzfunktion $\{\hat{\gamma}_\tau, 0, \pm 1, \dots, \pm(N-1)\}$ mit:

$$(2.59) \quad \hat{\gamma}(\tau) = \frac{1}{N} \sum_{t=1}^{N-|\tau|} (X_t - \bar{X})(X_{t+|\tau|} - \bar{X}).$$

Die entsprechenden empirischen Autokovarianzschätzer lauten:

$$(2.60) \quad c_\tau^u = \frac{1}{N - |\tau|} \sum_{t=1}^{N-|\tau|} (x_t - \bar{x})(x_{t+|\tau|} - \bar{x})$$

und

$$(2.61) \quad c_\tau = \frac{1}{N} \sum_{t=1}^{N-|\tau|} (x_t - \bar{x})(x_{t+|\tau|} - \bar{x}).$$

Die Gegenüberstellung von „unverzerrter“ und „verzerrter“ Schätzung ist insofern nicht ganz zutreffend, als beide Schätzfunktionen bei unbekanntem Erwartungswert μ einen Bias aufweisen. Die Schätzfunktion $\{\hat{\gamma}_\tau^u\}$ ist allerdings bei bekanntem μ erwartungstreu für $\{\gamma_\tau\}$. Beide Schätzfunktionen sind auch bei unbekanntem μ , zumindest asymptotisch, unverzerrt.

Inzwischen hat sich bei der Schätzung der Autokovarianzfunktion die „verzerrte“ Schätzfunktion $\{\hat{\gamma}_\tau\}$ durchgesetzt, was im wesentlichen auf zwei Gründe zurückzuführen ist: Zum einen hat die „verzerrte“ Schätzfunktion $\{\hat{\gamma}_\tau\}$ einen kleineren mittleren quadratischen Fehler als die „unverzerrte“ Schätzfunktion $\{\hat{\gamma}_\tau^u\}$ (vgl. JENKINS/WATTS[1968, S. 178ff], PRIESTLEY[1981, S. 323ff]).

Zum andern ist man nicht nur an einer konsistenten Schätzung für einen einzelnen Wert der Autokovarianzfunktion interessiert, sondern an einer Schätzung der gesamten Autokovarianzfunktion, die möglichst alle Eigenschaften der theoretischen Autokovarianzen besitzen soll. Unter Berücksichtigung von Satz 2.1.7 heißt dies insbesondere, daß eine geeignete empirische Autokovarianzfunktion positiv-semidefinit sein soll. Diese Eigenschaft kann für den „verzerrten“ Schätzer $\{\hat{\gamma}_\tau\}$ gezeigt werden, aber nicht für den „unverzerrten“ Schätzer $\{\hat{\gamma}_\tau^u\}$.

Hierzu faktorisiert man die empirische Autokovarianzmatrix:

$$(2.62) \quad \mathbf{C}_N = \begin{bmatrix} c_0 & c_1 & \cdots & c_{N-1} \\ c_1 & \ddots & \ddots & c_{N-2} \\ \vdots & \ddots & \ddots & c_1 \\ c_{N-1} & \cdots & c_1 & c_0 \end{bmatrix}$$

als

$$(2.63) \quad N\mathbf{C}_N = \mathbf{Z}\mathbf{Z}',$$

mit der $(N \times 2N - 1)$ -Matrix \mathbf{Z} :

$$(2.64) \quad \mathbf{Z} = \begin{bmatrix} 0 & \cdots & 0 & z_1 & z_2 & \cdots & z_N \\ \vdots & \ddots & z_1 & z_2 & \cdots & z_N & 0 \\ 0 & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ z_1 & z_2 & \cdots & z_N & 0 & \cdots & 0 \end{bmatrix},$$

wobei $z_i = x_i - \bar{x}$ für $i = 1, \dots, N$.

Dann gilt für jeden Vektor $\mathbf{a} = (a_1, \dots, a_N)' \in \mathcal{R}^N$ und $N \geq 1$:

$$(2.65) \quad \mathbf{a}'\mathbf{C}_N\mathbf{a} = (\mathbf{a}'\mathbf{Z})(\mathbf{a}'\mathbf{Z})' \geq 0,$$

so daß die Matrix \mathbf{C} positiv-semidefinit ist.

Die Bedingung $c_0 > 0$ ist hinreichend und notwendig für die strenge Positiv-Definitheit von \mathbf{C}_N . Dies ist der Fall, wenn nicht alle x_1, \dots, x_N identisch sind (vgl. NEWTON[1988, S. 165]), was bei praktischen Anwendungen immer als gegeben unterstellt werden kann.

Daß die Autokovarianzschätzer $\{c_\tau^u\}$ zu nicht positiv-semidefiniten Kovarianzmatrizen führen können, kann an Hand eines Beispiels gezeigt werden (vgl. AOKI[1987, S. 254]):

Sei $N = 3$ und $\mathbf{x} = (x_1, x_2, x_3)' = (1, 0, -1)'$. Dann ergeben sich bei Anwendung des „unverzerrten“ Autokovarianzschätzers die empirischen Autokovarianzen als:

$$c_0^u = 2/3 \quad c_1^u = 0 \quad c_2^u = -1$$

und damit die Autokovarianzmatrix \mathbf{C}_3^u :

$$\mathbf{C}_3^u = \begin{bmatrix} 2/3 & 0 & -1 \\ 0 & 2/3 & 0 \\ -1 & 0 & 2/3 \end{bmatrix}.$$

Die Matrix \mathbf{C}_3^u besitzt die Eigenwerte

$$\lambda_1 = -1/3 \quad \lambda_2 = 2/3 \quad \lambda_3 = 5/3$$

und ist somit indefinit.

Mit den „verzerrten“ Autokovarianzschätzern ergeben sich die empirischen Autokovarianzen als

$$c_0 = 2/3 \quad c_1 = 0 \quad c_2 = -1/3,$$

so daß die Autokovarianzmatrix lautet:

$$\mathbf{C}_3 = \begin{bmatrix} 2/3 & 0 & -1/3 \\ 0 & 2/3 & 0 \\ -1/3 & 0 & 2/3 \end{bmatrix}.$$

Die Eigenwerte von \mathbf{C}_3 ,

$$\lambda_1 = 1/3 \quad \lambda_2 = 2/3 \quad \lambda_3 = 1,$$

sind erwartungsgemäß alle positiv, so daß die Kovarianzmatrix \mathbf{C}_3 positiv-definit ist.

Positiv-(Semi)Definitheit von empirischen Autokovarianzschätzern ist nicht nur wünschenswert, da dies einer Fundamenteleigenschaft der theoretischen Autokovarianzen entspricht (vgl. Satz 2.2.2 (Theorem von Herglotz) und Satz 2.1.7). Von besonderer praktischer Bedeutung ist diese Eigenschaft auch bei der Entwicklung von numerisch stabilen Algorithmen. So verweisen z.B. SCHLITGEN/STREITBERG [1984, S. 176] auf die Notwendigkeit von positiv-definiten empirischen Autokovarianzen für die Konvergenz des Wilson-Algorithmus bei der iterativen Schätzung der Parameter eines MA-Prozesses. Es muß allerdings darauf hingewiesen werden, daß zahlreiche weitere Möglichkeiten zur positiv-(semidefiniten) Schätzung von Autokovarianzen bestehen, auf die im Rahmen der Schätzung von autoregressiven Modellen eingegangen wird. Dabei lassen sich auch die zum Teil erheblichen Nachteile der verzerrten Autokovarianzschätzer $\{\hat{\gamma}_\tau\}$ klarer und anschaulicher beschreiben.

Dessen ungeachtet wird, im weiteren, sofern nicht ausdrücklich anders angegeben, als direkte Autokovarianzschätzung die „verzerrte“ Form $\{\hat{\gamma}_\tau\}$ betrachtet.

Unter den vereinfachenden Annahmen eines Normalprozesses mit bekanntem Erwartungswert $\mu = 0$ und absolut-summierbarer Autokovarianzfunktion werden im folgenden Satz die Eigenschaften und die asymptotische Verteilung des Autokovarianzschätzers $\hat{\gamma}_\tau$ angegeben, die Aussagen lassen sich aber mit geeigneten Modifikationen generell auf allgemeine lineare Prozesse der Form:

$$X_t - \mu = \sum_{j=-\infty}^{\infty} \phi_j \epsilon_{t-j} \quad \epsilon_t \sim \text{IID}(0, \sigma^2)$$

mit

$$\sum_{j=-\infty}^{\infty} |\phi_j| < \infty \quad \text{und} \quad E(\epsilon_t^4) < \infty$$

übertragen (vgl. z.B. BROCKWELL/DAVIS[1991, S. 220ff]).

Satz 2.5.3 (*Eigenschaften von $\hat{\gamma}_\tau$*)

Sei $\{X_t, t \in \mathcal{Z}\}$ ein Normalprozeß mit $\mu = 0$ und absolut-summierbarer Autokovarianzfunktion $\{\gamma_\tau\}$.

Dann gilt:

$$(2.66) \quad E(\hat{\gamma}_N(\tau)) = \left(1 - \frac{|\tau|}{N}\right) \gamma(\tau)$$

$$(2.67) \quad \lim_{N \rightarrow \infty} N \text{Cov}(\hat{\gamma}_N(k), \hat{\gamma}_N(l)) = \sum_{m=-\infty}^{\infty} (\gamma_m \gamma_{m+(k-l)} + \gamma_{m-l} \gamma_{m+k}).$$

Darüberhinaus sind die Größen

$$\sqrt{N}(\hat{\gamma}_N(0) - \gamma_0), \dots, \sqrt{N}(\hat{\gamma}_N(l) - \gamma_l)$$

asymptotisch normalverteilt mit Erwartungswertvektor $\mathbf{0} = (0, \dots, 0)$ und den in (2.67) angegebenen Kovarianzen. ■

Die Autokorrelationsschätzer $\{\hat{\rho}_N(\tau), \tau = 0, \pm 1, \dots, \pm(N-1)\}$ erhält man durch Normierung der Autokovarianzschätzer als:

$$\hat{\rho}_N(\tau) = \frac{\hat{\gamma}_N(\tau)}{\hat{\gamma}_N(0)},$$

entsprechend ergeben sich die empirischen Autokorrelationen als:

$$r_\tau = \frac{c_\tau}{c_0}.$$

Die Eigenschaften der Autokorrelationsschätzer sind im folgenden Satz zusammengefaßt:

Satz 2.5.4 (*Eigenschaften von $\hat{\rho}_\tau$*)

Sei $\{X_t, t \in \mathcal{Z}\}$ ein Normalprozeß mit $\mu = 0$ und absolut-summierbarer Autokovarianzfunktion $\{\gamma_\tau\}$.

Dann gilt:

$$(2.68) \quad E(\hat{\rho}_N(\tau)) \doteq \left(1 - \frac{|\tau|}{N}\right)\rho(\tau)$$

und

$$(2.69) \quad \lim_{N \rightarrow \infty} N \operatorname{Cov}(\hat{\rho}_N(k), \hat{\rho}_N(l)) = \sum_{m=-\infty}^{\infty} \left[\rho_{m+l}\rho_{m+k} + \rho_{m-l}\rho_{m+k} - 2\rho_k\rho_m\rho_{m+l} - 2\rho_l\rho_m\rho_{m+k} + 2\rho_l\rho_k\rho_m^2 \right].$$

Die Größen

$$\sqrt{N}(\hat{\rho}_N(1) - \rho_1), \dots, \sqrt{N}(\hat{\rho}_N(l) - \rho_l)$$

sind gemeinsam asymptotisch normalverteilt mit dem Erwartungswertvektor $\mathbf{0} = (0, \dots, 0)'$ und den in (2.69) angegebenen Kovarianzen. ■

Analog wie bei den Autokovarianzschätzern gelten die Aussagen von Satz (2.5.4) mit weniger strengen Annahmen.

Gleichung (2.69) verweist auf ein großes Problem bei der praktischen Arbeit mit Schätzern der Autokorrelationsfunktion: Die Schätzer sind sehr stark miteinander korreliert,

was die Interpretation eines einzelnen Schätzers sehr schwierig macht und ein erhebliches Abweichen einer geschätzten Autokorrelationsfunktion von der zugrundeliegenden theoretischen Autokorrelationsfunktion bewirken kann.

Diese Probleme trugen u.a. mit dazu bei, daß im Rahmen der autoregressiven Spektralschätzung die direkte Schätzung der Autokovarianzen entweder umgangen wurde und weitere Alternativen entwickelt wurden, die sich aus speziellen Schätztechniken ableiten.

So werden z.B. bei der Burg-Maximum-Entropie-Schätzung die AR-Parameter unter Umgehung einer expliziten Schätzung von Autokovarianzen direkt aus den Daten geschätzt. Auf ULRYCH/CLAYTON[1976] ist der Ansatz einer Kleinst-Quadrate-Schätzung für autoregressive Modelle zurückzuführen, bei der der Vorwärts- und Rückwärtsprognosefehler minimiert wird, wobei Autokovarianzmatrizen auftreten, die keine Toeplitzstruktur besitzen.

Die beiden genannten Alternativen sowie zahlreiche weitere Variationen (vgl. z.B. MARPLE[1987, S. 206ff]) sind unerlässlich, um bei bestimmten Anwendungsgebieten überhaupt sinnvolle Spektralschätzungen durchführen zu können.

2.6 Entropie, Information und stochastische Prozesse

In verschiedenen Teilen der Arbeit wird es erforderlich sein, sich auf die Entropie von stochastischen Prozessen oder daraus abgeleiteten Kriterien zu beziehen. In diesem Abschnitt werden nach einem kurzen Überblick die wichtigsten Definitionen und Aussagen zusammengefaßt.

Der Begriff Entropie wurde zuerst in der Physik verwandt. Im klassischen physikalischen Kontext der statistischen Mechanik ist Entropie eine Zustandsgröße, welche die Unordnung eines Systems charakterisiert. Das physikalische Konzept der Entropie ist von fundamentaler Bedeutung (vgl. z.B. ALONSO/FINN[1992, S. 426ff, 454ff]): Nach dem zweiten Hauptsatz der Thermodynamik verlaufen die Prozesse innerhalb eines geschlossenen Systems so, daß die Entropie niemals abnimmt, sondern zunimmt oder im Gleichgewichtsfall unverändert bleibt. Daraus wird u.a. die Hypothese abgeleitet, daß die nicht umkehrbare Gerichtetheit der kosmologischen Zeit an die Zunahme der Entropie gekoppelt ist.

SHANNON[1948]² entwickelte im Rahmen der Informations- und Kommunikationstheorie auf *axiomatischer Grundlage* ein Maß für die a priori Unsicherheit bzw. a posteriori Information von Zufallsexperimenten. Aufgrund einer formalen Analogie mit der Darstellung in der statistischen Mechanik nannte er dieses Maß ebenfalls Entropie, obwohl sein Konzept als unabhängige und eigenständige Entwicklung betrachtet werden kann. Allgemeine Darstellungen der informationstheoretischen Entropie mit unterschiedlichen Erweiterungen enthalten z.B. REZA[1961], RÉNYI[1970, Kap. IX], GİAŞU[1977]. KLIR/FOLGER[1988, Kap. 5] diskutieren neben dem SHANNON'schen Entropiekonzept und dessen Verallgemeinerungen auch Maße für Unsicherheit, die sich nicht auf ein wahrscheinlichkeitstheoretisches Konzept stützen.

Die für inferenzstatistische Zwecke bedeutendste Erweiterung des Shannon'schen Informationskonzeptes stammt von KULLBACK/LEIBLER[1951]. KULLBACK[1959] gibt eine detaillierte Darstellung des Konzeptes mit verschiedenen Anwendungen.

JAYNES[1957a,b] begründet die moderne Formulierung eines allgemeinen Konzeptes der Maximum-Entropie-Schätzung, worauf in Abschnitt 3.3 näher eingegangen wird.

Für die folgenden Ausführungen ist das thermodynamische Konzept der Entropie von

²Im folgenden wird nach der deutschen Übersetzung aus dem Jahr 1976 zitiert.

untergeordneter Bedeutung. Im Vordergrund stehen hier die inferenzstatistischen Implikationen des Entropie- bzw. Informationsbegriffs, wobei auch in diesem enger gesetzten Rahmen die inhaltliche Mehrfachbesetzung der beiden Begriffe eine klare Darstellung erschwert (vgl. auch 3.3). Der Tagungsband von ZUREK[1990] (vgl. darin z.B. ZUREK[1990a], RISSANEN[1990]) verdeutlicht die Problematik, unterschiedliche wissenschaftliche Zugänge zum Entropie- und Informationsbegriff in einen gemeinsamen interdisziplinären Rahmen zu fassen. Wenn im weiteren der Begriff Entropie verwandt wird, dann ist damit die Shannon-Entropie gemeint, Abweichungen werden besonders gekennzeichnet.

Die Entropie einer diskreten Zufallsvariable wird definiert als (vgl. Shannon[1976, S. 62ff]):

Def. 2.6.1 (*Entropie*)

Sei X eine diskrete Zufallsvariable, mit $\Pr(X = x_k) = p_k$, $k = 1, \dots, K$. Dann heißt

$$(2.70) \quad S(X) = S(p_1, \dots, p_K) = - \sum_{k=1}^K p_k \ln p_k,$$

Entropie der Zufallsvariablen X , wobei $0 \ln 0$ über den Grenzwert $\lim_{p_k \rightarrow 0} p_k \ln p_k = 0$ abgeleitet wird. ■

Die Shannon-Entropie gemäß Def. 2.6.1 kann mit Hilfe verschiedener axiomatisch vorgegebener Eigenschaften eindeutig abgeleitet werden, wodurch sie im Vergleich zu anderen statistischen Kenngrößen für Unsicherheit einen besonderen Stellenwert besitzt. Auf eine Diskussion des axiomatischen Hintergrundes wird hier verzichtet (vgl. hierzu detailliert KLIR/FOLGER[1988, S. 155ff]).

Die folgenden Eigenschaften charakterisieren die Entropie $S(X)$ als sinnvolles Maß für Unsicherheit (SHANNON[1976, S. 62ff]):

Satz 2.6.1 (*Elementare Eigenschaften - diskrete Entropie*)

(i) Sei X eine diskrete Zufallsvariable mit $\Pr(X = x_k) = p_k$, $k = 1, \dots, K$. Dann gilt:

$$0 \leq S(X) \leq \ln K.$$

Das Maximum $S(X) = \ln K$ wird genau dann erreicht, wenn gilt:

$$p_1 = p_2 = \cdots = p_K = 1/K,$$

d.h. wenn X diskret gleichverteilt ist. Das Minimum $S(X) = 0$ wird genau dann erreicht wenn ein p_k gleich 1 ist und alle übrigen gleich Null sind.

(ii) Seien X und Y zwei diskrete Zufallsvariablen mit der gemeinsamen Entropie

$$S(X, Y) = \sum_{k=1}^K \sum_{l=1}^L \Pr(X = x_k, Y = y_l) \ln \Pr(X = x_k, Y = y_l).$$

Dann gilt:

$$S(X, Y) \leq S(X) + S(Y),$$

mit Gleichheit genau dann, wenn X und Y unabhängig sind.

■

Damit besitzt eine diskrete Zufallsvariable dann maximale Entropie, wenn alle ihre Ausprägungen mit gleicher Wahrscheinlichkeit auftreten können. Ist die Entropie einer diskreten Zufallsvariablen gleich Null, so gibt es nur ein einziges, sicheres Ereignis. Die Entropie unabhängiger Zufallsvariablen ist additiv.

Neben der oben gegebenen Interpretation der Entropie $S(X)$ als Maß für *a priori* Unsicherheit existiert auch eine Interpretation der Entropie $S(X)$ als Maß für mittlere *a posteriori* Information. Dabei wird unterstellt, daß die Realisation eines unwahrscheinlichen Ereignisses einen höheren Informationsgehalt besitzt, als die Realisation eines wahrscheinlichen Ereignisses. Die *a posteriori* Information $\text{Inf}(X = x_k)$ einer Realisation x_k wird definiert durch (vgl. z.B. REZA[1961, S. 8], GUIAŞU[1977, S. 2])

$$\text{Inf}(X = x_k) = -\ln p_k,$$

so daß u.a. gilt $\text{Inf}(X = x_l) > \text{Inf}(X = x_k)$ wenn $p_l < p_k$. Die mittlere *a posteriori* Information ergibt sich als

$$E[\text{Inf}(X)] = -\sum_{k=1}^K p_k \ln p_k$$

und ist somit gleich der Entropie $S(X)$. Die *a posteriori* Information einer Realisation x_k wird auch als ihre Selbstinformation bezeichnet.

Analog zur Definition 2.6.1 läßt sich für eine stetige Zufallsvariable X mit der Dichtefunktion $f(x)$ die Entropie $S(X)$ angeben als:

$$(2.71) \quad S(X) = S(f, f) = - \int_{-\infty}^{\infty} f(x) \ln f(x) dx = E[-\ln f(X)].$$

Die Übertragung des diskreten Entropiekonzeptes auf stetige Verteilungen bereitet allerdings Probleme (vgl. z.B. REZA[1961, S. 267ff]): Im stetigen Fall muß die Entropie $S(X)$ im allgemeinen weder positiv noch endlich sein. Im Gegensatz zum diskreten Fall ist die Entropie einer stetigen Zufallsvariablen X nicht invariant gegenüber Transformationen von X . Dadurch ist die Entropie stetiger Zufallsvariablen kein absolutes Maß für Unsicherheit wie im diskreten Fall, sondern abhängig von dem zugrundegelegten Koordinatensystem, wobei die Transformation einer Zufallsvariablen, wie üblich, als Übergang auf ein anderes Koordinatensystem interpretiert wird. Trotzdem ist auch im stetigen Fall die Entropie ein nützliches Konzept, insbesondere besteht, wie weiter unten gezeigt wird, eine enge Beziehung zu grundlegenden inferenzstatistischen Ansätzen. Zunächst werden einige Aussagen für stetige Zufallsvariablen zusammengefaßt (vgl. z.B. REZA[1961, S. 278ff]):

Satz 2.6.2 (*Elementare Eigenschaften - stetige Entropie*)

- (i) Sei X eine beliebige stetige Zufallsvariable auf dem Intervall $[a, b]$ mit der Dichte $f_X(x)$. Dann gilt: $S(X) \leq \ln(b - a)$, mit Gleichheit genau dann, wenn X auf dem Intervall $[a, b]$ gleichverteilt ist.
- (ii) Sei X eine stetige Zufallsvariable mit vorgegebenem Erwartungswert $E(X) = 0$ und vorgegebener Varianz $\text{Var}(X) = \sigma^2$. Dann gilt: $S(X) \leq \ln \sqrt{2\pi e \sigma^2}$, mit Gleichheit genau dann, wenn X normalverteilt ist.
- (iii) Sei $\mathbf{x} = (X_1, \dots, X_N)'$ ein N -dimensionaler Zufallsvektor mit stetigen Zufallsvariablen, vorgegebenem Erwartungswert $E(\mathbf{x}) = \mathbf{0}$ und vorgegebenen zweiten Momenten

$$\mathbf{\Omega}[r, s] = \gamma(r, s) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} x_r x_s f(x_1, \dots, x_N) dx_1 \dots dx_N.$$

Dann gilt: $S(\mathbf{x}) \leq \ln(2\pi e)^{N/2} + \ln(\det \mathbf{\Omega})^{1/2}$, mit Gleichheit genau dann, wenn \mathbf{x} eine N -dimensionale Normalverteilung besitzt. ■

Mit Satz 2.6.2 (i) gilt für stetige Dichten $f_X(x)$ auf einem endlichen Intervall $[a, b]$ weitgehend die Interpretation der diskreten Entropie, die Entropie mißt dann in beiden Fällen eine Abweichung von der Gleichverteilung. Satz 2.6.2 (ii) und (iii) verweist auf die Sonderstellung der Normalverteilung bei Dichten mit endlichen zweiten Momenten: In diesem Fall besitzt die Normalverteilung maximale Entropie und die Entropie ist proportional zur Varianz σ^2 im univariaten Fall bzw. proportional zur verallgemeinerten Varianz $\det \Omega$ im multivariaten Fall. Es ist allerdings darauf hinzuweisen, daß in Abhängigkeit von den gewählten Nebenbedingungen unterschiedliche Verteilungen maximale Entropie besitzen können, d.h. es existieren unterschiedliche maxentropie Verteilungen.

Bevor nun das Entropiekonzept auf stochastische Prozesse übertragen wird, ist es nützlich, eine wesentliche Modifikation des SHANNON'schen Informationsbegriffes durch KULLBACK/LEIBLER[1951] vorzustellen:

Def. 2.6.2 (*Kullback-Leibler-Informationskriterium*)

Sei X eine Zufallsvariable mit der Dichtefunktion $f(x)$ und sei $g(x)$ eine approximierende Dichte für $f(x)$. Dann heißt

$$\text{KLIC}(f, g) = \int_{-\infty}^{\infty} f(x) \ln \frac{f(x)}{g(x)} dx$$

Kullback-Leibler-Informationskriterium der Dichte $f(x)$ bezüglich der Dichte $g(x)$. ■

Die Definition des Kullback-Leibler-Informationskriteriums kann wie folgt motiviert werden: Betrachtet wird eine Zufallsvariable X , welche die Dichte $f(x)$ oder die Dichte $g(x)$ besitzen kann. Diese Annahmen lassen sich mit Hilfe der beiden Hypothesen

$$H_1 : X \sim f(x) \quad H_2 : X \sim g(x)$$

darstellen. Der Loglikelihoodquotient

$$\ln \frac{f(x)}{g(x)}$$

läßt sich nun als die durch die Beobachtung x erhaltene Information interpretieren, die zugunsten der Hypothese H_1 und gegen die Hypothese H_2 spricht. Im KULLBACK/LEIBER'schen Sinn ist Information also empirische Evidenz, die es erlaubt, zwischen alternativen Hypothesen zu diskriminieren. Das Kullback-Leibler-Informationskriterium $\text{KLIC}(f, g)$ kann nun als die mittlere diskriminierende Information zugunsten

von H_1 und gegen die Hypothese H_2 interpretiert werden, die man erhält, wenn die datenerzeugende Zufallsvariable X die Dichte $f(x)$ besitzt und als approximierende Dichte $g(x)$ unterstellt wird. Damit steht das Kullback-Leibler-Informationskriterium in engem Zusammenhang mit der Gütefunktion eines Testes der Hypothesen H_1 vs. H_2 . KULLBACK/LEIBLER[1951] bzw. KULLBACK[1959] bezeichnen ihr Informationskriterium als gerichtete Informationsdivergenz.

Das Kullback-Leibler-Informationskriterium läßt sich umformen zu:

$$(2.72) \quad \text{KLIC}(f, g) = \int_{-\infty}^{\infty} f(x) \ln f(x) dx - \int_{-\infty}^{\infty} f(x) \ln g(x) dx.$$

Definiert man den zweiten Term der rechten Seite in Gleichung (2.72) als Kreuzentropie der Dichte $f(x)$ bezüglich der Dichte $g(x)$, d.h.

$$S(f, g) = - \int_{-\infty}^{\infty} f(x) \ln g(x) dx,$$

dann ergibt sich für das Kullback-Leibler-Informationskriterium die Darstellung:

$$\text{KLIC}(f, g) = S(f, g) - S(f, f).$$

Das Kullback-Leibler-Informationskriterium erfüllt die folgende wichtige Eigenschaft (vgl. z.B. Kullback[1959, S. 14]):

Satz 2.6.3 (*Informations-Ungleichung*)

Für das Kullback-Leibler-Informationskriterium gilt:

$$\text{KLIC}(f, g) \geq 0,$$

mit Gleichheit dann und nur dann, wenn $f(x) = g(x)$.

■

Damit eignet sich das Kullback-Leibler-Informationskriterium als Abstands- bzw. Diskrepanzmaß für Dichtefunktionen.

Aus Satz 2.6.3 folgt unmittelbar:

$$S(f, f) \leq S(f, g).$$

Neben dem Kullback-Leibler-Informationskriterium ist noch auf das Rényi-Informationskriterium hinzuweisen (vgl. z.B. SUGIMOTO/WADA[1988]):

Def. 2.6.3 (*Rényi-Informationskriterium*)

Sei X eine Zufallsvariable mit der Dichtefunktion $f(x)$ und sei $g(x)$ eine approximierende Dichte für $f(x)$. Dann heißt

$$\text{RIC}_\alpha(f, g) = \frac{1}{\alpha - 1} \ln \int_{-\infty}^{\infty} f^\alpha(x) g^{(1-\alpha)}(x) dx$$

Rényi-Informationskriterium der Dichte $f(x)$ bezüglich der Dichte $g(x)$. Für den reellen Parameter α soll gelten: $\alpha > 0$ und $\alpha \neq 1$. ■

Es kann gezeigt werden, daß gilt (vgl. z.B. SUGIMOTO/WADA[1988]):

$$\lim_{\alpha \rightarrow 1} \text{RIC}_\alpha(f, g) = \text{KLIC}(f, g),$$

d.h. das Kullback-Leibler-Informationskriterium ergibt sich als Spezialfall des Rényi-Informationskriteriums.

PARZEN[1992] gibt eine Redefinition des Rényi-Informationskriteriums und beschreibt den Zusammenhang mit anderen Informationskriterien.

Im weiteren wird als Informationskriterium nur noch das Kullback-Leibler-Informationskriterium betrachtet. Es muß jedoch darauf hingewiesen werden daß, in Abhängigkeit von der jeweiligen Problemstellung, andere Informationskriterien vorteilhafter sein können (vgl. z.B. KAILATH[1967]).

Die besondere Bedeutung des Kullback-Leibler-Informationskriteriums ergibt sich durch seine unmittelbare Beziehung zum Konzept der Maximum-Likelihoodschätzung:

Wird durch $G_\theta = \{g_\theta(x), \theta \in \Theta\}$ eine Familie approximierender Dichten als parametrisches Modell für $f(x)$ spezifiziert, dann ist es sinnvoll, für die am besten approximierende Dichte $g(x, \tilde{\theta}) = g_{\tilde{\theta}}$ den Schätzer $\tilde{\theta}$ so festzulegen, daß die Bedingung

$$(2.73) \quad \tilde{\theta} = \arg \min_{\theta} \text{KLIC}(f, g_\theta) = S(f, g_\theta) - S(f, f)$$

erfüllt wird. Da der Term $S(f, f)$ für die Wahl eines optimalen Parameters θ vernachlässigt werden kann, ist (2.73) äquivalent zu:

$$(2.74) \quad \tilde{\theta} = \arg \min_{\theta} S(f, g_\theta).$$

Nach AKAIKE[1974] konvergiert die mittlere Loglikelihoodfunktion

$$(2.75) \quad 1/N \sum_{k=1}^N \ln g(x_k, \theta)$$

von unabhängigen Realisationen x_1, \dots, x_N mit Wahrscheinlichkeit 1 für $N \rightarrow \infty$ gegen die negative Kreuzentropie

$$\int_{-\infty}^{\infty} f(x) \ln g(x) dx.$$

Daraus ergibt sich, daß durch Anwendung des Maximum-Likelihood-Prinzips

$$(2.76) \quad \max_{\theta} \sum_{k=1}^N \ln g(x_k, \theta)$$

für hinreichend große N ein Parameter $\tilde{\theta}$ bestimmt werden kann, der in guter Näherung auch die Kreuzentropie $S(f, g_{\theta})$ und damit das Kullback-Leibler-Informationskriterium $\text{KLIC}(f, g_{\theta})$ minimiert. WHITE[1982] analysiert die Eigenschaften der (Quasi-)Maximum-Likelihood-Schätzung für den Fall der Modellfehlspezifikation, d.h. für den Fall daß die Dichte $f(x)$ nicht in der approximierenden Familie G_{θ} enthalten ist. Er gibt genauere Bedingungen an, unter denen die (Quasi-)Maximum-Likelihood-Schätzung konsistente Schätzungen der Parameter liefert, die das Kullback-Leibler-Informationskriterium minimieren. Es ist bei lediglich asymptotisch gültigen Aussagen jedoch zu beachten, daß die Güte der (Quasi-)Maximumlikelihood-Schätzung bei kleineren Stichprobenumfängen in Abhängigkeit von der konkreten Problemstellung stark variieren kann.

Die Verteilung eines schwach stationären Gauss-Prozesses $\{X_t, t \in \mathcal{Z}\}$ mit $E(X_t) = \mu = 0$ ist durch seine Spektraldichte vollständig spezifiziert. Dadurch ist es möglich, Entropie, Kreuzentropie sowie das Kullback-Leibler-Informationskriterium stochastischer Prozesse in Abhängigkeit von ihrer *Spektraldichte* $f(\omega)$ bzw. einer approximierenden *Spektraldichte* $g(\omega)$ anzugeben (vgl. PARZEN[1982, 1983]). Wie in der Literatur üblich, wird das gleiche Symbol f bzw. g kontextabhängig für eine Dichte- bzw. Spektraldichtefunktion verwandt. Grundlegend für die Darstellung im Frequenzbereich ist das Theorem von Szegö (vgl. z.B. NEWTON[1988, S. 143]):

Satz 2.6.4 (*Theorem von Szegö*)

Sei $\{X_t\}$ ein schwach stationärer Prozeß mit der Spektraldichte $f(\omega)$. Es soll gelten:

$$m = \inf_{\omega} f(\omega) > 0 \quad \text{und} \quad M = \sup_{\omega} f(\omega) < \infty.$$

Die in nicht-zunehmender Reihenfolge geordneten Eigenwerte der Autokovarianzmatrix Ω_N einer Stichprobe (X_1, \dots, X_N) seien

$$\lambda_{N,1} \geq \lambda_{N,2} \geq \dots \geq \lambda_{N,N}.$$

Weiter seien

$$f_{(1)} \geq f_{(2)} \geq \dots \geq f_{(N)}$$

die in nicht-zunehmender Reihenfolge geordneten Werte der Spektraldichte $f(\omega_j)$ an den Frequenzen $\omega_j = (j-1)/N$, $(j = 1, \dots, N)$. h sei eine stetige Funktion auf dem Intervall $[m, M]$. Dann gilt:

$$(2.77) \quad \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=1}^N [h(\lambda_{N,k}) - h(f_{(k)})] = 0$$

und

$$(2.78) \quad \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=1}^N h(\lambda_{N,k}) = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=1}^N h(f_{(k)}) = \int_{-0.5}^{0.5} h(f(\omega)) d\omega.$$

■

Aus Satz 2.6.4 folgt mit der identischen Abbildung $h(x) = x$, daß für einen hinreichend großen Stichprobenumfang N die geordneten Eigenwerte der Autokovarianzmatrix Ω_N annähernd gleich den geordneten Werten der Spektraldichte $f(\omega_j)$ sind. Darüberhinaus gibt es einen asymptotischen Zusammenhang zwischen dem Integral einer stetigen Funktion der Spektraldichte $h(f(\omega))$ und den entsprechend transformierten Eigenwerten von Ω_N .

Satz 2.6.4 bildet die allgemeine theoretische Grundlage des folgenden Satzes:

Satz 2.6.5 (*Entropie und Information stochastischer Prozesse*)

Sei $\{X_t, t \in \mathcal{Z}\}$ ein schwach stationärer Gauss-Prozeß mit $E(X_t) = \mu = 0$ und der Spektraldichte $f(\omega)$. Es sei $g(\omega)$ eine approximierende Spektraldichte. Dann gilt:

(i) (*Entropie*)

$$(2.79) \quad S(f, f) = \frac{1}{2} \int_{-0.5}^{0.5} [\ln f(\omega) + 1 + \ln 2\pi] d\omega,$$

(ii) (*Kreuzentropie*)

$$(2.80) \quad S(f, g) = \frac{1}{2} \int_{-0.5}^{0.5} \left[\ln g(\omega) + \frac{f(\omega)}{g(\omega)} + \ln 2\pi \right] d\omega,$$

(iii) (Kullback-Leibler-Informationskriterium)

$$(2.81) \quad \text{KLIC}(f, g) = \frac{1}{2} \int_{-0.5}^{0.5} \left[\frac{f(\omega)}{g(\omega)} - \ln \frac{f(\omega)}{g(\omega)} - 1 \right] d\omega.$$

Beweiskizze für (iii):

Für eine endliche Stichprobe (X_1, \dots, X_N) des Gauss-Prozesses $\{X_t\}$ mit der Dichte $f(x_1, \dots, x_N)$ sowie einer approximierenden Dichte $g(x_1, \dots, x_N)$, gilt (vgl. z.B. KULLBACK[1959, S. 189]):

(2.82)

$$\begin{aligned} \text{KLIC}(f(x_1, \dots, x_N), g(x_1, \dots, x_N)) &= \frac{1}{2} \left[\ln \frac{\det \mathbf{\Omega}_g}{\det \mathbf{\Omega}_f} + \text{sp } \mathbf{\Omega}_f (\mathbf{\Omega}_g^{-1} - \mathbf{\Omega}_f^{-1}) \right] \\ &= \frac{1}{2} \left[\ln \frac{\det \mathbf{\Omega}_g}{\det \mathbf{\Omega}_f} + \text{sp}(\mathbf{\Omega}_f \mathbf{\Omega}_g^{-1}) - N \right], \end{aligned}$$

wobei mit $\mathbf{\Omega}_f$ bzw. $\mathbf{\Omega}_g$ die Autokovarianzmatrizen der Dichten $f(x_1, \dots, x_N)$ bzw. $g(x_1, \dots, x_N)$ bezeichnet werden.

Die Darstellung im Frequenzbereich ergibt sich im wesentlichen mit Hilfe des Theorems von Szegö (vgl. Satz 2.6.4) durch den Ansatz:

$$(2.83) \quad \text{KLIC}(f(\omega), g(\omega)) = \lim_{N \rightarrow \infty} \frac{1}{N} \text{KLIC}(f(x_1, \dots, x_N), g(x_1, \dots, x_N)).$$

Die Autokovarianzmatrizen lassen sich mit Hilfe ihrer Spektralzerlegung darstellen als:

$$\mathbf{\Omega}_f = \mathbf{\Gamma}_f \mathbf{\Lambda}_f \mathbf{\Gamma}_f' \quad \mathbf{\Omega}_g = \mathbf{\Gamma}_g \mathbf{\Lambda}_g \mathbf{\Gamma}_g',$$

mit den Diagonalmatrizen der Eigenwerte

$$\mathbf{\Lambda}_f = \text{diag}(\lambda_{f,1}, \dots, \lambda_{f,N}) \quad \mathbf{\Lambda}_g = \text{diag}(\lambda_{g,1}, \dots, \lambda_{g,N}),$$

sowie den Orthogonalmatrizen $\mathbf{\Gamma}_g$ bzw. $\mathbf{\Gamma}_f$ der jeweils zugehörigen Eigenvektoren.

Nach dem Theorem von Szegö gilt:

$$(2.84) \quad \lim_{N \rightarrow \infty} \frac{1}{N} \ln \frac{\det \mathbf{\Omega}_g}{\det \mathbf{\Omega}_f} = \lim_{N \rightarrow \infty} \frac{1}{N} \left(\sum_{k=1}^N \ln \lambda_{g,k} - \sum_{k=1}^N \ln \lambda_{f,k} \right) = \int_{-0.5}^{0.5} \ln \frac{g(\omega)}{f(\omega)} d\omega.$$

Aus den Spektralzerlegungen der Autokovarianzmatrizen ergibt sich:

$$\text{sp } \mathbf{\Omega}_f \mathbf{\Omega}_g^{-1} = \text{sp } \mathbf{\Gamma}_f \mathbf{\Lambda}_f \mathbf{\Gamma}_f' \mathbf{\Gamma}_g \mathbf{\Lambda}_g^{-1} \mathbf{\Gamma}_g' = \text{sp } \mathbf{\Gamma}_g' \mathbf{\Gamma}_f \mathbf{\Lambda}_f \mathbf{\Gamma}_f' \mathbf{\Gamma}_g \mathbf{\Lambda}_g^{-1}.$$

Asymptotisch besitzen alle Toeplitzmatrizen gleiche Eigenvektoren (vgl. z.B. BROCKWELL/DAVIS[1991, S. 136]), so daß für einen hinreichend großen Stichprobenumfang N gilt:

$$\mathbf{\Gamma}'_g \mathbf{\Gamma}_f \approx \mathbf{\Gamma}'_f \mathbf{\Gamma}_g \approx \mathbf{I}_N.$$

\mathbf{I}_N ist die Einheitsmatrix N -ter Ordnung.

Damit läßt sich analog wie in (2.84) zeigen, daß gilt:

(2.85)

$$\lim_{N \rightarrow \infty} \frac{1}{N} \text{sp } \mathbf{\Omega}_f \mathbf{\Omega}_g^{-1} = \lim_{N \rightarrow \infty} \frac{1}{N} \text{sp } \mathbf{\Lambda}_f \mathbf{\Lambda}_g^{-1} = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=1}^N \frac{\lambda_{f,k}}{\lambda_{g,k}} = \int_{-0.5}^{0.5} \frac{f(\omega)}{g(\omega)} d\omega.$$

Mit (2.84) und (2.85) erhält man über den Ansatz (2.83) die Gleichung (2.81) .

■

Da die Entropie $S(f, f)$ bzw. die Kreuzentropie $S(f, g)$ häufig nur im Kontext von Maximierungs- bzw. Minimierungsproblemen betrachtet werden, ist es einfacher, die in diesen Zusammenhang nicht relevanten Terme wegzulassen und die folgenden vereinfachten Definitionen zu verwenden:

$$\acute{S}(f, f) = \int_{-0.5}^{0.5} \ln f(\omega) d\omega,$$

$$\acute{S}(f, g) = \int_{-0.5}^{0.5} \left[\ln g(\omega) + \frac{f(\omega)}{g(\omega)} \right] d\omega.$$

Es ist darauf hinzuweisen, daß im Zusammenhang mit den sogenannten Maximum-Entropie-Schätzern auch andere Definitionen der Entropie sinnvoll sein können (vgl. Abschnitt 3.3).

Aus der Ungleichung $u - \ln u - 1 \geq 0$ folgt, daß das Kullback-Leibler-Informationskriterium auch in der Form (2.81) für zwei Spektraldichten $f(\omega)$ und $g(\omega)$ die wichtige Bedingung

$$\text{KLIC}(f, g) \geq 0$$

erfüllt. Die Gleichheit gilt dann und nur dann, wenn $f(\omega) = g(\omega)$ ist (vgl. z.B. PARZEN[1983]).

Die Zusammenhänge zwischen dem bekanntesten statistischen Informationsbegriff, der Fisher-Information bzw. der Fisher'schen Informationsmatrix, und dem Kullback-Leibler-Informationskriterium sind z.B. in KULLBACK[1959, S. 26ff] beschrieben.

Eine allgemeinere Darstellung geben SUGIMOTO/WADA[1988]; Sie diskutieren u.a. die Spektraldarstellung von verschiedenen Informationskriterien für schwach stationäre Gauss'sche Prozesse sowie deren approximativen Zusammenhang mit einer asymptotischen Form der Fisher'schen Informationsmatrix, die in Analogie zu dem Ansatz (2.83) hergeleitet wird.

Kapitel 3

Autoregressive Spektralschätzung

Autoregressive Spektralschätzung war bereits 1927 von YULE zur Modellierung des Sonnenfleckenzyklus verwandt worden. Die Motivation, autoregressive Beziehungen zur Modellierung periodischer Vorgänge einzusetzen, ergab sich aus der trigonometrischen Identität (vgl. z.B. Marple[1987, S. 8ff])

$$(3.1) \quad \sin(tx) = 2 \cos(x) \sin((t-1)x) - \sin((t-2)x),$$

woraus mit $\alpha_1 = 2 \cos(x)$ und $x = 2\pi\omega$ sowie der Berücksichtigung von Zufallsstörungen der spezielle diskrete autoregressive Ansatz,

$$(3.2) \quad X_t = \alpha_1 X_{t-1} - X_{t-2} + \epsilon_t,$$

sowie dessen unmittelbare Erweiterung,

$$(3.3) \quad X_t = \alpha_1 X_{t-1} - \alpha_2 X_{t-2} + \epsilon_t,$$

abgeleitet werden konnte. Die Parameterschätzung für die Modelle nach Gleichung (3.2) bzw. (3.3) erfolgte nach dem Prinzip der kleinsten Quadrate.

Inzwischen existieren auf der Grundlage der Theorie stochastischer Prozesse zahlreiche verschiedene Verfahren der autoregressiven Spektralschätzung.

Für die Verbreitung der autoregressiven Spektralschätzung im ingenieurwissenschaftlichen Bereich spielten die Arbeiten von BURG[1967, 1968] eine besondere Rolle. Der Beitrag von Burg umfaßte zwei wesentliche Innovationen: Zum einen verknüpfte er das Konzept der autoregressiven Spektralschätzung mit dem Konzept der maximalen

Entropie. Damit lieferte er eine neue theoretische Sichtweise der Spektralschätzung auf der Basis von AR-Modellen. Zum andern stellte BURG[1968] auf der Anwendungsebene einen neuen rekursiven Algorithmus zur Schätzung der AR-Parameter vor, der auch bei sehr kleinen Stichprobenumfängen hochauflösende Spektralschätzungen ermöglichte.

Die Verbreitung der Anwendung von autoregressiven Spektralschätzern wurde weiterhin durch die Arbeiten von AKAIKE[1969a, 1969b, 1974] unterstützt, der nahezu zeitgleich mit den Veröffentlichungen von BURG Kriterien vorschlug, mit denen es möglich war, die Ordnung eines $AR(p)$ -Prozesses anhand der Daten zu bestimmen. Im Gegensatz zu den nichtparametrischen Alternativen besaßen autoregressive Spektralschätzer damit den Vorteil, sich bei der Modellierung auf objektive, datenbezogene Kriterien stützen zu können.

Die Anwendung von autoregressiven Spektralschätzern auf die Schätzung von harmonischen Prozessen in Noise warf allerdings auch neuartige Probleme auf, die mit den Termini „Peak-Shifting“ und „Spectral-Line-Splitting“ bezeichnet werden:¹

- Peak-Shifting

Hierunter versteht man einen Bias, der sich aus der Differenz zwischen den tatsächlichen Frequenzen eines harmonischen Prozesses in Noise und den Frequenzen, an denen bei autoregressiven Spektralschätzern Peaks auftreten, ergibt. Es kann gezeigt werden, daß dieser Bias für verschiedene autoregressive Spektralschätzer in unterschiedlichem Ausmaß phasenabhängig ist (vgl. z.B. CHEN/STEGEN[1974], ULRICH/CLAYTON[1976]).

- Spectral-Line-Splitting

Ein einziger tatsächlicher Peak des datenerzeugenden harmonischen Prozesses erscheint zwei oder mehrfach (vgl. z.B. FOUGERE ET AL.[1976], KAY/MARPLE[1979], HERRING[1980], FOUGERE[1985]).

Dies führte zur Entwicklung von weiteren modifizierten autoregressiven Schätzverfahren (vgl. z.B. ULRICH/CLAYTON[1976], SWINGLER[1979]); einen Erklärungsansatz für diese Phänomene lieferte u.a. HURVICH[1986], indem er für autoregressive Spektralschätzer datenabhängige Spektralfenster ableitet, die Sidelobes wie klassische Spektralschätzer aufweisen können.

¹Beispiele hierzu werden im Abschnitt 3.4.5 vorgestellt.

In den folgenden Abschnitten werden zunächst allgemeine Eigenschaften von autoregressiven Modellen und ihre Bedeutung für die Spektralschätzung behandelt. Anschließend werden die vier gängigsten Varianten der autoregressiven Spektralschätzung, Yule-Walker-, Burg-, FBLP(Forward-Backward-Linear-Prediction)- und FLP(Forward-Linear-Prediction)-Schätzer dargestellt. Anhand von einfachen Simulationsbeispielen erfolgt dann abschließend ein Vergleich der vorgestellten autoregressiven Schätzverfahren.

3.1 Autoregressive Prozesse und lineare Prädiktion

Die Interdependenz zwischen kausalen und antikausalen AR-Prozessen sowie Methoden der optimalen Vorwärts- und Rückwärtsprädiktion bildet zusammen mit rekursiven Algorithmen, die weitere theoretische Einblicke ermöglichen, den Gestaltungshintergrund für die Entwicklung von zahlreichen Verfahren der autoregressiven Spektralschätzung (vgl. z.B. MARPLE[1987, S. 191ff], THERRIEN[1992, S. 409ff]).

Im Falle einer Einschnitt-Vorwärtsprognose sollen die Koeffizienten eines kausalen linearen Prognosefilters $\phi_p^f = (\phi_1^f, \dots, \phi_p^f)'$ p -ter Ordnung,

$$(3.4) \quad \hat{X}_t^f = -\phi_1^f X_{t-1} - \dots - \phi_p^f X_{t-p},$$

so bestimmt werden, daß der mittlere quadratische Vorwärtsprognosefehler

$$(3.5) \quad P^f = E(|X_t - \hat{X}_t^f|^2) = E(|\epsilon_t^f|^2)$$

minimiert wird.

Gelegentlich soll die Ordnung p eines Prognosefilters besonders hervorgehoben werden, so daß z.B. geschrieben wird: $\hat{X}_{p,t}^f$ oder auch $E(|\epsilon_{p,t}^f|^2)$.

Aus Gleichung (3.5) folgt:

$$\begin{aligned}
P^f &= E(X_t - \hat{X}_t^f)(X_t^* - \hat{X}_t^{f*}) \\
&= E(X_t + \sum_{j=1}^p \phi_j^f X_{t-j})(X_t^* + \sum_{k=1}^p \phi_k^{f*} X_{t-k}^*) \\
&= E(X_t X_t^* + \sum_{k=1}^p \phi_k^{f*} X_t X_{t-k}^* + \sum_{j=1}^p \phi_j^f X_{t-j} X_t^* + \sum_{j=1}^p \sum_{k=1}^p \phi_j^f \phi_k^{f*} X_{t-j} X_{t-k}^*) \\
(3.6) \quad &= \gamma_0 + (\phi_p^f)^H \gamma_p + \gamma_p^H \phi_p^f + (\phi_p^f)^H \Omega_{p-1} \phi_p^f
\end{aligned}$$

mit:

$$\gamma_p = \begin{bmatrix} \gamma_1 \\ \vdots \\ \gamma_{p-1} \\ \gamma_p \end{bmatrix} \quad \Omega_{p-1} = \begin{bmatrix} \gamma_0 & \gamma_1^* & \cdots & \gamma_{p-1}^* \\ \gamma_1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \gamma_1^* \\ \gamma_{p-1} & \cdots & \gamma_1 & \gamma_0 \end{bmatrix}.$$

Die Lösung des Minimierungsansatzes

$$\min_{\phi_p^f} P^f = E(|\epsilon_t^f|^2)$$

wird im folgenden Satz unter Verwendung der erweiterten Normalgleichungen zusammengefaßt (vgl. z.B. Marple[1987, S. 191]):

Satz 3.1.1 (*Optimale Vorwärtsprognose*)

Sei $\{X_t, t \in \mathcal{Z}\}$ ein schwach stationärer Prozeß. Dann sind die Koeffizienten $\phi_p^f = (\phi_1^f, \dots, \phi_p^f)'$ des Vorwärtsprognosefilters, der den mittleren quadratischen Prognosefehler P^f minimiert, Lösung des Gleichungssystems:

$$(3.7) \quad \begin{bmatrix} \gamma_0 & \gamma_1^* & \cdots & \gamma_p^* \\ \gamma_1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \gamma_1^* \\ \gamma_p & \cdots & \gamma_1 & \gamma_0 \end{bmatrix} \begin{bmatrix} 1 \\ \phi_1^f \\ \vdots \\ \phi_p^f \end{bmatrix} = \begin{bmatrix} P^f \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

bzw. in Kurzform:

$$(3.8) \quad \begin{bmatrix} \gamma_0 & \gamma_p^H \\ \gamma_p & \Omega_{p-1} \end{bmatrix} \begin{bmatrix} 1 \\ \phi_p^f \end{bmatrix} = \begin{bmatrix} P^f \\ \mathbf{0}_p \end{bmatrix}$$

■

Für die Einschnitt-Rückwärtsprognose

$$(3.9) \quad \hat{X}_t^b = -\phi_1^b X_{t+1} - \dots - \phi_p^b X_{t+p},$$

können entsprechend die Koeffizienten des antikausalen Prognosefilters $\phi_p^b = (\phi_1^b, \dots, \phi_p^b)'$ so bestimmt werden, daß der mittlere quadratische Rückwärtsprognosefehler

$$(3.10) \quad E(|X_t - \hat{X}_t^b|^2)$$

minimiert wird. Es ist allerdings einfacher, für den Vorwärts- und Rückwärtsprognosefilter die gleiche Menge $\{X_t, \dots, X_{t-p}\}$ zugrunde zu legen, so daß der Rückwärtsprognosefehler ϵ_t^b unter leichtem Mißbrauch der Notation üblicherweise definiert wird als:

$$(3.11) \quad \epsilon_t^b = X_{t-p} - \hat{X}_{t-p}^b$$

$$(3.12) \quad = X_{t-p} + \phi_1^b X_{t-p+1} + \dots + \phi_p^b X_t.$$

Für den mittleren quadratischen Rückwärtsprognosefehler P^b ergibt sich mit $s = t - p$:

$$\begin{aligned} P^b &= E(|\epsilon_t^b|^2) \\ &= E(X_s - \hat{X}_s^b)(X_s^* - \hat{X}_s^{b*}) \\ &= E(X_s + \sum_{j=1}^p \phi_j^b X_{s+j})(X_s^* + \sum_{k=1}^p \phi_k^{b*} X_{s+k}^*) \\ &= E(X_s X_s^* + \sum_{k=1}^p \phi_k^{b*} X_s X_{s+k}^* + \sum_{j=1}^p \phi_j^b X_{s+j} X_s^* + \sum_{j=1}^p \sum_{k=1}^p \phi_j^b \phi_k^{b*} X_{s+j} X_{s+k}^*) \\ (3.13) \quad &= \gamma_0 + (\phi_p^b)^H \gamma_p^* + \gamma_p' \phi_p^b + (\phi_p^b)^H \Omega_{p-1}^* \phi_p^b, \end{aligned}$$

wobei γ_p und Ω_{p-1} wie in (3.6) definiert sind.

Die Lösung für das Minimierungsproblem:

$$\min_{\phi_p^b} P^b = E(|\epsilon_t^b|^2)$$

ergibt sich als (vgl. z.B. Marple[1987, S. 193]):

Satz 3.1.2 (*Optimale Rückwärtsprognose*)

Sei $\{X_t, t \in \mathcal{Z}\}$ ein schwach stationärer Prozeß. Dann sind die Koeffizienten $\phi_p^b = (\phi_1^b, \dots, \phi_p^b)'$ des Rückwärtsprognosefilters, der den mittleren quadratischen Prognosefehler P^b minimiert, Lösung des Gleichungssystems:

$$(3.14) \quad \begin{bmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_p \\ \gamma_1^* & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \gamma_1 \\ \gamma_p^* & \cdots & \gamma_1^* & \gamma_0 \end{bmatrix} \begin{bmatrix} 1 \\ \phi_1^b \\ \vdots \\ \phi_p^b \end{bmatrix} = \begin{bmatrix} P^b \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

bzw. in Kurzform:

$$(3.15) \quad \begin{bmatrix} \gamma_0 & \gamma_p' \\ \gamma_p^* & \Omega_{p-1}^* \end{bmatrix} \begin{bmatrix} 1 \\ \phi_p^b \end{bmatrix} = \begin{bmatrix} P^b \\ \mathbf{0}_p \end{bmatrix}.$$

■

Ordnet man Gleichung (3.14) um:

$$(3.16) \quad \begin{bmatrix} \gamma_0 & \gamma_1^* & \cdots & \gamma_p^* \\ \gamma_1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \gamma_1^* \\ \gamma_p & \cdots & \gamma_1 & \gamma_0 \end{bmatrix} \begin{bmatrix} \phi_p^b \\ \vdots \\ \phi_1^b \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ P^b \end{bmatrix},$$

so zeigt sich, daß sowohl in den Vorwärtsgleichungen (3.7) als auch in den Rückwärtsgleichungen (3.14) die gleiche zu invertierende hermitesche Toeplitzmatrix auftritt.

Aus den Eigenschaften von (inversen) hermiteschen Toeplitzmatrizen (vgl. Anhang A, (A.4), (A.5) und (A.29)) ergibt sich:

Satz 3.1.3 (*Eigenschaften der optimalen Vorwärts- und Rückwärtsprognose*)

Seien $\phi_p^f = (\phi_1^f, \dots, \phi_p^f)'$ und $\phi_p^b = (\phi_1^b, \dots, \phi_p^b)'$ die Koeffizienten des Vorwärts- bzw. des Rückwärtsprognosefilters, die den mittleren quadratischen Prognosefehler P^f bzw. P^b eines schwach stationären stochastischen Prozesses $\{X_t, t \in \mathcal{Z}\}$ minimieren. Dann gilt:

$$(3.17) \quad \phi_p^f = \phi_p^{b*}$$

und

$$(3.18) \quad P^f = P^b = P$$

■

Die Koeffizienten des Vorwärtsprognosefilters ϕ_p^f und die Koeffizienten des Rückwärtsprognosefilters ϕ_p^b sind zueinander konjugiert komplex und die mittleren quadratischen Prognosefehler P^f und P^b sind identisch. Damit kann zur Bestimmung der optimalen Filterkoeffizienten optional von der Betrachtung des kausalen oder des antikausalen Systems ausgegangen werden. Dies bildet u.a. den entscheidenden Ausgangspunkt für die Entwicklung von verschiedenen hochauflösenden Verfahren der autoregressiven Spektralschätzung, wie z.B. dem Burg- oder dem FBLP-Schätzer.

Handelt es sich bei dem zugrundeliegenden stochastischen Prozeß um einen kausalen AR(p)-Prozeß

$$X_t + \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} = \epsilon_t,$$

dann ist der Output des Filters $(1, \phi_1^f, \dots, \phi_p^f)'$ ein White-Noise-Prozeß, dessen Varianz σ_ϵ^2 mit dem mittleren quadratischen Prognosefehler P^f übereinstimmt. Das Gleichungssystem (3.7) entspricht in diesem Fall den erweiterten Yule-Walker-Gleichungen. Diese Aussagen gelten analog für den entsprechenden antikausalen AR(p)-Prozeß

$$X_t + \phi_1^* X_{t+1} + \dots + \phi_p^* X_{t+p} = \epsilon_t.$$

Zur Vereinfachung der Notation wird im weiteren nur noch da, wo es erforderlich ist, zwischen den Koeffizienten ϕ_p eines kausalen AR-Prozesses und den Koeffizienten des optimalen Vorwärtsprognosefilters ϕ_p^f unterschieden, entsprechend werden auch σ_ϵ^2 und $P^f = P^b = P$ als äquivalent betrachtet.

3.2 Levinson-Durbin-Rekursion

Das Gleichungssystem des optimalen Vorwärtsprognosefilters

$$(3.19) \quad \begin{bmatrix} \gamma_0 & \boldsymbol{\gamma}_p^H \\ \boldsymbol{\gamma}_p & \boldsymbol{\Omega}_{p-1} \end{bmatrix} \begin{bmatrix} 1 \\ \boldsymbol{\phi}_p \end{bmatrix} = \begin{bmatrix} P \\ \mathbf{0}_p \end{bmatrix}$$

ist ein Spezialfall des im Anhang A beschriebenen Toeplitz-Gleichungssystems, so daß die Parameter ϕ_1, \dots, ϕ_p und P rekursiv berechnet werden können. Die Formulierung des Algorithmus erfolgt so, daß die Übertragung in eine matrizenorientierte Programmiersprache unmittelbar möglich ist. Die Notation wird dementsprechend geringfügig erweitert.

Satz 3.2.1 (*Levinson-Durbin-Rekursion*)

Sei $\phi_{j,k}$ der Koeffizient mit Index k im j -ten Rekursionschritt, $\boldsymbol{\phi}_j[1:j] = (\phi_{1,1}, \dots, \phi_{j,j})'$, $\boldsymbol{\gamma}[1:j] = (\gamma_1, \dots, \gamma_j)'$ und \mathbf{J} eine ordnungskonforme Reflektionsmatrix. P_j sei der mittlere quadratische Prognosefehler im j -ten Rekursionschritt. Dann läßt sich die rekursive Lösung des Gleichungssystems (3.19) angeben als:

Initialisierung:

$$(3.20) \quad \phi_{11} = -\gamma_1/\gamma_0, \quad P_1 = \gamma_0(1 - |\phi_{11}|^2)$$

für $j = 2, \dots, p$:

$$(3.21) \quad \phi_{jj} = \frac{\boldsymbol{\phi}_{j-1}[1:j-1]'\mathbf{J}\boldsymbol{\gamma}[1:j-1] + \gamma_j}{P_{j-1}}$$

$$(3.22) \quad \boldsymbol{\phi}_j[1:j-1] = \boldsymbol{\phi}_{j-1}[1:j-1] + \phi_{jj}\mathbf{J}\boldsymbol{\phi}_{j-1}^*[1:j-1]$$

$$(3.23) \quad P_j = P_{j-1}(1 - |\phi_{jj}|^2).$$

■

Wie grundsätzlich bei rekursiven Berechnungen (vgl. z.B. RICE[1983, S. 46ff]), ist auf numerische Stabilität zu achten. BOX/JENKINS[1976, S. 86] verweisen auf numerische Instabilität der rekursiven Vorgehensweise, falls Wurzeln des charakteristischen AR-Polynoms nahe dem Einheitskreis liegen. Nach PAGANO[1972] ist die Berechnung der Koeffizienten mit Hilfe einer Cholesky-Zerlegung numerisch stabiler.

Die Koeffizienten $\{\phi_{j,j}, j = 1, \dots, p\}$, die in der Levinson-Durbin-Rekursion auftreten, werden aufgrund ihrer besonderen Bedeutung mit einem eigenem Symbol bezeichnet:

$$\kappa_j = \phi_{jj} \quad j = 1, \dots, p.$$

In geophysikalischen oder akustischen Modellen (vgl. z.B. ROBINSON[1967, S. 117ff] bzw. OPPENHEIM[1978a, S. 152]) können die Koeffizienten κ_j physikalisch interpretiert werden, sie werden in diesen Fällen auch als Reflektionskoeffizienten bezeichnet.

Da für die Varianzen des Prognosefehlers in (3.23) gilt:

$$(3.24) \quad P_j \geq 0 \quad \text{und} \quad P_{j-1} \geq 0$$

folgt:

$$(3.25) \quad |\kappa_j| \leq 1,$$

d.h. die Reflektionskoeffizienten sind betragsmäßig durch Eins beschränkt. Darüberhinaus ergibt sich:

$$(3.26) \quad P_j \leq P_{j-1},$$

d.h. der mittlere quadratische Prognosefehler ist in Abhängigkeit von der Ordnung j des Prognosefilters eine monoton nicht-steigende Funktion.

Für eine statistische Interpretation der Reflektionskoeffizienten κ_j wird folgende Definition benötigt:

Def. 3.2.1 (*Partielle Autokorrelation*)

Seien X_t und X_{t-j} zwei Zufallsvariablen eines stochastischen Prozesses $\{X_t, t \in \mathcal{Z}\}$. Dann heißt der um den linearen Einfluß der Zufallsvariablen $X_{t-1}, \dots, X_{t-j+1}$ bereinigte Korrelationskoeffizient ($j \geq 2$):

$$(3.27) \quad \rho_{X_t X_{t-j} | X_{t-1}, \dots, X_{t-j+1}} = \text{Corr}(X_t - \hat{X}_{j-1,t}^f, X_{t-j} - \hat{X}_{j-1,t-j}^b) = \rho_{Par}(j)$$

partieller oder direkter Autokorrelationskoeffizient, wobei $\hat{X}_{j-1,t}^f$ und $\hat{X}_{j-1,t-j}^b$ die beste lineare Vorwärts- bzw. Rückwärtsprognose der Ordnung $j-1$ durch $\{X_{t-1}, \dots, X_{t-j+1}\}$ ist.

Für $j = 0, 1$ wird der partielle Autokorrelationskoeffizient festgelegt als:

$$\rho_{Par}(0) = 1 \quad \rho_{Par}(1) = \rho(1),$$

für $j < 0$ soll gelten:

$$\rho_{Par}(-j) = \rho_{Par}^*(j).$$

■

Da für einen AR(p)-Prozeß gilt:

$$\rho_{Par}(j) = 0 \quad \text{für } j > p,$$

kann die geschätzte partielle Autokorrelationsfunktion als (grobes) Hilfsmittel zur Bestimmung der Ordnung p eines AR-Prozesses verwandt werden.

Der Zusammenhang zwischen partiellen Autokorrelationskoeffizienten und den Reflektionskoeffizienten ergibt sich als (vgl. z.B. MARPLE[1987, S. 196]):

$$(3.28) \quad \kappa_j = -\rho_{X_t X_{t-j} | X_{t-1}, \dots, X_{t-j+1}},$$

bzw. in der expliziten Darstellung:

$$(3.29) \quad \begin{aligned} \kappa_j &= -\text{Corr}(X_t - \hat{X}_{j-1,t}^f, X_{t-j} - \hat{X}_{j-1,t-j}^b) \\ &= -\text{Corr}(\epsilon_{j-1,t}^f, \epsilon_{j-1,t-j}^b) \\ &= -\frac{E(\epsilon_{j-1,t}^f \epsilon_{j-1,t-j}^{b*})}{\sqrt{E(|\epsilon_{j-1,t}^f|^2)} \sqrt{E(|\epsilon_{j-1,t-j}^b|^2)}}. \end{aligned}$$

Die Darstellung (3.29) ist der Ausgangspunkt für verschiedene Verfahren der autoregressiven Spektralschätzung, wobei der Reflektionskoeffizient κ_j unmittelbar aus den Daten geschätzt wird. Die bekannteste Variante dieser Verfahren ist der Burg-Maximum-Entropie-Spektralschätzer.

Die verschiedenen Arten und Weisen wie AR-Prozesse alternativ parametrisch dargestellt werden können, sind mit den notwendigen Nebenbedingungen im folgenden Satz zusammengefaßt (vgl. z.B. MARPLE[1987, S. 197ff]):

Satz 3.2.2 (*Darstellung von AR-Prozessen*)

Die folgenden Darstellungen eines kausal-invertierbaren, stabilen AR(p)-Prozesses sind äquivalent:

1. $\{\gamma_0, \gamma_1, \dots, \gamma_p\}$, mit $(p+1)$ positiv-definiten Autokovarianzen
2. $\{\gamma_0, \phi_1, \dots, \phi_p\}$, wobei die Wurzel des charakteristischen Polynoms $1 + \sum_{j=1}^p \phi_j z^{-j}$ innerhalb des Einheitskreises liegen und $\gamma_0 > 0$
3. $\{\gamma_0, \kappa_1, \dots, \kappa_p\}$ mit $|\kappa_j| < 1$ und $\gamma_0 > 0$.

■

Mit Satz (3.2.2) ist es möglich, je nach Problemstellung, sich für eine geeignete Darstellungsform eines AR-Prozesses zu entscheiden.

3.3 AR-Prozesse und das Prinzip der maximalen Entropie

Das Prinzip der maximalen Entropie begründet einen allgemeinen Ansatz der statistischen Modellbildung und Datenanalyse, dessen Ursprung in der Physik, insbesondere der statistischen Mechanik und Thermodynamik, zu finden ist (vgl. Abschnitt 2.6). Seine moderne Formulierung ist wesentlich auf JAYNES[1957a,b] zurückzuführen.

Für die Ausführungen in diesem Abschnitt ist es sinnvoll, die informationstheoretische SHANNON'sche Definition der Entropie (vgl. Def. 2.6.1) als wichtigen Spezialfall eines Entropiebegriffes aufzufassen, bei dem unter Entropie ganz allgemein Unsicherheit bzw. Ungewißheit verstanden wird.

Das Prinzip der maximalen Entropie kann wie folgt zusammengefaßt werden (vgl. z.B. ABLES[1972]):

Wenn experimentellen Daten ein Modell zugeordnet werden soll, dann hat dieses Modell sowohl mit den gegebenen Daten übereinzustimmen als auch möglichst neutral und unverbindlich („maximally non-committal“) bezüglich der (noch) nicht verfügbaren Daten zu sein.

Für praktische Anwendungen läßt sich diese Forderung dadurch konkretisieren, indem eine Entropiefunktion aufgestellt wird, die jedem in Frage kommenden Modell einen skalaren Wert zuweist, der den Grad an Ungewissheit und Unsicherheit zum Ausdruck bringt, der mit diesem Modell verbunden ist. Ein Modell, das in hohem Maße neutral und unverbindlich ist, besitzt einen hohen Grad an Ungewißheit und Unsicherheit und umgekehrt. Ein konkretes Modell wird dadurch bestimmt, daß diese Entropiefunktion unter der Nebenbedingung maximiert wird, daß das Modell, gegebenenfalls unter Berücksichtigung von Meßfehlern, mit den gegebenen Daten übereinzustimmen hat. Die allgemeine Formulierung des Prinzips der maximalen Entropie läßt grundsätzlich offen, welche Entropiefunktion und welche Nebenbedingungen zu wählen sind.

Bei der Auswahl einer geeigneten Entropiefunktion bedient man sich meist der informationstheoretischen Interpretation der Entropie (vgl. Definition 2.6.1). BURG[1967] wandte das Prinzip der maximalen Entropie auf das Problem der Spektralschätzung an. Er lehnte die zu dieser Zeit weit verbreiteten indirekten Spektralschätzer ab, da die übliche Art der Berechnung der verzerrten Autokovarianzschätzer $\{\hat{\gamma}_\tau\}$ impliziert, daß die Daten außerhalb des durch die Stichprobe gegebenen Stützbereiches gleich Null gesetzt werden, was im groben Widerspruch zum Prinzip der maximalen Entropie steht: Die Extrapolation der unbekanntenen Daten mit Null stellt eine willkürliche Erweiterung der Daten dar, die nicht aus der zugrundeliegenden Stichprobe abgeleitet werden kann. Aus dem gleichen Grund lehnte er direkte Spektralschätzer ab, da die Anwendung der Fouriertransformation eine unendlich periodische Fortsetzung der Stichprobe impliziert.

Zur Ableitung von Schätzalternativen geht BURG[1967] zunächst von der folgenden, vereinfachten Fragestellung aus:

Von einem normalverteilten, stochastischen Prozeß seien die ersten $(p+1)$ Autokovarianzen $\{\gamma_0, \dots, \gamma_p\}$ bekannt. Wie sind die Autokovarianzen fortzuschreiben, damit die Entropie $S(f)$ des stochastischen Prozesses $\{X_t\}$ mit der Spektraldichte $f(\omega)$ und den Autokovarianzen $\{\gamma_0, \dots, \gamma_p, \gamma_{p+1}, \gamma_{p+2}, \dots\}$ maximal wird ?

Das Problem läßt sich mit Hilfe der Entropie $S(f)$ aus Satz 2.6.5 formal darstellen als:

$$(3.30) \quad \max_{\gamma_\tau, \tau > p} \int_{-0.5}^{0.5} \ln f(\omega) d\omega$$

unter den $(p + 1)$ Nebenbedingungen:

$$\gamma_\tau = \int_{-0.5}^{0.5} e^{i2\pi\tau\omega} f(\omega) d\omega, \quad \tau = 0, \dots, p.$$

Die für das Maximierungsproblem nicht relevanten Konstanten der Entropie $S(f)$ sind in (3.30) außer Acht gelassen worden.

Die Lösung wird im folgenden Satz angegeben:

Satz 3.3.1 (*Maximale-Entropie-Extrapolation*)

Eine Folge von positiv-definiten Autokovarianzen $\{\gamma_p^, \dots, \gamma_0, \dots, \gamma_p\}$ besitzt eine eindeutige Maximum-Entropie-Extrapolation, die durch die Autokovarianzen des autoregressiven stochastischen Prozesses p -ter Ordnung bestimmt wird, dessen Koeffizienten durch die erweiterten Yule-Walker-Gleichungen:*

$$(3.31) \quad \begin{bmatrix} \gamma_0 & \gamma_p^H \\ \gamma_p & \Omega_{p-1} \end{bmatrix} \begin{bmatrix} 1 \\ \phi_p \end{bmatrix} = \begin{bmatrix} P \\ \mathbf{0}_p \end{bmatrix}$$

bestimmt sind.

■

Verschiedene Beweise für Satz 3.3.1 befinden sich z.B. in VAN DEN BOS[1971], PRIESTLEY[1981, S. 604ff] oder PARZEN[1982,1983].

Die Aussage von Satz 3.3.1 kann leicht dadurch veranschaulicht werden, daß wenn bei gegebenen $(p + 1)$ Autokovarianzen ein AR-Prozeß p -ter Ordnung als erzeugender Prozeß unterstellt wird, eine Extrapolation der Autokovarianzen so erfolgt, daß der mittlere quadratische Einschnitt-Prognosefehler P konstant bleibt und nicht weiter abnimmt. Dies ist äquivalent dazu, daß die Entropie des zugeordneten Prozesses nicht über das Maß hinaus verringert wird, als es durch die gegebenen Autokovarianzen zu rechtfertigen ist. Die Maximum-Entropie-Extrapolation mit vorgegebenen $(p + 1)$ Autokovarianzen führt zu keinem anderen Ergebnis, wie das Problem des optimalen Vorwärtsprognosefilters p -ter Ordnung aus Satz 3.1.1. Der Zusammenhang zwischen der Entropie $S(f)$ eines schwach stationären stochastischen Prozesses $\{X_t\}$ mit der MA(∞)-Darstellung

$$X_t = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j} \quad \psi_0 = 1, \quad \epsilon_t \sim \text{WN}(0, \sigma_\epsilon^2)$$

und dem mittleren quadratischen Einschnitt-Prognosefehler P ergibt sich, ohne Berücksichtigung konstanter Terme, als (vgl. z.B. BROCKWELL/DAVIS[1991, S. 191]):

$$(3.32) \quad P = \sigma_\epsilon^2 = \exp\left[\int_{-0.5}^{0.5} \ln f(\omega) d\omega\right] = \exp[S(f)],$$

bzw.

$$(3.33) \quad \ln P = \ln \sigma_\epsilon^2 = S(f).$$

Gleichung (3.32) wird auch als Formel von Kolmogorov bezeichnet.

Bei der Interpretation von Satz 3.3.1 sind verschiedene Umstände zu berücksichtigen, die dessen praktische Bedeutung erheblich mindern:

Das Ergebnis von Satz 3.3.1 gilt nur dann, wenn die ersten $(p + 1)$ Autokovarianzen bekannt sind oder ohne Meßfehler ermittelt wurden. Das für praktische Anwendungen relevante Problem, ein geeignetes Schätzverfahren für die unbekanntenen Autokovarianzen $\{\gamma_\tau\}$ zu finden, wird durch Satz 3.3.1 in keiner Weise einer Lösung näher gebracht. Darüberhinaus macht Satz 3.3.1 keinerlei Angaben darüber, wie anhand einer Stichprobe $(X_1, \dots, X_N)'$ die Ordnung p des anzupassenden AR-Prozesses zu bestimmen ist. Wird die Ordnung p zu klein gewählt, bleiben relevante Merkmale des zu modellierenden Prozesses unberücksichtigt, wird die Ordnung p zu groß gewählt, ergeben sich Artefakte. PARZEN[1983] weist darauf hin, daß es sich bei der autoregressiven Maximum-Entropie-Methode um ein Konzept zur Ableitung eines stochastischen Modells und nicht um ein Schätzverfahren handelt. Satz 3.3.1 motiviert lediglich ganz allgemein die Verwendung autoregressiver Modelle, ohne daß zwischen verschiedenen autoregressiven Schätzverfahren differenziert werden kann. Dem entspricht der in der statistischen Literatur üblich gewordene Sprachgebrauch, pauschal alle autoregressiven Verfahren der Spektralschätzung als Maximum-Entropie-Schätzer zu bezeichnen. Dies sollte aber ausschließlich denjenigen Verfahren vorbehalten sein, bei denen versucht wird, das oben angegebene Prinzip der maximalen Entropie ausdrücklich auch bei der *Schätzung* zu berücksichtigen, wie z.B. bei dem Verfahren von BURG[1968],

oder neueren Methoden der direkten Maximum-Entropie-Spektralschätzung (vgl. z.B. DREHER[1990]), auf die weiter unten kurz eingegangen wird.

Die Aussage von Satz 3.3.1 ist auch insofern überraschend, als das Maximum-Entropie-Prinzip, das ja gemäß seiner allgemeinen Formulierung bei der Modellbildung möglichst neutral und unverbindlich sein will, zu einem speziellen parametrischen Ansatz in Form eines autoregressiven Modelles führt. Autoregressive Spektralschätzungen haben sich zwar bei vielen praktischen Anwendungen bewährt, aber die Approximation der Spektraldichte eines MA-Prozesses mit Hilfe eines autoregressiven Modells kann z.B. Probleme bereiten (vgl. z.B. das Beispiel in PRIESTLEY[1981, S. 611]).

Angesichts der vielfältigen Möglichkeiten zur Ausgestaltung des Maximum-Entropie-Konzeptes erscheint es nicht verwunderlich, daß sich hierzu wissenschaftlich kontroverse Positionen finden lassen. Ein wichtiger Diskussionspunkt besteht darin, ob die SHANNON'sche Definition einer Entropiefunktion die einzig sinnvolle Formalisierung für Ungewißheit und Unsicherheit darstellt. Unterschiedliche Positionen betonen z.B. die axiomatische Fundierung der SHANNON'schen Entropiefunktion (vgl. z.B. SHORE/JOHNSON[1983]) oder vertreten eine pragmatische, anwendungsorientierte Sichtweise (vgl. z.B. NARAYAN/NITYANANDA[1983]).

DREHER[1990, Abschnitt 4], der sich mit einem speziellen Anwendungsgebiet der Spektralanalyse beschäftigt, bringt eine kurze Übersicht verschiedener Konzepte der Entropie und stellt verschiedene relevante Entropiefunktionen vor, deren (nicht-lineare) Maximierung unter geeigneten Nebenbedingungen z.B. mit Hilfe des Cambridge-Algorithmus (vgl. SKILLING/BRYAN[1984]) numerisch erfolgt. Im Gegensatz zur autoregressiven Maximum-Entropie-Methode handelt es sich bei diesen Verfahren um nicht-parametrische Spektralschätzungen, die aber trotzdem für hochauflösende Schätzungen geeignet sind. Diese Klasse von Verfahren soll als direkte Maximum-Entropie-Schätzer bezeichnet werden. SKILLING/GULL[1985] übertragen ihre in der Bilddatenverarbeitung gemachten Erfahrungen auf das Problem der Spektralschätzung und fordern, daß wenn die Ermittlung der Gestalt einer Spektraldichte und nicht die Prognose der zugrundeliegenden Zeitreihe von Bedeutung ist, unmittelbar die normierte und diskretisierte Spektraldichte $h(\omega_j)$ unter Verwendung der Entropiefunktion

$$S(h) = - \sum_j h(\omega_j) \ln h(\omega_j)$$

mit Hilfe einer direkten Maximum-Entropie-Schätzung modelliert werden sollte. INGMANN/MERLIS[1992] zeigen, wie sich der Ansatz der direkten Maximum-Entropie-

Spektralschätzung mit Hilfe eines neuronalen Netzes implementieren läßt. Eine ausführlichere Betrachtung dieser Verfahren, die vom Ansatz her der ursprünglichen Intention des Maximum-Entropie-Konzeptes näher stehen als das rein theoretisch fundierte autoregressive Konzept, kann in dem hier gesetzten Rahmen nicht erfolgen.

Eine weitere Variante der Maximum-Entropie-Schätzung stammt von AKAIKE[1977a,b], der die Maximierung des Erwartungswertes des negativen Kullback-Leibler-Informationskriteriums als Maximum-Entropie-Schätzung definiert (vgl. Abschnitt 2.6). Dieser Ansatz kann als eine Verallgemeinerung der Maximum-Likelihood-Schätzung interpretiert werden, der es auch ermöglicht, Hilfskriterien zur Modellspezifikation abzuleiten.

Zusammenfassend kann festgehalten werden, daß sich aus dem Prinzip der maximalen Entropie kein einheitliches Konzept der Spektralschätzung ableiten läßt. Berücksichtigt man die vielen unterschiedlichen Ansätze, dann besagt das Prinzip der maximalen Entropie nicht mehr, als daß eine „geeignete“ Funktion unter „geeigneten“ Nebenbedingungen zu maximieren ist. Die bekannteste praktische Lösung ist der Burg-Algorithmus (vgl. Abschnitt 3.4.2), der allerdings als spezieller autoregressiver Schätzalgorithmus auch ohne Bezug auf das Konzept der maximalen Entropie abgeleitet werden kann. Wie andere autoregressive Schätzverfahren, bleibt der Burg-Algorithmus insofern unvollständig, als auch hierbei die Ordnung p des zu schätzenden AR-Prozesses vorgegeben werden muß.

3.4 Autoregressive Schätzverfahren

In diesem Abschnitt werden vier Methoden diskutiert, die unter noch zu spezifizierenden Bedingungen geeignet sind, Schätzwerte $(\hat{\phi}_1, \dots, \hat{\phi}_p, \hat{\sigma}^2)'$ für die Parameter eines autoregressiven Prozesses

$$(3.34) \quad X_t + \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} = \epsilon_t$$

mit

$$\phi_0 = 1, \phi_p \neq 0 \quad \text{und} \quad \{\epsilon_t\} \sim \text{WN}(0, \sigma^2).$$

zu liefern, so daß eine Spektralschätzung gemäß

$$(3.35) \quad \hat{f}^{AR}(\omega) = \frac{\hat{\sigma}^2}{|1 + \sum_{k=1}^p \hat{\phi}_k e^{-i2\pi\omega k}|^2} \quad \omega \in [-0.5, 0.5].$$

berechnet werden kann.

Alle vier Verfahren,

- Yule-Walker-Schätzung
- Burg-Maximum-Entropie-Schätzung
- FBLP(Forward-Backward-Linear-Prediction)-Schätzung
- FLP(Forward-Linear-Prediction)-Schätzung,

können unter der Normalverteilungsannahme als approximative Maximum-Likelihood-Schätzungen betrachtet werden, die für „große“ Stichprobenumfänge Resultate liefern, die sich nicht wesentlich voneinander unterscheiden. Asymptotisch besitzen Yule-Walker-, FLP- und Maximum-Likelihood-Schätzer die gleichen Eigenschaften (vgl. z.B. SCHLITZGEN/STREITBERG[1984, S.165ff]). Für den Burg-Schätzer, der ausschließlich rekursiv definiert ist, zeigt HAINZ[1992] für den allgemeineren multivariaten Fall, daß er die gleichen asymptotischen Eigenschaften wie der Yule-Walker-Schätzer besitzt. Für den FBLP-Schätzer liegen keine entsprechenden analytischen asymptotischen Aussagen vor. Die Ergebnisse eigener Simulationen verweisen jedoch darauf, daß für Stichprobenumfänge ab $N = 256$ die Unterschiede zwischen FBLP-Schätzer und Burg- bzw. FLP-Schätzer vernachlässigbar sind, sofern es sich bei dem datenerzeugenden Prozeß tatsächlich um einen $AR(p)$ -Prozeß handelt.

Für „kleine“ Stichprobenumfänge können die Schätzergebnisse jedoch erheblich voneinander abweichen. Es muß darauf hingewiesen werden, daß die Charakterisierung eines Stichprobenumfanges als „groß“ bzw. „klein“ von den Eigenschaften des datenerzeugenden Prozesses abhängt, wie im folgenden noch dargelegt wird. Ohne Kenntnis dieser Eigenschaften sind konkretere Zahlenangaben nicht sinnvoll. Da die unterschiedlichen Schätzeigenschaften bei „kleinen“ Stichprobenumfängen von allgemeinem theoretischen Interesse sind und bei praktischen Anwendungen der Algorithmen erhebliche Relevanz besitzen können, wird diesem Aspekt bei der Darstellung besondere Aufmerksamkeit gewidmet.

Bei der Berechnung des Spektrums gemäß (3.35) war unterstellt worden, daß der Output des Filters $(1 + \sum_{j=1}^p \hat{\phi}_j L^j)$ tatsächlich ein White-Noise-Prozeß ist. Ein allgemeinerer Ansatz lautet:

$$(3.36) \quad X_t + \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} = v_t,$$

wobei $\{v_t\}$ ein schwach stationärer Prozeß mit der Spektraldichte $f_v(\omega)$ sein soll. Eine Schätzung der Spektraldichte $f_X(\omega)$ kann z.B. berechnet werden, indem zunächst ein AR-Prozeß p -ter Ordnung angepaßt und für die Residuen

$$(3.37) \quad \hat{v}_t = X_t + \hat{\phi}_1 X_{t-1} + \dots + \hat{\phi}_p X_{t-p}$$

eine Spektraldichte $\hat{f}_v(\omega)$ z.B. mit Hilfe eines indirekten Spektralschätzers berechnet wird.

Die Spektralschätzung $\hat{f}_X(\omega)$ ergibt sich dann als (vgl. z.B. SCHLITTEGEN/STREITBERG[1984, S. 330]):

$$(3.38) \quad \hat{f}_X(\omega) = \frac{\hat{f}_v(\omega)}{|1 + \sum_{k=1}^p \hat{\phi}_k e^{-i2\pi\omega k}|^2} \quad \omega \in [-0.5, 0.5].$$

Diese Art der Spektralschätzung wird als Prewhitening und Recolouring bezeichnet: Im ersten Schritt (Gleichung (3.37)) wird eine Reihe erzeugt, die im allgemeinen ein flacheres Spektrum besitzt wie die Originalreihe, so daß die Spektralschätzung $\hat{f}_v(\omega)$ einen erheblich geringeren Bias aufweisen wird. Im zweiten Schritt (Gleichung (3.37)) erfolgt auf der Grundlage des bekannten Filters $(1 + \sum_{k=1}^p \hat{\phi}_k L^k)$ und Anwendung von Satz 2.4.1 eine Rücktransformation, die das Spektrum der Originalreihe $\{X_t\}$ liefert.

In einigen Fällen kann sogar auf eine Schätzung der Koeffizienten verzichtet werden, da a priori Informationen über geeignete Prewhitening-Filter vorliegen. Dies ist z.B. bei vielen makroökonomischen Reihen der Fall, wo bekannt ist, daß Peaks in der Nähe der Frequenz $\omega = 0$ und/oder an bestimmten Saisonfrequenzen auftreten. Geeignete Prewhitening-Filter auf der Basis von sog. Quasidifferenzen diskutieren z.B. KÖNIG/WOLTERS[1972, S. 78ff].

3.4.1 Yule-Walker-Schätzung

Bei dem bekanntesten Verfahren der autoregressiven Spektralschätzung werden in den erweiterten Yule-Walker-Gleichungen die Autokovarianzen durch die verzerrten Autokovarianzschätzer

$$(3.39) \quad \hat{\gamma}(\tau) = \begin{cases} \frac{1}{N} \sum_{t=1}^{N-\tau} (X_t - \bar{X})(X_{t-\tau}^* - \bar{X}^*) & \tau \geq 0 \\ \hat{\gamma}(\tau)^* & \tau < 0 \end{cases}$$

ersetzt. Das resultierende Gleichungssystem

$$(3.40) \quad \begin{bmatrix} \hat{\gamma}_0 & \hat{\gamma}_1^* & \dots & \hat{\gamma}_p^* \\ \hat{\gamma}_1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \hat{\gamma}_1^* \\ \hat{\gamma}_p & \dots & \hat{\gamma}_1 & \hat{\gamma}_0 \end{bmatrix} \begin{bmatrix} 1 \\ \phi_1 \\ \vdots \\ \phi_p \end{bmatrix} = \begin{bmatrix} \sigma_\epsilon^2 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

kann rekursiv mit Hilfe der Levinson-Durbin-Rekursion gelöst werden. In der ingenieurwissenschaftlichen Literatur wird das Schätzverfahren nach Yule-Walker meist als Autokorrelationsmethode bezeichnet.

Da bei der Yule-Walker-Schätzung die positiv-definiten Autokovarianzschätzer $\{\hat{\gamma}_\tau\}$ verwandt werden, ergibt sich nach Satz 3.2.2, daß die Wurzeln des charakteristischen Polynoms

$$1 + \hat{\phi}_1 z^{-1} + \dots + \hat{\phi}_p z^{-p} = 0$$

innerhalb des Einheitskreises liegen. Damit ist gewährleistet, daß es sich bei

$$\hat{\phi}(L) = 1 + \sum_{j=1}^p \hat{\phi}_j L^j$$

um einen kausal-invertierbaren, stabilen Filter handelt. Diese Eigenschaft wird bei den Yule-Walker-Schätzungen aber nur zu Lasten andererer wünschenswerter Schätzeigenschaften erreicht:

Es ist allgemein bekannt, daß Spektralschätzungen nach der Yule-Walker-Methode eine geringere Auflösung besitzen als andere autoregressive Schätzverfahren. Das schlechte

Auflösungsvermögen der Yule-Walker-Spektralschätzungen ist auf den zum Teil erheblichen Bias der Schätzungen (vgl. z.B. TJOSTHEIM/PAULSEN[1983], LYSNE/TJOSTHEIM[1987], PUKKILA[1988], HUANG[1990]) bei „kleinen“ Stichprobenumfängen zurückzuführen. Während z.B. TJOSTHEIM/PAULSEN[1983] für einfache AR-Prozesse den Bias der Yule-Walker-Schätzer auch auf analytischem Weg ableiten, ergibt sich ein wesentlicher Hinweis zur Erklärung der unbefriedigenden Schätzeigenschaften der Yule-Walker-Lösungen dadurch, daß die Yule-Walker-Gleichungen als Normalgleichungen eines speziellen Kleinst-Quadrate-Problems dargestellt werden (vgl. z.B. MARPLE [1987, S. 216ff], NEWTON[1988, S. 206ff]). Hierzu betrachtet man die Regression:

$$(3.41) \quad \mathbf{x}_0 = -\mathbf{X}\boldsymbol{\phi} + \boldsymbol{\epsilon},$$

mit

$$\boldsymbol{\epsilon} = (\epsilon_1, \dots, \epsilon_{N+p})', \quad E(\boldsymbol{\epsilon}) = \mathbf{0}, \quad \text{Var}(\boldsymbol{\epsilon}) = \sigma^2 \mathbf{I},$$

$$\boldsymbol{\phi} = (\phi_1, \dots, \phi_p)',$$

$$\mathbf{X}_p = [\mathbf{x}_0 \ \mathbf{X}] = \begin{bmatrix} X_1 & 0 & \cdots & 0 \\ X_2 & X_1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ X_{p+1} & & X_2 & X_1 \\ \vdots & \ddots & \ddots & \vdots \\ X_N & X_{N-1} & & X_{N-p} \\ 0 & X_N & \ddots & X_{N-p+1} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & X_N \end{bmatrix}.$$

In der $(N+p) \times (p+1)$ -Matrix \mathbf{X}_p sind die für das Gleichungssystem (3.41) erforderlichen, nicht in der Stichprobe verfügbaren Werte

$$\{X_0, \dots, X_{N-p+1}\} \quad \text{und} \quad \{X_{N+1}, \dots, X_{N+p}\}$$

durch Nullen ersetzt. Die Normalgleichungen der Regression (3.41) ergeben sich als:

$$(3.42) \quad \mathbf{X}^H \mathbf{X} \phi = -\mathbf{X}^H \mathbf{x}_0.$$

Da sich für die optimale Lösung $\hat{\phi}$ der Normalgleichungen (3.42) aufgrund der Orthogonalitätsbeziehungen im linearen Regressionsmodell die Quadratsumme der Residuen darstellen läßt als:

$$(3.43) \quad \hat{\epsilon}^H \hat{\epsilon} = \mathbf{x}_0^H \mathbf{x}_0 + \mathbf{x}_0^H \mathbf{X} \hat{\phi},$$

können (3.42) und (3.43) zu den erweiterten Normalgleichungen zusammengefaßt werden:

$$(3.44) \quad \begin{bmatrix} \mathbf{x}_0^H \mathbf{x}_0 & \mathbf{x}_0^H \mathbf{X} \\ \mathbf{X}^H \mathbf{x}_0 & \mathbf{X}^H \mathbf{X} \end{bmatrix} \begin{bmatrix} 1 \\ \hat{\phi} \end{bmatrix} = [\mathbf{x}_0 \ \mathbf{X}]^H [\mathbf{x}_0 \ \mathbf{X}] \begin{bmatrix} 1 \\ \hat{\phi} \end{bmatrix} = \begin{bmatrix} \hat{\epsilon}^H \hat{\epsilon} \\ \mathbf{0} \end{bmatrix}.$$

Aufgrund der Struktur der Matrix $\mathbf{X}_p = [\mathbf{x}_0 \ \mathbf{X}]$ (vgl. hierzu auch S. 39) ergibt sich aus (3.44) nach Division durch N , daß die optimale Lösung des Regressionsproblems (3.41) und die Lösung der erweiterten Yule-Walker-Gleichungen (3.40) mit den verzerrten Autokovarianzschätzern $\{\hat{\gamma}_\tau\}$ äquivalent sind.

Die Darstellung als Kleinst-Quadrate-Problem zeigt, daß die Yule-Walker-Schätzung eine Prä- und Postextrapolation der Stichprobe mit Nullen impliziert. Dies erzwingt zwar die Positiv-Definitheit der Autokovarianzschätzer, bewirkt aber andererseits einen Bias, der eine geringere Spektralaufösung der über die Yule-Walker-Lösungen geschätzten Spektraldichten nach sich zieht. Der Bias hängt von den Eigenschaften des datenerzeugenden AR-Prozesses ab. Relevant ist insbesondere der Abstand der Wurzeln des charakteristischen Polynoms vom Einheitskreis. PUKKILA[1988] hebt als wesentliches Resultat einer Simulationsstudie, bei der ein AR(4)-Prozeß mit Wurzeln nahe dem Einheitskreis zugrunde gelegt wurde, hervor, daß nicht nur bei einem Stichprobenumfang von $N = 200$ noch ein erheblicher Bias der Yule-Walker-Schätzungen zu beobachten war, sondern daß sich sogar eine Zunahme der Standardabweichungen der Parameterschätzungen in Abhängigkeit von den Stichprobenumfängen $N = 50, 100, 200$ ergab. DAHLHAUS[1984] zeigt, daß sich der Bias des Yule-Walker-Schätzers durch Anwendung eines Datentapers mindern läßt. Es wird im allgemeinen aber vorteilhafter sein, einen

autoregressiven Schätzer zu wählen, bei dem diese Probleme gar nicht erst auftreten können.

Drei Alternativen zu den Yule-Walker-Schätzungen, mit denen sich im allgemeinen bessere Ergebnisse erzielen lassen, werden in den nächsten Unterabschnitten vorgestellt.

3.4.2 Burg-Schätzung

Bei der Burg-Schätzung werden die nach dem Prinzip der maximalen Entropie nicht akzeptablen Autokovarianzschätzer $\{\hat{\gamma}_\tau\}$ vermieden. Vielmehr erfolgt eine Modifikation der Levinson-Durbin-Rekursion, indem für die Reflektionskoeffizienten κ_j eine direkte Schätzung aus der Stichprobe $(X_1, \dots, X_N)'$ eingeführt wird, die mit dem Grundkonzept der maximalen Entropie verträglich ist.

Ausgangspunkt der Burg-Schätzung ist, daß sich bei Anwendung der Levinson-Durbin-Rekursion (vgl. Gleichung (3.22)) sowie der Definitionen für den Vorwärts- bzw. Rückwärtsprognosefehler, die Vorwärts- und Rückwärtsprognosefehler auf der j -ten Stufe der Rekursion darstellen lassen als (vgl. z.B. MARPLE [1987, S. 211]):

$$(3.45) \quad \epsilon_{j,t}^f = \hat{\epsilon}_{j-1,t}^f + \kappa_j \hat{\epsilon}_{j-1,t-1}^b$$

$$(3.46) \quad \epsilon_{j,t}^b = \hat{\epsilon}_{j-1,t-1}^b + \kappa_j^* \hat{\epsilon}_{j-1,t}^f.$$

Mit den Gleichungen (3.45) und (3.46) lassen sich die Vorwärts- bzw. Rückwärtsprognosefehler der Stufe j in Abhängigkeit von den bekannten Vorwärts- bzw. Rückwärtsprognosefehlern der Stufe $j - 1$ sowie dem unbekanntem Reflektionskoeffizienten κ_j darstellen. Die Initialisierung der Rekursionen erfolgt mit

$$\hat{\epsilon}_{0,t}^f = \hat{\epsilon}_{0,t}^b = X_t \quad \text{für } t = 1, \dots, N.$$

Unter Verwendung der Rekursionen (3.45) und (3.46) lassen sich der mittlere quadratische Vorwärts- und Rückwärtsprognosefehler in Abhängigkeit von dem Parameter κ_j darstellen als:

$$(3.47) \quad \tilde{P}_j^f(\kappa_j) = \frac{1}{N} \sum_{t=j+1}^N |\epsilon_{j,t}^f|^2 = \frac{1}{N} \sum_{t=j+1}^N |\hat{\epsilon}_{j-1,t}^f + \kappa_j \hat{\epsilon}_{j-1,t-1}^b|^2$$

bzw.

$$(3.48) \quad \tilde{P}_j^b(\kappa_j) = \frac{1}{N} \sum_{t=j+1}^N |\epsilon_{j,t}^b|^2 = \frac{1}{N} \sum_{t=j+1}^N |\hat{\epsilon}_{j-1,t-1}^b + \kappa_j^* \hat{\epsilon}_{j-1,t}^f|^2$$

Die Minimierungsansätze

$$\min_{\kappa_j} \tilde{P}_j^f(\kappa_j) \quad \text{bzw.} \quad \min_{\kappa_j} \tilde{P}_j^b(\kappa_j)$$

führen zu den Lösungen:

$$(3.49) \quad \hat{\kappa}_j^f = - \frac{\sum_{t=j+1}^N \hat{\epsilon}_{j-1,t}^f \hat{\epsilon}_{j-1,t-1}^{b*}}{\sum_{t=j+1}^N |\hat{\epsilon}_{j,t-1}^b|^2}$$

bzw.

$$(3.50) \quad \hat{\kappa}_j^b = - \frac{\sum_{t=j+1}^N \hat{\epsilon}_{j-1,t}^f \hat{\epsilon}_{j-1,t-1}^{b*}}{\sum_{t=j+1}^N |\hat{\epsilon}_{j,t}^f|^2}.$$

Die Besonderheit des Burg-Algorithmus besteht nun darin, beide Lösungsansätze zu integrieren, indem auf jeder Stufe j das arithmetische Mittel des mittleren quadratischen Vorwärts- und Rückwärtsprognosefehlers minimiert wird:

$$(3.51) \quad \min_{\kappa_j} \tilde{P}_j^{fb}(\kappa_j) = \frac{1}{2} \left(\frac{1}{N} \sum_{t=j+1}^N |\epsilon_{j,t}^f|^2 + \frac{1}{N} \sum_{t=j+1}^N |\epsilon_{j,t}^b|^2 \right).$$

Dies führt zu folgender Schätzung für κ_j (vgl. z.B. MARPLE[1987, S. 213]):

$$(3.52) \quad \hat{\kappa}_j^{fb} = - \frac{2 \sum_{t=j+1}^N \hat{\epsilon}_{j-1,t}^f \hat{\epsilon}_{j-1,t-1}^{b*}}{\sum_{t=j+1}^N |\hat{\epsilon}_{j,t}^f|^2 + \sum_{t=j+1}^N |\hat{\epsilon}_{j,t-1}^b|^2}.$$

Die Reflektionskoeffizienten $\hat{\kappa}_j^{fb}$ der Burg-Schätzung können als harmonisches Mittel der Reflektionskoeffizienten $\hat{\kappa}_j^f$ und $\hat{\kappa}_j^b$ interpretiert werden, die sich bei alleiniger Minimierung von Vorwärts- bzw. Rückwärtsprognosefehler ergeben, d.h. es gilt:

$$(3.53) \quad \hat{\kappa}_j^{fb} = \frac{2}{1/\hat{\kappa}_j^f + 1/\hat{\kappa}_j^b}.$$

Die Burg-Maximum-Entropie-Methode ist also ein iteratives, lokales Optimierungsverfahren, bei dem auf jeder Stufe j ein optimaler Reflektionskoeffizient $\hat{\kappa}_j$ unter gleichzeitiger Berücksichtigung von Vorwärts- und Rückwärtsprognosefehlern ermittelt wird. Die Lösung entspricht dem Prinzip der maximalen Entropie, da in der Summation der Prognosefehler die für die autoregressive Prognose erforderlichen j Startwerte unberücksichtigt bleiben, so daß nur tatsächlich berechenbare Prognosefehler berücksichtigt werden und keinerlei Annahmen über Werte außerhalb des durch die Stichprobe gegebenen Stützbereiches gemacht werden müssen.

Durch die Einbindung in die Levinson-Durbin-Rekursion (vgl. (3.24) und (3.25)) gilt auch bei der Burg-Methode für die Reflektionskoeffizienten

$$\hat{\kappa}_j^{fb} \leq 1.$$

Dies ist gleichbedeutend damit, daß die Wurzeln des charakteristischen Polynoms eines nach der Burg-Methode geschätzten AR-Prozesses innerhalb oder auf dem Einheitskreis liegen können.

Der Burg-Algorithmus repräsentiert nur eine von zahlreichen Möglichkeiten, Schätzer für die Reflektionskoeffizienten κ_j zu konstruieren.

Eine naheliegende Erweiterung ist der gewichtete Burg-Schätzer:

$$(3.54) \quad \hat{\kappa}_j^{fb} = -\frac{2 \sum_{t=j+1}^N w_{j-1,t} \hat{\epsilon}_{j-1,t}^f \hat{\epsilon}_{j-1,t}^{b*}}{\sum_{t=j+1}^N w_{j-1,t} |\hat{\epsilon}_{j,t}^f|^2 + \sum_{t=j+1}^N w_{j-1,t} |\hat{\epsilon}_{j,t-1}^b|^2},$$

der u.a. dazu verwandt wurde, die Schätzeigenschaften des Burg-Schätzers für harmonische Prozesse in Noise zu verbessern. Hinweise zu geeigneten Gewichten $w_{j-1,t} \geq 0$ finden sich z.B. in SWINGLER[1979] oder HELME/NIKIAS[1985].

Eine andere Art der Verallgemeinerung findet sich bei MAKHOUL[1977], der eine Klasse von Schätzern für reellwertige Reflektionskoeffizienten als verallgemeinertes m -tes Mittel aus $\hat{\kappa}_j^f$ und $\hat{\kappa}_j^b$ definiert:

$$(3.55) \quad \hat{\kappa}_j^{fb}(m) = S \left[\frac{1}{2} (|\hat{\kappa}_j^f|^m + |\hat{\kappa}_j^b|^m) \right]^{1/m} \quad m \leq 0,$$

mit $S = \text{sign}(\kappa_j^f) = \text{sign}(\kappa_j^b)$.

Mit $m = -1$ ergibt sich der Burg-Schätzer als Spezialfall. Andere a priori Vorgaben für m sind möglich, allerdings sind die daraus resultierenden Reflektionskoeffizienten im allgemeinen nicht als Lösung eines Fehlerminimierungsansatzes interpretierbar.

Auf der Grundlage des Burg-Algorithmus können Spektraldichten mit hoher Auflösung geschätzt werden. Probleme mit dem Burg-Algorithmus ergeben sich bei der Schätzung von harmonischen Prozessen in Noise, wo phasenabhängiges Peak-Shifting und Spectral-Line-Splitting auftreten kann. Beispiele sind in Abschnitt 3.4.5 angegeben. Darüberhinaus sind zahlreiche Verallgemeinerungen des Burg-Algorithmus möglich, wodurch in Kombination mit der Levinson-Durbin-Rekursion zahlreiche weitere sinnvolle Schätzungen der Reflektionskoeffizienten und damit auch der Autokovarianzen möglich sind.

3.4.3 FBLP-Schätzung

Bei der Forward-Backward-Linear-Prediction-(FBLP)Schätzung, die auf unabhängige Arbeiten von ULRICH/CLAYTON[1976] und NUTTAL[1976] zurückgeht, wird das Konzept der gemeinsamen Minimierung von Vorwärts- und Rückwärtsprognosefehler auf den Kleinst-Quadrat-Ansatz zur Schätzung autoregressiver Modelle übertragen. Neben der Bezeichnung FBLP-Schätzung sind auch die Bezeichnungen Least-Squares-Schätzung, FBLS-(Forward-Backward-Least-Squares)-Schätzung und modifizierte Kovarianzmethode zu finden.

Analog wie bei der Burg-Schätzung ist das arithmetische Mittel der summierten quadratischen Vorwärts- und Rückwärtsprognosefehler zu minimieren:

$$(3.56) \quad \min_{\phi^{fb}} \tilde{P}_{ss}^{fb} = \frac{1}{2} (\tilde{P}_{ss}^f + \tilde{P}_{ss}^b).$$

mit:

$$(3.57) \quad \tilde{P}_{ss}^f = \sum_{t=p+1}^N |\epsilon_t^f|^2$$

$$(3.58) \quad = \sum_{t=p+1}^N |X_t + \phi_1 X_{t-1} + \dots + \phi_p X_{t-p}|^2.$$

und

$$(3.59) \quad \tilde{P}_{ss}^b = \sum_{t=p+1}^N |\epsilon_t^b|^2$$

$$(3.60) \quad = \sum_{t=p+1}^N |X_{t-p} + \phi_1^* X_{t-p+1} + \dots + \phi_p^* X_t|^2.$$

Mit den Vektoren

$$\boldsymbol{\epsilon}^f = (\epsilon_{p+1}^f, \dots, \epsilon_N^f)', \quad \boldsymbol{\epsilon}^b = (\epsilon_{p+1}^b, \dots, \epsilon_N^b)', \quad \boldsymbol{\phi} = (1, \phi_1, \dots, \phi_p)'$$

und der $(N-p) \times (p+1)$ Matrix

$$\mathbf{X}_p = [\mathbf{x}_0 \ \mathbf{X}] = \begin{bmatrix} X_{p+1} & \cdots & X_2 & X_1 \\ \vdots & \ddots & \ddots & \vdots \\ X_N & \cdots & X_{N-p+1} & X_{N-p} \end{bmatrix}$$

lassen sich die Prognosefehler darstellen als:

$$(3.61) \quad \boldsymbol{\epsilon}^f = \mathbf{X}_p \boldsymbol{\phi}$$

$$(3.62) \quad \boldsymbol{\epsilon}^b = \mathbf{X}_p \mathbf{J} \boldsymbol{\phi}^* \quad \text{bzw.} \quad \boldsymbol{\epsilon}^{b*} = \mathbf{X}_p^* \mathbf{J} \boldsymbol{\phi},$$

wobei \mathbf{J} eine ordnungskonforme Reflektionsmatrix ist.

Die summierten quadratischen Prognosefehler \tilde{P}_{ss}^{fb} ergeben sich in Matrixschreibweise als:

$$(3.63) \quad \begin{aligned} \tilde{P}_{ss}^{fb} &= \frac{1}{2} [(\boldsymbol{\epsilon}^f)^H (\boldsymbol{\epsilon}^f) + (\boldsymbol{\epsilon}^b)^H (\boldsymbol{\epsilon}^b)] \\ &= \frac{1}{2} [(\mathbf{X}_p \boldsymbol{\phi})^H (\mathbf{X}_p \boldsymbol{\phi}) + (\mathbf{X}_p^* \mathbf{J} \boldsymbol{\phi})^H (\mathbf{X}_p^* \mathbf{J} \boldsymbol{\phi})] \\ &= \frac{1}{2} [\boldsymbol{\phi}^H (\mathbf{X}_p^H \mathbf{X}_p + \mathbf{J} \mathbf{X}_p' \mathbf{X}_p^* \mathbf{J}) \boldsymbol{\phi}] \end{aligned}$$

Minimierung von (3.63) bezüglich $\boldsymbol{\phi} = (\phi_0, \dots, \phi_p)'$ unter der Nebenbedingung $\phi_0 = 1$ führt zu dem Lagrangeansatz

$$(3.64) \quad \mathcal{L}(\boldsymbol{\phi}, \lambda) = \frac{1}{2} \left[\boldsymbol{\phi}^H (\mathbf{X}_p^H \mathbf{X}_p + \mathbf{J} \mathbf{X}_p' \mathbf{X}_p^* \mathbf{J}) \boldsymbol{\phi} \right] + \lambda (1 - \boldsymbol{\phi}^H \boldsymbol{\iota})$$

mit $\boldsymbol{\iota} = (1, 0, \dots, 0)'$.

Die Bedingungen erster Ordnung lauten:

$$(3.65) \quad \frac{\partial \mathcal{L}}{\partial \boldsymbol{\phi}} = (\mathbf{X}_p^H \mathbf{X}_p + \mathbf{J} \mathbf{X}_p' \mathbf{X}_p^* \mathbf{J}) \boldsymbol{\phi} - \lambda \boldsymbol{\iota} = 0$$

$$(3.66) \quad \frac{\partial \mathcal{L}}{\partial \lambda} = (1 - \boldsymbol{\phi}^H \boldsymbol{\iota}) = 0.$$

Aus der Prämultiplication von Gleichung (3.66) mit $\boldsymbol{\phi}^H$ ergibt sich $\lambda = \tilde{P}_{ss}^{fb}$, so daß die erweiterten Normalgleichungen sich ergeben als:

$$(3.67) \quad (\mathbf{X}_p^H \mathbf{X}_p + \mathbf{J} \mathbf{X}_p' \mathbf{X}_p^* \mathbf{J}) \boldsymbol{\phi} = \begin{bmatrix} \tilde{P}_{ss}^{fb} \\ \mathbf{0} \end{bmatrix}$$

Der mittlere quadratische Prognosefehler wird geschätzt als:

$$\hat{P} = \frac{\tilde{P}_{ss}^{fb}}{N - p}.$$

Das Gleichungssystem (3.67) kann nicht unter Verwendung der Levinson-Durbin-Rekursion gelöst werden, da die Matrix $(\mathbf{X}_p^H \mathbf{X}_p + \mathbf{J} \mathbf{X}_p' \mathbf{X}_p^* \mathbf{J})$ zwar hermitisch ist, aber keine Toeplitzstruktur besitzt. Allerdings ist \mathbf{X}_p eine (Rechteck-)Toeplitzmatrix und $\mathbf{X}_p^* \mathbf{J}$ eine Hankel-Matrix, so daß spezielle Eigenschaften vorhanden sind, welche die Entwicklung eines effizienten, rekursiven Algorithmus erlauben. Auf die Darstellung des rekursiven Algorithmus wird hier verzichtet. Bei der Ableitung der Rekursion ergibt sich jedoch u.a., daß für den quadratischen Prognosefehler die Beziehung

$$\tilde{P}_{ss,j}^{fb} = \tilde{P}_{ss,j-1}^{fb} (1 - |\phi_{jj}|^2)$$

erfüllt ist (vgl. MARPLE[1987, S. 251ff]), so daß aufgrund von

$$\tilde{P}_{ss,j}^{fb} \geq 0 \quad \text{und} \quad \tilde{P}_{ss,j-1}^{fb} \geq 0,$$

analog wie bei der Levinson-Durbin-Rekursion, folgt:

$$|\phi_{jj}| \leq 1,$$

so daß bei der FBLP-Schätzung die Wurzeln des charakteristischen Polynoms innerhalb oder auf dem Einheitskreis liegen.

MARPLE[1980] verweist darauf, daß in seltenen Fällen Wurzeln des charakteristischen Polynoms eines autoregressiven Prozesses, dessen Koeffizienten mit Hilfe der FBLP-Methode geschätzt wurden, außerhalb des Einheitskreises liegen können. Nach eigenen Erfahrungen trat dieses Problem nur dann auf, wenn die Matrix $(\mathbf{X}_p^H \mathbf{X}_p + \mathbf{J} \mathbf{X}_p' \mathbf{X}_p^* \mathbf{J})$ sehr schlecht konditioniert war.

Vorteile gegenüber anderen autoregressiven Schätzverfahren besitzt der FBLP-Algorithmus insbesondere bei der Schätzung von harmonischen Prozessen in Noise (vgl. auch Abschnitt 3.4.5): Phasenabhängiges Peak-Shifting ist vernachlässigbar gering, Line-Splitting ist bei Anwendung des FBLP-Algorithmus noch nicht beobachtet worden (vgl. z.B. KAY[1988, S. 227ff], MARPLE[1987, S. 224]). Die guten Ergebnisse des FBLP-Algorithmus bei der Schätzung von harmonischen Prozessen in Noise führten zu einer Weiterentwicklung des Verfahrens durch TUFTS/KUMARESAN[1982], wobei der Vektor ϕ durch eine geeignete Linearkombination der Eigenvektoren der Matrix $(\mathbf{X}_p^H \mathbf{X}_p + \mathbf{J} \mathbf{X}_p' \mathbf{X}_p^* \mathbf{J})$ geschätzt wird.

3.4.4 FLP-Schätzung

Bei der FLP-Methode wird nur der Vorwärtsprognosefehler minimiert, so daß dieser Ansatz der Anwendung des allgemeinen Kleinst-Quadrate-Prinzips entspricht. Im Vergleich zu den anderen drei Verfahren wird der geringste Gebrauch von der Annahme der schwachen Stationarität des datenerzeugenden Prozesses gemacht.

Der Minimierungsansatz lautet:

$$(3.68) \quad \min_{\phi^f} \tilde{P}_{ss}^f = (\epsilon^f)^H \epsilon^f.$$

Analog wie bei der FBLP-Methode, lassen sich die erweiterten Normalgleichungen ableiten als:

$$(3.69) \quad \mathbf{X}_p^H \mathbf{X}_p \phi = \begin{bmatrix} \tilde{P}_{ss}^f \\ \mathbf{0} \end{bmatrix}.$$

Der mittlere quadratische Prognosefehler wird geschätzt als:

$$\hat{P} = \frac{\tilde{P}_{ss}^f}{N - p}.$$

Die Wurzeln des charakteristischen Polynoms eines nach der FLP-Methode geschätzten Prognosefilters müssen nicht innerhalb des Einheitskreises liegen (vgl. z.B. MARPLE[1987, S. 221]). Diese Eigenschaft ist allerdings in Hinblick auf die Spektralschätzung von untergeordneter Bedeutung. Bezüglich Spectral-Line-Splittung bzw. Peak-Shifting verhält sich die FLP-Methode genauso unproblematisch wie die FBLP-Methode.

Nach eigenen Erfahrungen ergaben sich jedoch erhebliche Unterschiede bei der Schätzung von harmonischen Prozessen in Noise, wenn in dem oben erwähnten Verfahren von TUFTS/KUMARESAN[1982] statt Linearkombinationen von Eigenvektoren der Matrix $(\mathbf{X}_p^H \mathbf{X}_p + \mathbf{J} \mathbf{X}_p' \mathbf{X}_p^* \mathbf{J})$ lediglich Linearkombinationen von Eigenvektoren der Matrix $\mathbf{X}_p^H \mathbf{X}_p$ verwandt wurden. Beide Matrizen sind lediglich symmetrisch und besitzen keine Toeplitzstruktur, wie es unter der Annahme schwach stationärer stochastischer Prozesse notwendig wäre. Die Abweichung von der Toeplitzstruktur bei der Matrix $\mathbf{X}_p^H \mathbf{X}_p$ scheint jedoch gravierender zu sein als bei der Matrix $(\mathbf{X}_p^H \mathbf{X}_p + \mathbf{J} \mathbf{X}_p' \mathbf{X}_p^* \mathbf{J})$. Eine detaillierte Untersuchung dieses Problems liegt außerhalb des hier gesetzten Rahmens.

3.4.5 Vergleich autoregressiver Schätzverfahren

Wie bereits ausgeführt worden war, können die vier autoregressiven Schätzverfahren nach der Yule-Walker-, Burg-, FBLP- bzw. FLP-Methode als approximative Maximum-Likelihood-Schätzer betrachtet werden, deren Schätzeigenschaften sich für „große“ Stichprobenumfänge nicht wesentlich unterscheiden sollten. Bei „kleinen“ Stichprobenumfängen können sich jedoch erhebliche Unterschiede zwischen den Schätzern ergeben. Dies wird im folgenden mit einfachen Simulationsbeispielen für die Eigenschaften Auflösungsvermögen, Spectral-Line-Splittung und Peak-Shifting demonstriert. Die praktische Relevanz dieser Eigenschaften variiert mit dem substanzwissenschaftlichen Einsatzbereich der Schätzverfahren. Es ist allerdings von grundsätzlichem theoretischen Interesse, daß Unterschiede zwischen den Schätzern existieren, die eher qualitativen Charakter haben.

Auflösungsvermögen. Auf das schlechte Auflösungsvermögen des Yule-Walker-Verfahrens, im Gegensatz zur Burg-, FBLP- und FLB-Schätzung, war bereits hingewiesen worden. Dieses Verhalten wird hier an Hand eines reellwertigen AR(8)-Prozesses

$$(3.70) \quad X_t + \sum_{j=1}^8 \phi_j X_{t-j} = \epsilon_t \quad \epsilon_t \sim \text{WN}(0, 1),$$

demonstriert, für dessen charakteristisches Polynom $\prod_{j=1}^8 (1 - z_j z^{-1})$ die folgenden vier Paare konjugiert-komplexer Wurzeln gewählt wurden:

$$(3.71) \quad \begin{aligned} z_{1/2} &= 0.999e^{\pm i2\pi 0.125} & z_{3/4} &= 0.999e^{\pm i2\pi 0.135} \\ z_{5/6} &= 0.999e^{\pm i2\pi 0.145} & z_{7/8} &= 0.999e^{\pm i2\pi 0.155} . \end{aligned}$$

In Abbildung 3.1 (a) und Abbildung 3.1 (b) sind das theoretische Spektrum bzw. die Lage der vier Wurzelpaare bezüglich des Einheitskreises graphisch dargestellt. Das theoretische Spektrum des Prozesses weist vier scharfe Peaks an den Frequenzen $\omega = 0.125, 0.135, 0.145, 0.155$ auf.

Für die Schätzungen nach der Yule-Walker-, Burg-, FBLP-, und FLP-Methode wurden für die Stichprobenumfänge $N = 64$ und $N = 256$ jeweils 100 unabhängige Realisationen des in (3.70) und (3.71) beschriebenen AR-Prozesses erzeugt. Die Anzahl der Einschwingwerte betrug $N_E = 512$, um den Einfluß der mit Null initialisierten acht Anfangswerte geringzuhalten. Zur Auswertung wurden zum einen die aus jeweils 100 Realisationen gemittelten Spektren herangezogen (vgl. Abb. 3.2 (a) - (d), 3.4(a) - (d)). Zum anderen wurden, um einen besseren Eindruck über die Streuung der einzelnen Schätzungen zu erhalten, in den Abbildungen 3.3 (a) - (d) und 3.5 (a) - (d) Wurzelplots für jeden Schätzer dargestellt, wobei für jeden Stichprobenumfang und jeden Schätzer jeweils $8 \times 100 = 800$ Realizationen der Wurzeln z_j geplottet wurden. Auf eine tabellarische Darstellung der mittleren geschätzten Koeffizienten, deren Standardabweichungen sowie weiterer anderer ermittelter Größen wurde aus Platzgründen verzichtet.

Um numerische Probleme bei der Yule-Walker-Schätzung mitzubersichtigen, wurde die Lösung der Yule-Walker-Gleichungen sowohl nach der Levinson-Durbin-Rekursion als auch mit Hilfe einer Cholesky-Zerlegung berechnet. Obwohl die Wurzeln des datenerzeugenden Prozesses sehr nahe am Einheitskreis lagen, ergab ein Vergleich der

Koeffizienten jedoch lediglich unwesentliche Abweichungen ($< 10E-14$), so daß nur die Lösung nach der Levinson-Durbin-Rekursion berücksichtigt wurde.

Für den Stichprobenumfang $N = 64$ zeigt der graphische Vergleich der geschätzten Spektren (siehe Abb. 3.2 (a) - (d)), daß der Burg-Schätzer die vier Peaks des theoretischen Spektrums am besten wiedergibt, etwas weniger stark ausgeprägt sind die Peaks bei der FBLP-Schätzung. Bei der FLP-Schätzung ist lediglich ein Peak klar zu erkennen, die restlichen drei Peaks sind aber, wenn auch mit Mühe, wahrnehmbar. Ganz im Gegensatz dazu weist das gemittelte Spektrum der Yule-Walker-Schätzung nur einen einzigen Peak aus. Die dazugehörigen Wurzelplots (siehe Abb. 3.3 (a) - (d)) bestätigen die aus den gemittelten Spektren gewonnenen Informationen: Die Wurzeln des Burg-Schätzers liegen im Vergleich zu den anderen drei Schätzern dem Einheitskreis am nächsten. Beim FBLP- und FLP-Schätzer streuen die Wurzeln leicht in den Einheitskreis hinein, beim FLP-Schätzer stärker ausgeprägt als beim FBLP-Schätzer. Ein völlig anderes Bild ergibt sich beim Yule-Walker-Schätzer: Es existieren verschiedene Cluster von Wurzeln auch weit im inneren des Einheitskreises und es treten zahlreiche reelle Wurzeln auf, obwohl der datenerzeugende Prozeß nur konjugiert-komplexe Wurzeln besitzt. Am Wurzelplot des Burg-Schätzers läßt sich noch eine abweichende Eigenart im Vergleich zur FBLP- und FLP-Schätzung erkennen. Während die Wurzeln der beiden letzteren fast ausschließlich in den durch die Lage der theoretischen Wurzeln sowie dem Ursprung markierten Kreissegmenten streuen, reicht die Streuung der Wurzeln des Burg-Schätzers über diese Kreissegmente hinaus, wobei jedoch die Wurzeln in jedem Fall sehr nahe am Einheitskreis liegen. Diese Eigenart, bei kleinen Stichprobenumfängen die Peaks eines Spektrums nicht nur deutlicher als andere Schätzer wiederzugeben, sondern diese auch noch gegenüber den theoretischen Peaks mit größeren Abständen auszuweisen, sind vermutlich der Grund dafür, daß der Burg-Schätzer bei Praktikern als hochauflösender Schätzer sehr beliebt ist. Diese Eigenart läßt sich auch anhand der Darstellung einzelner Realisationen zeigen, auf deren Wiedergabe hier verzichtet wird.

Für den Stichprobenumfang $N = 256$ zeigt sowohl der graphische Vergleich der geschätzten Spektren (siehe Abb. 3.4 (b) - (d)), als auch der Wurzelplots (siehe Abb. 3.5 (b) - (d)), daß zwischen Burg-, FBLP- und FLP-Schätzer kaum noch visuelle Unterschiede erkennbar sind. Im Gegensatz dazu weist der Yule-Walker-Schätzer auch bei dem Stichprobenumfang $N = 256$ lediglich nur einen Peak im Spektrum auf (vgl. Abb. 3.4 (a)), bei dem zugehörigen Wurzelplot (siehe Abb. 3.5 (a)), ergibt sich im Vergleich zum

Stichprobenumfang $N = 64$ eine geringere Streuung um die verschiedenen wahrnehmbaren Cluster. Für den Yule-Walker-Schätzer ergab sich auch bei einem Stichprobenumfang von $N = 512$ keinerlei Veränderung, auf eine Wiedergabe der Graphik wird verzichtet.

Spectral-Line-Splitting. Das Phänomen Spectral-Line-Splitting und das unterschiedliche Verhalten der Schätzverfahren ist in Abbildung 3.6 dargestellt. Hierbei wurde als datenerzeugender Prozeß

$$X_t = \sin(2\pi t 0.2 + \pi/4) + \epsilon_t \quad \text{mit} \quad \{\epsilon_t\} \sim \text{WN}(0, 0.0005)$$

zugrundegelegt und jeweils 100 unabhängige Realisationen erzeugt.

Wie in den Vorbemerkungen zur autoregressiven Spektralschätzung bereits ausgeführt worden war, kann ein Prozeß dieser Art sehr gut durch einen AR(2)-Prozeß beschrieben werden. Wird in Unkenntnis des wahren datenerzeugenden harmonischen Prozesses ein AR(8)-Prozeß für die Schätzung verwandt, so treten bei dem Burg-Schätzer zwei nahe beieinanderliegende ausgeprägte Peaks in der Umgebung der Frequenz $\omega = 0.2$ auf, die fälschlicherweise auf die Existenz einer zweiten stark ausgeprägten Schwingungskomponente verweisen (vgl. Abb.3.6 (b)). Bei der Yule-Walker-, FBLP- und der FLP-Schätzung tritt auch in diesem Fall nur jeweils ein ausgeprägter Peak auf (vgl. Abb.3.6 (a) und (d)). Allerdings läßt sich dieses Phänomen auch unter geeigneten Extrembedingungen für den Yule-Walker-Schätzer sowie für weitere hier nicht vorgestellte autoregressive Spektralschätzer zeigen, lediglich von dem FBLP-Schätzer liegen bisher keine Berichte über das Auftreten von Spectral-Line-Splitting vor (vgl. z.B. MARPLE[1987, S. 224]). Für die autoregressive Spektralschätzung mit datengestützten Kriterien zur Bestimmung der Ordnung p eines AR-Prozesses ist das Auftreten von Spectral-Line-Splitting insofern von Interesse, als hierbei eine so starke Überschätzung der geeigneten Modellordnung beobachtet werden kann, so daß Spectral-Line-Splitting auftritt (KAY/MARPLE[1979]). Die besondere Eigenheit des Spectral-Line-Splitting liegt darin, daß es gerade nicht bei niedrigen Signal-to-Noise-Ratios (Signal-Rausch-Abständen)²auftritt, sondern bei sehr hohen Signal-to-Noise-Ratios. Dies hat die unerwartete Konsequenz, daß die Schätzung beim Auftreten von Spectral-Line-Splitting dadurch verbessert werden kann, daß man die Ausgangsdaten zusätzlich mit der Reali-

²Für einen harmonischen Prozeß $X_t = A \sin(2\pi t \omega + \varphi) + \epsilon_t$ läßt sich das Signal-to-Noise-Ratio (SNR) angeben als: $\text{SNR} = 10 \log_{10}(A/2\sigma_\epsilon^2)$ [dB].

sation eines White-Noise-Prozesses überlagert. Bestimmte Werte der Phasenverschiebung φ sowie bestimmte Kombinationen der Frequenz ω und der Anzahl der Beobachtungen N begünstigen das Auftreten von Spectral-Line-Splitting, dies gilt ebenso, wenn eine im Verhältnis zur Anzahl der vorliegenden Beobachtungen N relativ hohe Ordnung p des geschätzten AR-Prozesses gewählt wird (vgl. z.B. HAYKIN[1989, S. 206]). Praktische Relevanz besitzt das Problem des Spectral-Line-Splitting z.B. in der Kernresonanz-Spektroskopie, wo eine Verwechslungsgefahr zwischen artifiziell gespaltenen Spektrallinien und physikalisch bedingten sog. Dubletten bzw. Multipletten des geschätzten Spektrums besteht (vgl. SCHMIDT[1989, S. 39ff]).

Peak-Shifting. In Abbildung 3.7 ist anhand des harmonischen Prozesses

$$X_t = \sin(2\pi t 0.2 + \varphi) + \epsilon_t \quad \{\epsilon_t\} \sim \text{WN}(0, 0.0005)$$

mit

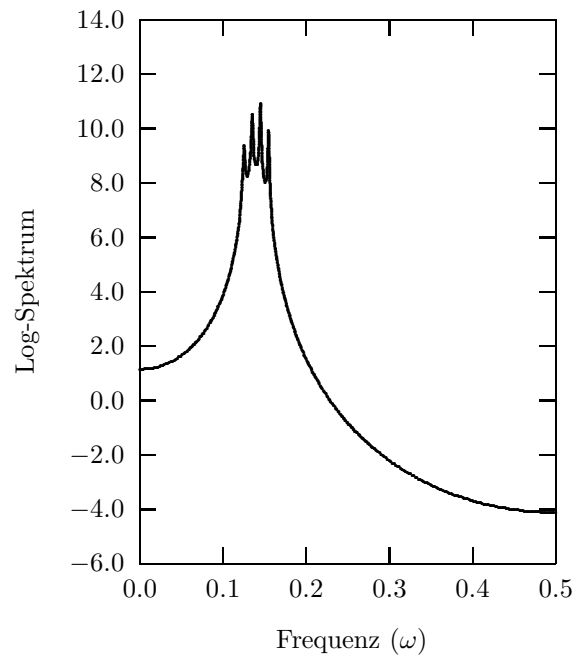
$$t = 1, \dots, 20, \quad (180/\pi)\varphi = -180(4)180$$

die unterschiedliche Phasenabhängigkeit der Yule-Walker-, Burg- und FBLP-Schätzung dokumentiert. Für jede Phase φ wurden 20 unabhängige Realisationen erzeugt, ein AR(2)-Prozeß geschätzt und ein Spektrum im Bereich $0.0 \leq \omega \leq 0.5$ an 512 äquidistanten Punkten berechnet. Die Frequenz mit dem maximalen Peak wurde jeweils ermittelt und eine mittlere Frequenz als Durchschnitt aus 20 Werten berechnet.

In den Abbildungen 3.7 (a) - (d) ist der Zusammenhang zwischen der mittleren Frequenz des maximalen Peaks und der Phase des datenerzeugenden Cosinoidprozesses dargestellt. Es ergibt sich eine stark ausgeprägte Phasenabhängigkeit bei der Yule-Walker- und der Burg-Schätzung. Bei der FBLP- und FLP-Schätzung tritt diese Phasenabhängigkeit nur in stark geminderter Form auf, in dem hier gewählten Beispiel sind die Ergebnisse numerisch identisch.

Zusammenfassung. Wie die Diskussion gezeigt hat, kann die Yule-Walker-Schätzung unter Verwendung der verzerrten Autokovarianzschätzer erhebliche Nachteile aufweisen. Einer Übersicht in HUANG[1990] ist allerdings zu entnehmen, daß bei der Yule-Walker-Methode Vorteile bezüglich der notwendigen Anzahl an Rechenoperationen und des benötigten Speicherbedarfes, insbesondere im Vergleich zum Burg-Algorithmus, bestehen. Da dies nur in Ausnahmefällen von Bedeutung sein wird, sind Burg-, FBLP- oder FLP-Algorithmus im allgemeinen vorzuziehen.

(a) Spektrum



(b) Wurzeln

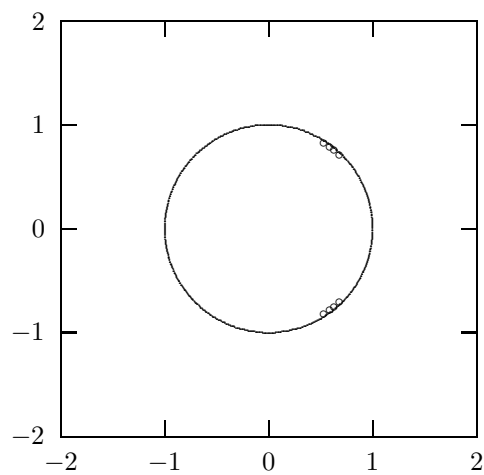


Abb. 3.1: Theoretisches Spektrum und Lage der Wurzeln gegenüber dem Einheitskreis, AR(8)-Prozeß nach (3.70) und (3.71).

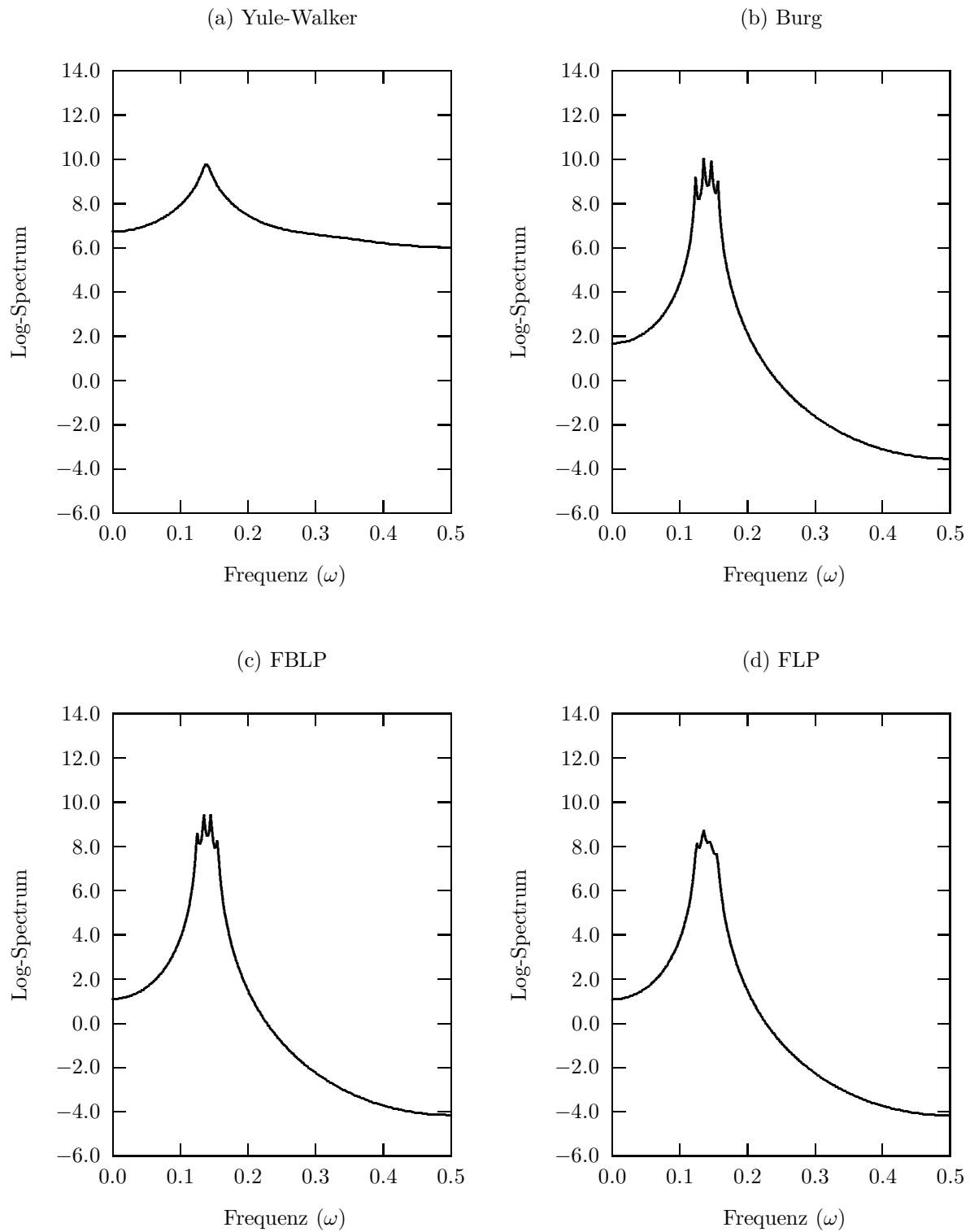


Abb. 3.2: Vergleich von Yule-Walker-, Burg- und FBLP- und FLP-Methode, AR(8)-Prozeß nach (3.70) und (3.71), gemittelte Spektren aus 100 unabhängigen Realisationen, $N = 64$, $N_E = 512$, $N_\omega = 512$.

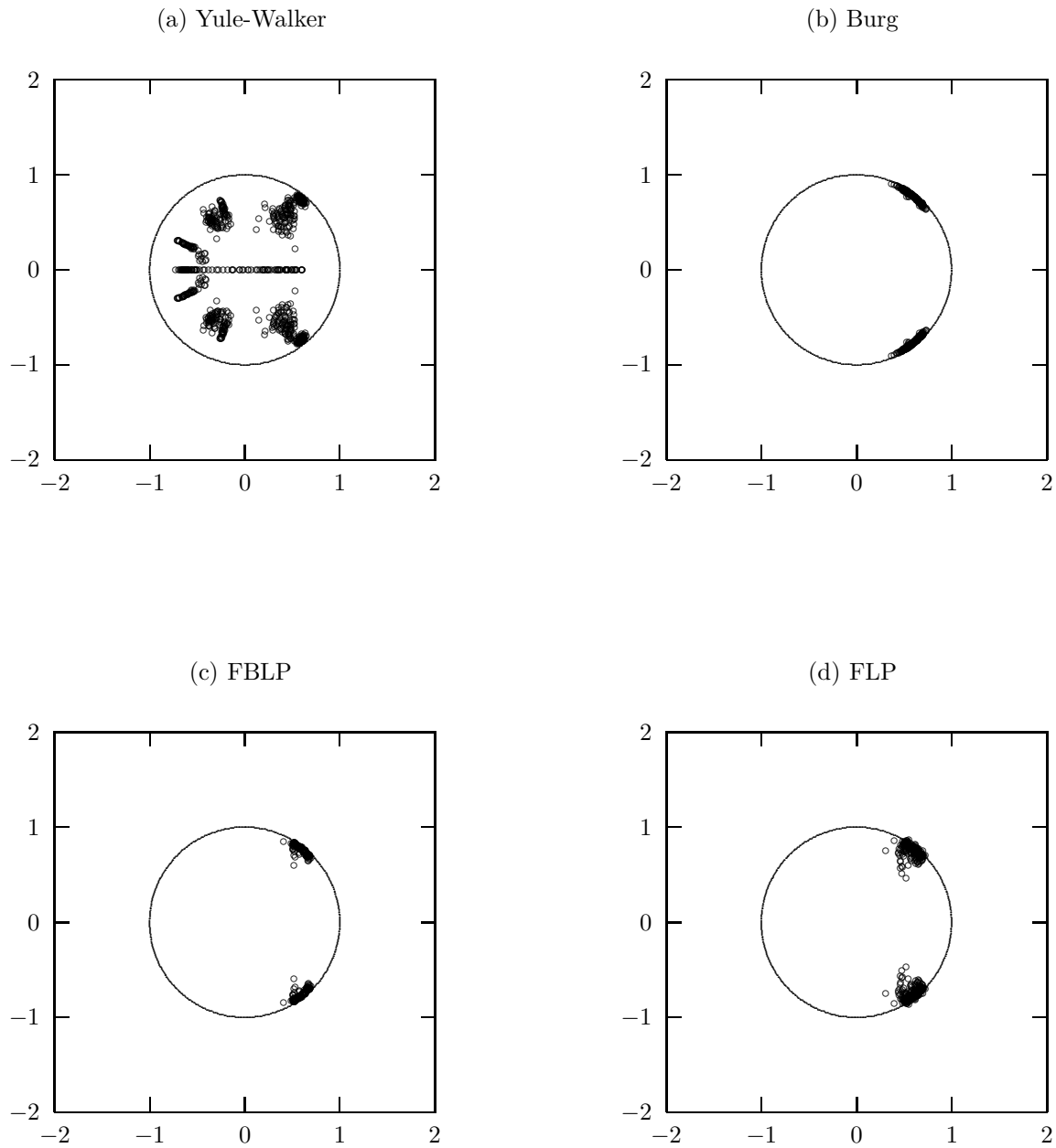


Abb. 3.3: Vergleich von Yule-Walker-, Burg-, FBLP- und FLP-Schätzung: AR(8)-Prozeß nach (3.70) und (3.71), Wurzelplots von 100 unabhängigen Realisationen, $N = 64$, $N_E = 512$.

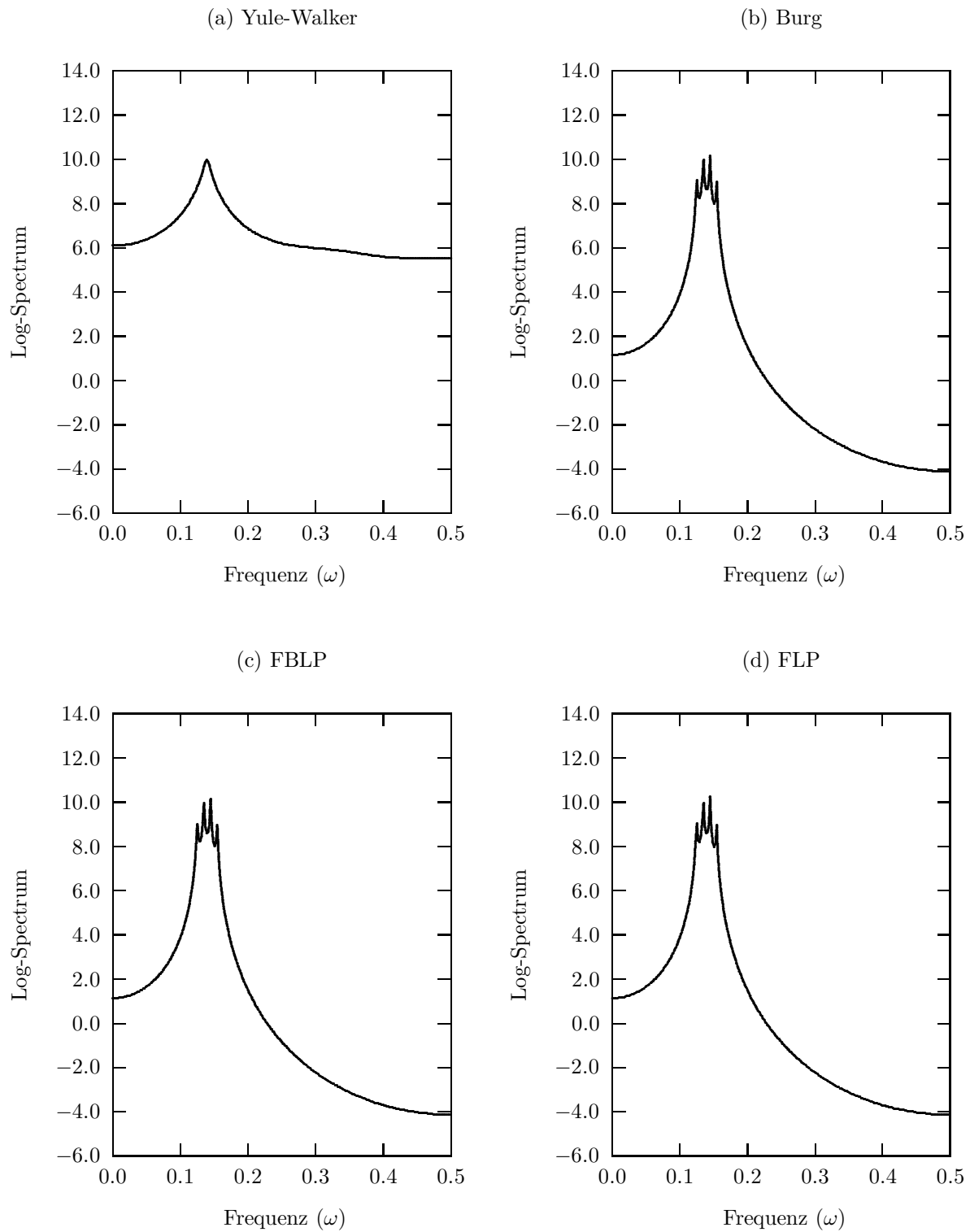


Abb. 3.4: Vergleich von Yule-Walker-, Burg- und FBLP- und FLP-Methode: AR(8)-Prozeß nach (3.70) und (3.71), gemittelte Spektren aus 100 unabhängigen Realisationen, $N = 256$, $N_E = 512$, $N_\omega = 512$.

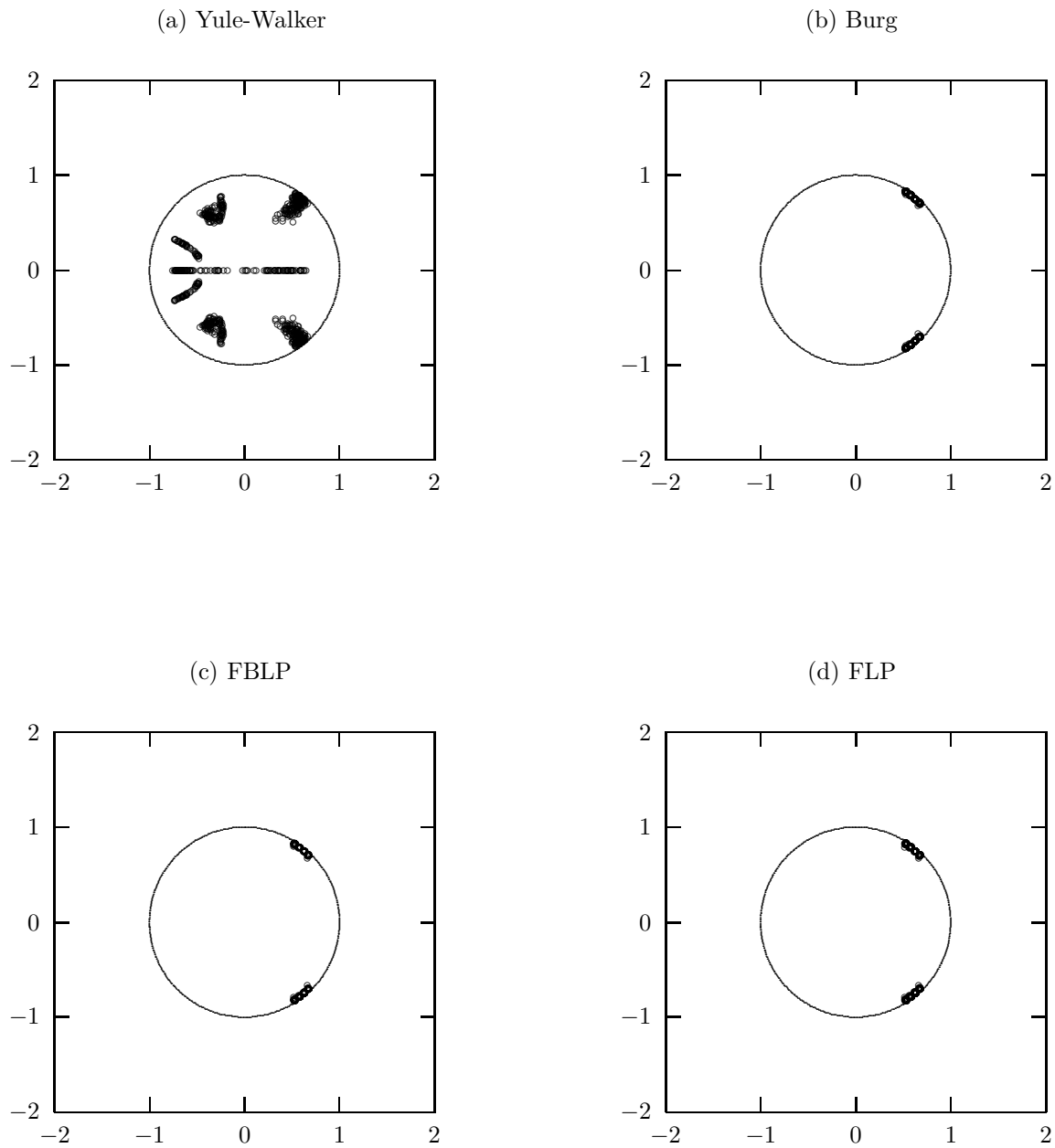


Abb. 3.5: Vergleich von Yule-Walker-, Burg-, FBLP- und FLP-Schätzung: AR(8)-Prozeß nach (3.70) und (3.71), Wurzelplots von 100 unabhängigen Realisationen, $N = 256$, $N_E = 512$.

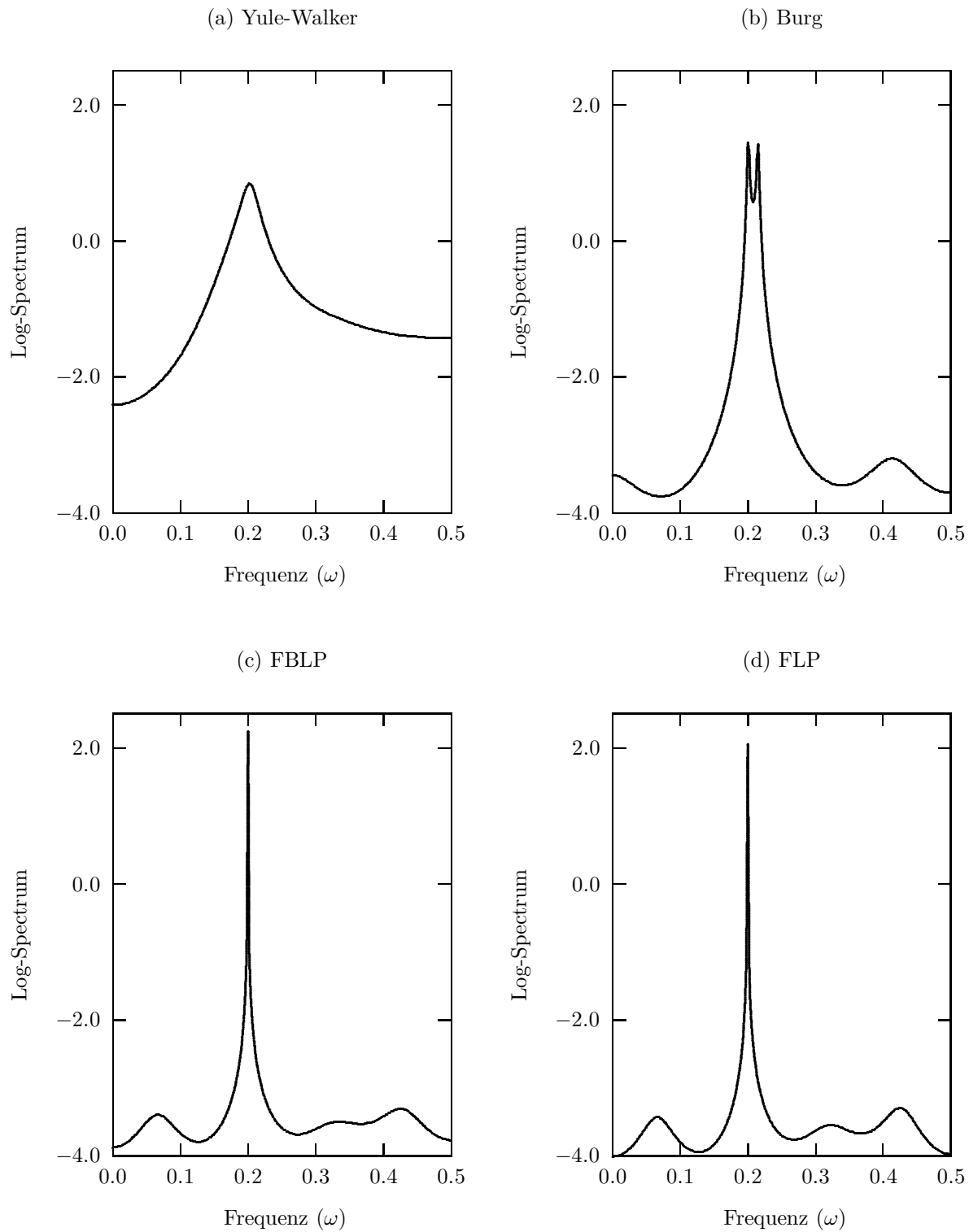


Abb. 3.6: Spectral-Line-Splitting: $X_t = \sin(2\pi t 0.2 + \pi/4) + \epsilon_t$, $\{\epsilon_t\} \sim \text{WN}(0, 0.0005)$ $t = 1, \dots, 20$, Spektren gemittelt aus jeweils 100 autoregressiven Schätzungen mit $p = 8$ nach der Yule-Walker, Burg-, FBLP- und FLP-Methode.

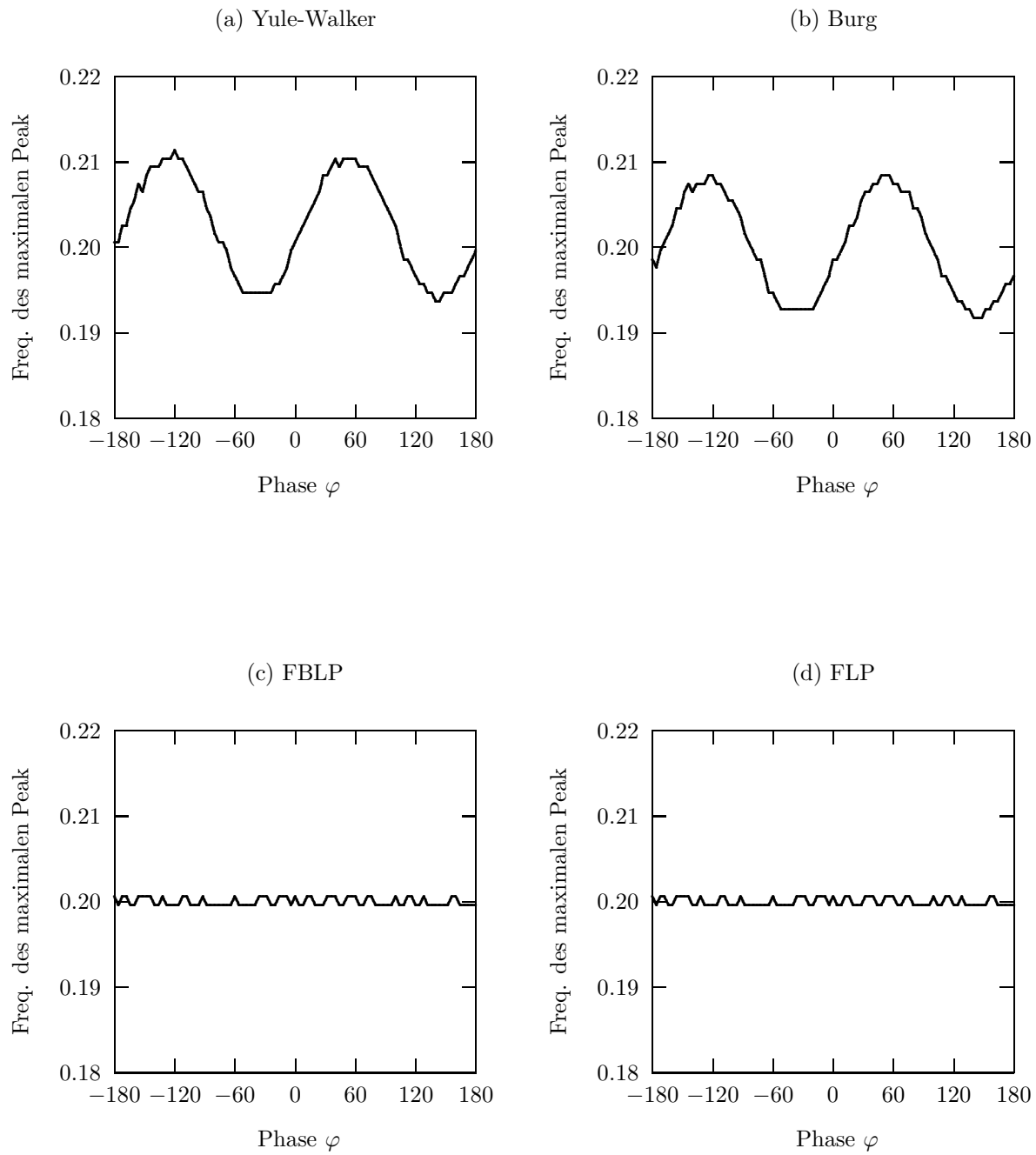


Abb. 3.7: Peak-Shifting: $X_t = \sin(2\pi t 0.2 + \varphi) + \epsilon_t$, $\{\epsilon_t\} \sim \text{WN}(0, 0.0005)$ $t = 1, \dots, 20$, $(180/\pi)\varphi = -180(4)180$, zugrundeliegende Spektren gemittelt aus jeweils 20 autoregressiven Schätzungen mit $p = 2$ nach der Yule-Walker, Burg-, FBLP- und FLP-Methode.

Kapitel 4

Nichtparametrische Spektralschätzung

4.1 Das Periodogramm

Betrachtet man die theoretische Spektraldichte

$$(4.1) \quad f(\omega) = \sum_{\tau=-\infty}^{\infty} \gamma_{\tau} e^{-i2\pi\omega\tau}$$

eines (schwach stationären) stochastischen Prozesses, dann bietet es sich unmittelbar an, die unendliche Fouriertransformation mit den Autokovarianzen γ_{τ} in (4.1) durch eine endliche Fouriertransformation mit den Autokovarianzschätzern $\hat{\gamma}_{\tau}$ zu ersetzen, um eine Schätzung der Spektraldichte $f(\omega)$ zu erhalten:

Def. 4.1.1 (*Periodogramm*)

Sei $\{X_t, t \in \mathcal{Z}\}$ ein stationärer stochastischer Prozeß mit Erwartungswert μ , absolut-summierbarer Autokovarianzfunktion und der Spektraldichte $f(\omega)$. Gegeben sei eine Stichprobe (X_1, \dots, X_N) und die Autokovarianzschätzer $\{\hat{\gamma}_{\tau}, \tau = 0, \pm 1, \dots, \pm(N-1)\}$. Dann heißt

$$(4.2) \quad I_N(\omega) = \sum_{\tau=-(N-1)}^{(N-1)} \hat{\gamma}_{\tau} e^{-i2\pi\omega\tau} \quad \omega \in [-0.5, 0.5]$$

Periodogramm.

■

Alternativ läßt sich (4.2) darstellen als (vgl. z.B. NEWTON[1988, S. 168]):

$$(4.3) \quad I_N(\omega) = \frac{1}{N} \left| \sum_{t=1}^N X_t e^{-i2\pi\omega t} \right|^2.$$

Häufig ist es einfacher, sich bei der Betrachtung des Periodogramms auf die sog. Fourierfrequenzen

$$\{\omega_k | \omega_k = k/N, k = 0, \pm 1, \dots, \omega_k \in [-0.5, 0.5]\}$$

zu beschränken.

Das Periodogramm eines *reellwertigen* stochastischen Prozesses ist symmetrisch, d.h. es gilt

$$(4.4) \quad I_N(\omega) = I_N(-\omega).$$

Die Unterschiede bei der Analyse von reellwertigen bzw. komplexwertigen stochastischen Prozessen bestehen im wesentlichen darin, daß das Periodogramm eines komplexwertigen stochastischen Prozesses nicht symmetrisch ist und an den Frequenzen $\omega = 0.0, \pm 0.5$ andere statistische Eigenschaften besitzt. Um die Darstellung zu vereinfachen, werden bei vielen Aussagen nur die Fourierfrequenzen $\omega_k \neq 0, \pm 0.5$ zugrundegelegt, so daß die Resultate sowohl für reell- als auch für komplexwertige stochastische Prozesse gelten. Es muß allerdings darauf hingewiesen werden, daß spezielle spektralanalytische Fragestellungen existieren, bei denen gerade der Wert eines Spektrums an der Frequenz $\omega = 0$ eine besondere Rolle spielt (vgl. z.B. ANDREWS[1991], WOLTERS[1991], ROBINSON[1991]).

Aufgrund der Symmetrieeigenschaft ist es bei reellwertigen Prozessen ausreichend, Frequenzen im Intervall $[0.0, 0.5]$ zu betrachten.

Unabhängig davon, ob ein reell- oder komplexwertiger Prozeß zugrundeliegt, gilt:

$$(4.5) \quad I_N(\omega + 1) = I_N(\omega),$$

d.h. das Periodogramm besitzt die Periode 1 und kann, sofern erforderlich, periodisch fortgesetzt und damit für alle $\omega \in \mathcal{R}$ definiert werden.

Das empirische Periodogramm $\tilde{I}_N(\omega)$ für die Realisation eines stochastischen Prozesses $\mathbf{x} = (x_1, \dots, x_N)'$ ergibt sich mit den empirischen Autokovarianzen c_τ als:

$$(4.6) \quad \tilde{I}_N(\omega) = \sum_{\tau=-(N-1)}^{(N-1)} c_\tau e^{-i2\pi\omega\tau}.$$

Der Erwartungswert des Periodogramms an den Fourierfrequenzen $\omega_k \neq 0, \pm 0.5$ läßt sich darstellen als (vgl. z.B. Schlittgen/Streitberg[1984, S. 268]):

$$(4.7) \quad E(I_N(\omega_k)) = \int_{-0.5}^{0.5} f(\lambda) F_N(\omega_k - \lambda) d\lambda,$$

wobei $F_N(\cdot)$ der Fejerkern mit $F_N(\lambda) = (1/N) \frac{\sin^2(\pi\lambda N)}{\sin^2(\pi\lambda)}$ ist.

Der Erwartungswert des Periodogramms ergibt sich also als Faltung der theoretischen Spektraldichte mit dem Fejerkern $F_N(\lambda)$, so daß das Periodogramm kein erwartungstreuer Schätzer, sondern ein gewichtetes Mittel der theoretischen Spektraldichte ist. Da der Fejerkern sich für große N wie eine Dirac-Deltafunktion verhält, ist das Periodogramm zumindest asymptotisch erwartungstreu. Diese Aussage und weitere grundlegende asymptotische Eigenschaften des Periodogramms sind unter vereinfachenden Annahmen im folgenden Satz zusammengefaßt (detaillierter z.B. in BROCKWELL/DAVIS[1991, S. 342ff]):

Satz 4.1.1 (*Asymptotische Eigenschaften des Periodogramms*)

Sei $I_N(\omega)$ das Periodogramm eines stationären stochastischen Prozesses $\{X_t, t \in \mathcal{Z}\}$ mit $\mu = 0$, absolut-summierbarer Autokovarianzfunktion und der Spektraldichte $f(\omega)$.

Dann gilt an den Fourierfrequenzen $\omega_k \neq 0, \pm 0.5$:

$$(i) \quad \lim_{N \rightarrow \infty} E[I_N(\omega_k)] = f(\omega_k)$$

$$(ii) \quad \lim_{N \rightarrow \infty} \text{Var}[I_N(\omega_k)] = f^2(\omega_k)$$

$$(iii) \quad \lim_{N \rightarrow \infty} \text{Cov}[I_N(\omega_r), I_N(\omega_s)] = 0 \quad r \neq s.$$

■

Aus Satz 4.1.1 (i) und (ii) folgt, daß das Periodogramm $I_N(\omega_k)$ zwar ein asymptotisch erwartungstreuer Schätzer, aber kein konsistenter Schätzer ist, da die Varianz für $N \rightarrow \infty$ nicht gegen Null, sondern gegen eine von N unabhängige Konstante strebt.

Punkt (iii) aus Satz 4.1.1 gibt einen Hinweis, wie aus dem inkonsistenten Periodogramm konsistente Schätzer abgeleitet werden können: Aufgrund der asymptotischen Unkorreliertheit von $I_N(\omega_r), I_N(\omega_s)$ für $r \neq s$ kann durch Mittelung von benachbarten Werten des inkonsistenten Periodogramms ein Schätzer berechnet werden, dessen Varianz kleiner ist, als die Varianz einzelner Werte des Periodogramms.

Die relevanten Aussagen über die Verteilung des Periodogramms sind im nächsten Satz zusammengefaßt (vgl. z.B. NEWTON[1988, S. 168ff]):

Satz 4.1.2 (*Verteilung des Periodogramms*)

- (i) Sei $X_t = \sum_{j=-\infty}^{\infty} \psi_j \epsilon_{t-j}$ ein stationärer Prozeß mit $\{\epsilon_t\} \sim \text{IID}(0, \sigma_\epsilon^2)$ und $\sum_{j=-\infty}^{\infty} (1 + |j|)|\psi_j| < \infty$.

Dann gilt: An den Fourierfrequenzen sind die Werte des Periodogramms $I_N(\omega_k)$ asymptotisch unabhängig verteilt mit:

$$(4.8) \quad \frac{2I_N(\omega_k)}{f(\omega_k)} \sim \chi_2^2 \quad \omega \neq 0.0, \pm 0.5$$

- (ii) Ist $\{X_t\}$ ein unabhängig und identisch normalverteilter stochastischer Prozeß mit $\mu_X = 0$ und der Varianz σ_X^2 , dann gilt $f(\omega_k) = \sigma_X^2$ und die Verteilungsaussage aus (4.8) ist nicht nur asymptotisch, sondern exakt gültig.

■

Die Bedingung $\sum_{-\infty}^{\infty} (1 + |j|)|\psi_j| < \infty$ ist erforderlich, damit die Koeffizienten ψ_j hinreichend schnell gegen Null konvergieren.

Auf der Grundlage von Satz 4.1.2 können Konfidenzintervalle für nichtparametrische Spektralschätzer abgeleitet werden. Darüberhinaus bildet die Transformation einer Stichprobe (X_1, \dots, X_N) mit stochastisch abhängigen Zufallsvariablen in zumindest asymptotisch unabhängige Periodogrammordinaten den Ausgangspunkt für die Übertragung von statistischen Methoden, die zunächst für unabhängige Stichproben entwickelt worden waren, auf Anwendungen in der Zeitreihenanalyse. Erste Ansätze zur automatischen, datengestützten Bestimmung von Glättungsparametern für die nichtparametrische Spektralschätzung benutzten das Periodogramm als Grundlage zur Implementierung von Kreuzvalidierungsverfahren (vgl. BELTRÃO/BLOOMFIELD[1987],

HURVICH[1985], WHABA[1980]). Einen ähnlichen Ansatz, wobei jedoch der Glättungsparameter für eine nichtparametrische Regression von autokorrelierten Zeitreihendaten zu ermitteln ist, beschreiben HURVICH/ZEGGER[1990].

4.2 Direkte Spektralschätzer

Im Anschluß an Satz 4.1.1 war bereits darauf hingewiesen worden, daß die asymptotische Unkorreliertheit von verschiedenen Werten des Periodogramms es erlaubt, durch Mittelung benachbarter Werte Schätzer der Spektraldichte abzuleiten, deren Varianz geringer ist als die Varianz von einzelnen Periodogrammordinaten. Direkte Spektralschätzer bauen auf diesem Ansatz auf und ermöglichen es, insbesondere wenn das Periodogramm an den Fourierfrequenzen unmittelbar aus den Beobachtungsdaten mit Hilfe einer schnellen Fouriertransformation berechnet wird, auf einfache Art und Weise konsistente Schätzungen zu erhalten.

Def. 4.2.1 (*Direkte Spektralschätzer*)

Sei $\omega_k \in [-0.5, 0.5]$ eine Fourierfrequenz. Dann heißt

$$(4.9) \quad \hat{f}^D(\omega_k) = \sum_r g_r I(\omega_{k-r})$$

direkter Spektralschätzer mit dem Glättungsfilter $\{g_r\}$. ■

Der Glättungsfilter $\{g_r\}$ heißt auch diskretes Periodogrammfenster. Liegt ω_{k-r} nicht im Bereich $[-0.5, 0.5]$, dann wird das Periodogramm entsprechend periodisch erweitert. Bei praktischen Anwendungen ist zu empfehlen, eine leicht modifizierte Form der Schätzung (4.9) durchzuführen. Aus (4.3) folgt, daß gilt:

$$(4.10) \quad I_N(0) = \frac{1}{N} \left| \sum_{t=1}^N X_t \right|^2 = N |\bar{X}|^2.$$

An der Frequenz $\omega_k = 0$ mißt das Periodogramm also nicht den Beitrag einer Schwingungskomponente, sondern den Einfluß des Mittelwertes des zugrundeliegenden stochastischen Prozesses. Wird wie allgemein üblich, das empirische Periodogramm $\tilde{I}_N(\omega_k)$ von einer mittelwertbereinigten Reihe berechnet, dann gilt: $\tilde{I}_N(0) = 0$. Um unerwünschte Verzerrungen zu vermeiden, sollte daher vor Berechnung der direkten Spektralschätzung gemäß (4.9) eine der beiden folgenden Korrekturen für das Periodogramm

an der Frequenz $\omega = 0$ vorgenommen werden (vgl. z.B. FULLER[1976, S. 304], BROCKWELL/DAVIS[1991, S. 353]):

$$(4.11) \quad I_N(0) = \frac{1}{2} [I_N(\omega_1) + I_N(\omega_{-1})]$$

oder

$$(4.12) \quad I_N(0) = g_0 I_N(\omega_1) + \sum_{r=1}^{m_N} g_r I_N(\omega_{r+1}) + \sum_{r=1}^{m_N} g_r I_N(\omega_{-r-1}).$$

Die Korrekturen wurden so umformuliert, daß sie sowohl für den Fall eines symmetrischen Periodogramms (reellwertiger Prozeß) als auch für den Fall eines asymmetrischen Periodogramms (komplexwertiger Prozeß) angewendet werden können.

Direkte Spektralschätzer besitzen u.a. die folgenden asymptotischen Eigenschaften (vgl. z.B. SCHLITGEN/STREITBERG[1984, S. 284]):

Satz 4.2.1 (*Asymptotische Eigenschaften direkter Spektralschätzer*)

Sei $\hat{f}^D(\omega_k)$ ein direkter Spektralschätzer mit dem Glättungsfilter $\{g_r\}$. Dann gilt an den Fourierfrequenzen $\omega_k \neq 0.0, \pm 0.5$:

$$(4.13) \quad \lim_{N \rightarrow \infty} E [\hat{f}^D(\omega_k)] = f(\omega_k) \sum_r g_r$$

$$(4.14) \quad \lim_{N \rightarrow \infty} \text{Var} [\hat{f}^D(\omega_k)] = f^2(\omega_k) \sum_r g_r^2.$$

■

Setzt man $r = -m_N, \dots, m_N$ und trifft für die Zahl m_N und den Glättungsfilter $\{g_r\}$ die folgenden Annahmen:

$$(4.15) \quad m_N \rightarrow \infty \quad \text{und} \quad \frac{m_N}{N} \rightarrow 0 \quad \text{für} \quad N \rightarrow \infty,$$

$$(4.16) \quad g_r = g_{-r}, \quad g_r > 0 \quad \forall r,$$

$$(4.17) \quad \sum_{r=-m_N}^{m_N} g_r = 1,$$

$$(4.18) \quad \sum_{r=-m_N}^{m_N} g_r^2 \rightarrow 0 \quad \text{für } N \rightarrow \infty,$$

dann sind direkte Spektralschätzer konsistente Schätzer der Spektraldichte $f(\omega_k)$ (vgl. z.B. BROCKWELL/DAVIS[1991, S. 351ff]).

Für die Interpretation des Bias eines Spektralschätzers ist es nützlich, die für den Erwartungswert des Periodogramms gefundene Darstellung (4.7) zu verallgemeinern:

Def. 4.2.2 (*Spektralfenster*)

Sei $\hat{f}(\omega_k)$ ein Spektralschätzer, dessen Erwartungswert sich darstellen läßt als:

$$(4.19) \quad E[\hat{f}(\omega_k)] = \int_{-0.5}^{0.5} f(\lambda) S_N(\omega_k - \lambda) d\lambda.$$

Dann heißt $S_N(\lambda)$ Spektralfenster des Spektralschätzers $\hat{f}(\omega_k)$. ■

Für das Periodogramm ist das Spektralfenster $S_N(\lambda)$ mit dem Fejerkern $F_N(\lambda)$ identisch.

Der Erwartungswert eines direkten Spektralschätzers ist im folgenden Satz angegeben (vgl. SCHLITTEGEN/STREITBERG[1984, S. 291]):

Satz 4.2.2 (*Erwartungswert direkter Spektralschätzer*)

Sei $\hat{f}^D(\omega_k)$ ein direkter Spektralschätzer mit dem Glättungsfilter $\{g_k\}$. Dann gilt an den Fourierfrequenzen $\omega_k \neq 0, \pm 0.5$:

$$(4.20) \quad E[\hat{f}^D(\omega_k)] = \int_{-0.5}^{0.5} f(\lambda) S_N(\omega_k - \lambda) d\lambda,$$

mit:

$$(4.21) \quad S_N(\lambda) = \sum_{r=-m_N}^{m_N} g_r F_N(\lambda - \frac{r}{N}).$$

■

Das Phänomen, daß der Schätzwert eines Spektrums an einer bestimmten Frequenz ω auch durch Werte des Spektrums beeinflußt wird, die an möglicherweise weit entfernten Frequenzen liegen, wie es durch das Faltungsintegral (4.20) zum Ausdruck kommt, wird Leakage genannt.

Das Spektralfenster direkter Spektralschätzer und damit deren Bias wird im wesentlichen von zwei Komponenten beeinflußt:

- (i) Über den Fejerkern $F_N(\omega)$ überträgt sich der Bias des Periodogramms auf den Schätzer $\hat{f}^D(\omega)$. Dieser Effekt („Periodogramm-Leakage“, „broad-band“-Bias) ist darauf zurückzuführen, daß nur eine endliche Zahl von Beobachtungen in der Schätzung berücksichtigt wird. Von dem gesamten stochastischen Prozeß, der ja grundsätzlich für alle $t \in \mathcal{Z}$ definiert ist, geht über ein Datenfenster (Taper) mit rechteckiger Gestalt nur eine Teilmenge der möglichen Werte in die Schätzung ein. Eine Verminderung des „broad-band“-Bias kann dadurch erreicht werden, daß der Rechteck-Taper durch einen geeigneteren Taper ersetzt wird. DAHLHAUS[1983] verallgemeinert einige spektralanalytische Aussagen auf Anwendungen mit (Nicht-Rechteck-)Datentapern. Bei der Anwendung eines Tapers ist allerdings zu berücksichtigen, daß die Verminderung des „broad-band“-Bias grundsätzlich mit einem Anstieg der Varianz verbunden ist (vgl. z.B. BRILLINGER[1975, S. 151]). Die Bestimmung eines optimalen Tapers kann wiederum mit objektiven Methoden anhand der gegebenen Daten erfolgen (vgl. HURVICH[1988]). Im Rahmen dieser Arbeit wird auf die Anwendung von Datentapern verzichtet.
- (ii) Die Mittelung mit dem diskreten Periodogrammfenster $\{g_r\}$ über benachbarte Periodogrammwerte verursacht bei nicht konstanter Spektraldichte ebenfalls einen Bias („Window-Leakage“, „narrow-band“-Bias). Diese Komponente kann durch die Wahl der Filtergewichte $\{g_r\}$ beeinflusst werden.

Geht man von den asymptotischen Aussagen des Satzes 4.2.1 aus und minimiert für fest vorgegebenes $M = 2m_N + 1$ den Faktor $\sum_{|r| \leq m_N} g_r^2$ in (4.14) unter der Nebenbedingung der asymptotischen Erwartungstreue, $\sum_{|r| \leq m_N} g_r = 1$, so folgt aus dem Lagrangeansatz

$$(4.22) \quad \min_{\mathbf{g}, \lambda} \mathcal{L}(\mathbf{g}, \lambda) = \mathbf{g}'\mathbf{g} - \lambda(\mathbf{g}'\mathbf{1} - 1)$$

mit $\mathbf{g} = (g_{-m_N}, \dots, g_0, \dots, g_{m_N})'$ und $\mathbf{1} = (1, \dots, 1)'$, daß für einen varianzminimalen Schätzer gelten muß:

$$g_r = \frac{1}{M} \quad \forall r.$$

Dieser Schätzer mit gleichgewichtetem Glättungsfilter $\{g_r\}$ heißt direkter Daniell-Schätzer:

$$(4.23) \quad \hat{f}^{Dan}(\omega_k) = \frac{1}{M} \sum_{r=-m_N}^{m_N} I(\omega_{k-r}).$$

Die asymptotische Varianz des Daniell-Schätzers ergibt sich als

$$\text{Var}[\hat{f}^{Dan}(\omega_k)] \doteq \frac{1}{M} f^2(\omega_k).$$

Der Parameter M des Daniell-Schätzers, entscheidet über die Güte der Schätzung: Für $M = 1$ bzw. $m_N = 0$ ergibt sich das Periodogramm, das aufgrund der hohen Varianz instabil ist. Andererseits ist es mit dem Periodogramm möglich, unterschiedliche Peaks im Spektrum zu lokalisieren, die einen Frequenzabstand von mindestens $1/N$ besitzen. Größere Werte für M stabilisieren die Schätzung und reduzieren die Varianz. Gleichzeitig verursacht die Glättung aber einen Bias, der das Auflösungsvermögen des Spektralschätzers herabsetzt, so daß nur noch solche unterschiedlichen Peaks differenziert werden können, die einen Frequenzabstand von mindestens M/N besitzen. Es bietet sich für den direkten Daniell-Schätzer an, das Auflösungsvermögen bzw. den („narrow-band“) Bias durch die sog. Bandbreite M/N zu messen, die angibt, welcher Anteil der Fourierfrequenzen in die Schätzung eingeht. Der Trade-Off zwischen Bandbreite und Varianz des direkten Daniell-Schätzers läßt sich darstellen als:

$$(4.24) \quad \frac{M}{N} \times \frac{1}{M} f^2(\omega_k) = \frac{1}{N} f^2(\omega_k),$$

bzw.

$$(4.25) \quad \text{Bandbreite} \times \text{Varianz} = \text{konstant.}$$

Für gegebenen Stichprobenumfang N führt eine Verminderung des Bias, was die Wahl eines kleineren Wertes für M erforderlich macht, zwangsläufig zu einem Anstieg der Varianz, und umgekehrt.

Da sich die Beziehung (4.25) auf analoge Weise auf alle Arten von Spektraldichteschätzungen übertragen läßt, wird sie auch als *Unschärferelation der Spektralschätzung* bezeichnet.

Da sich Varianz und Bias nicht gleichzeitig minimieren lassen, muß versucht werden, M so festzulegen, daß ein aggregiertes Kriterium, in das beide Größen eingehen, wie etwa der mittlere quadratische Fehler

$$\text{MSE}(M) = E [\hat{f}(\omega) - f(\omega)]^2 = \text{Var} [\hat{f}(\omega)] + [E[\hat{f}(\omega)] - f(\omega)]^2,$$

minimiert wird. Unter der Annahme, daß der optimale Glättungsparameter M für jede Frequenz ω gleich ist, kann ein globales Kriterium, wie etwa der integrierte mittlere quadratische Fehler

$$\text{MISE}(M) = \int_{-0.5}^{0.5} E [\hat{f}(\omega) - f(\omega)]^2 d\omega,$$

als Auswahlkriterium herangezogen werden. Unglücklicherweise hängen nicht nur die beiden genannten Kriterien, sondern grundsätzlich jedes Auswahlkriterium von der wahren aber unbekanntem Spektraldichtefunktion $f(\omega)$ ab. Eine Lösung dieses Dilemmas bieten Verfahren, die es erlauben, mit Hilfe von datengestützten Kriterien einen zumindest „approximativ“ optimalen Parameter M zu finden. In Kapitel 5 wird gezeigt, daß der integrierte quadratische Fehler (ISE) nicht notwendigerweise ein sinnvolles Abstandsmaß für Spektren ist. Im allgemeinen wird ein Abstandsmaß aus der in Kapitel 5 neu definierten Klasse der global korrigierten Diskrepanzmaße sinnvollere Ergebnisse liefern. Theoretische Aussagen über statistische Eigenschaften von global korrigierten Diskrepanzmaßen liegen allerdings noch nicht vor.

Wird kein Verfahren angewandt, um den Glättungsparameter M anhand der vorliegenden Daten zu ermitteln, so ist man darauf angewiesen, durch Ausprobieren von verschiedenen Werten von M eine akzeptable Spektralschätzung zu ermitteln. Allerdings ist der direkte Daniellschätzer aufgrund seiner einfachen Struktur als theoretischer Referenzpunkt bedeutender als für praktische Spektraldichteschätzungen. Die Gleichgewichtung der Werte an den Nachbarfrequenzen kann bei nicht konstant verlaufendem Spektrum einen erheblichen Bias bewirken. Im allgemeinen sind Glättungsfilter, die an den Rändern weniger stark gewichtet werden, vorzuziehen. BLOOMFIELD[1976, S. 173ff] verwendet z.B. modifizierte Daniell-Schätzer, die eine gerade Anzahl von Gewichtungskoeffizienten besitzen, wobei die beiden äußersten Randgewichte die halbe Gewichtung der inneren Gewichte bekommen. Dieser modifizierte Daniell-Schätzer wird gegebenenfalls mehrmals iterativ auf ein Periodogramm angewandt. Eine einfache Alternative für praktische Anwendungen, insbesondere wenn interaktiv am Bildschirm gearbeitet werden kann, ist wie folgt:

Man wählt einen direkten Spektralschätzer, z.B. den direkten von-Hann-Schätzer¹

¹Wie TUKEY[1967, S. 27] berichtet, wurden diese Gewichte spontan von R. HAMMING für die Glättung von Spektren vorgeschlagen. Sie sind auf den österreichischen Meteorologen JULIUS VON HANN zurückzuführen, der diese zur Datenglättung eingesetzt hatte. Da kein einheitlicher Sprachgebrauch besteht, werden in dieser Arbeit die Begriffe iterative Hanning-Schätzung bzw. iterative Hamming-Schätzung synonym gebraucht.

$$\{g_r\}^{Hann} = \{1/4, 1/2, 1/4\}$$

und wendet diesen *iterativ* auf ein vorliegendes Periodogramm an.

Die Filtergewichte für den Hanning-Schätzer nach z.B. fünf Iterationen ergeben sich als fünffache Faltung von $\{g_r\}^{Hann}$ mit sich selbst:

$$(4.26) \quad \{g_r\}^{Hann} \star \{g_r\}^{Hann} \star \{g_r\}^{Hann} \star \{g_r\}^{Hann} \star \{g_r\}^{Hann} \\ = \left\{ \frac{1}{1024}, \frac{10}{1024}, \frac{45}{1024}, \frac{120}{1024}, \frac{210}{1024}, \frac{252}{1024}, \frac{210}{1024}, \frac{120}{1024}, \frac{45}{1024}, \frac{10}{1024}, \frac{1}{1024} \right\}.$$

Bei dem Filter (4.26) handelt es sich nicht um Rechteck-Filter, vielmehr fällt er von einem Maximum bei $r = 0$ zu den Rändern hin ab. Da es sich nicht um einen gleichgewichteten Filter handelt, ist es nicht ohne weiteres möglich, eine sinnvolle Größe für seine Bandweite anzugeben. Von den zahlreichen Möglichkeiten, Bandweiten für beliebige Filter zu definieren (detailliert z.B. in PRIESTLEY[1981, S. 513ff]) wird hier die einfachste Art gewählt: Einem beliebigen direkten Spektralschätzer wird als sog. äquivalente Bandweite (EBW) die Bandweite des direkten Daniell-Schätzers zugeordnet, der die gleiche Varianz besitzt. Nach diesem Ansatz ergibt sich die äquivalente Bandweite für einen beliebigen direkten Spektralschätzer $\hat{f}^D(\omega)$ als:

$$(4.27) \quad \text{EBW} [\hat{f}^D(\omega)] = \frac{1}{N \sum_{|r| \leq m_N} g_r^2}.$$

Für den Filter (4.26) ergibt sich die äquivalente Bandbreite z.B. als $5.56/N$.

Bei indirekten Spektralschätzern, die von einer Fouriertransformation der gewichteten Autokovarianzfunktion ausgehen, wird die Vorgehensweise, daß sukzessiv verschiedene Spektralschätzer eines Datensatzes mit immer kleiner werdender Bandweite betrachtet werden als „window-closing“ bezeichnet. Dementsprechend kann die hier vorgeschlagene Vorgehensweise für direkte Spektralschätzer als „window-opening“ bezeichnet werden, da die Bandweite der Schätzung bei jeder Iteration zunimmt. Die Bandweite wird bei der iterativen direkten Spektralschätzung indirekt über die gewählte

Anzahl der Iterationen festgelegt. Im allgemeinen ist der Bestimmung einer „optimalen“ Bandweite eine größere Bedeutung zuzumessen als der Auswahl eines möglichst gut auf einen Anwendungsfall zugeschnittenen Glättungsfilters bzw. Periodogrammfensters („window-carpentry“). Bei einigen Anwendungen haben sich jedoch spezielle Glättungsfiler als besonderes geeignet erwiesen: Die sog. Savitzky-Golay-Filer spielen z.B. in der chemischen Spektroskopie eine wichtige Rolle (vgl. z.B. PRESS ET AL.[1992, S. 650ff]).

Aus Satz 4.1.2 folgt, daß direkte Spektralschätzer als gewichtete Summe von χ^2 -verteilten Zufallsvariablen interpretiert werden können. Die Approximation von Erwartungswert- und Varianz dieser gewichteten Summe durch das c -fache einer χ^2 -verteilten Zufallsvariable mit ν Freiheitsgraden mit gleichem Erwartungswert und gleicher Varianz führt über den Ansatz (vgl. z.B. BROCKWELL/DAVIS[1991, S. 362]):

$$(4.28) \quad c\nu = f(\omega_k) \sum_{r=-m_N}^{m_N} g_r$$

$$(4.29) \quad 2c^2\nu = f^2(\omega_k) \sum_{r=-m_N}^{m_N} g_r^2$$

zu der approximativen Verteilung:

$$(4.30) \quad \frac{\nu \hat{f}^D(\omega_k)}{f(\omega_k)} \sim \chi_\nu^2 \quad \omega_k \neq 0.0, \pm 0.5$$

wobei ν berechnet wird als:

$$(4.31) \quad \nu = \frac{2}{\sum_{r=-m_N}^{m_N} g_r^2}.$$

Die Zahl ν heißt äquivalente Zahl von Freiheitsgraden (Equivalent Degrees of Freedom, EDF).

Ein $100(1-\alpha)\%$ -Konfidenzintervall für $\hat{f}^D(\omega_k)$ an einer Fourierfrequenz $\omega_k \neq 0.0, \pm 0.5$ läßt sich damit angeben als (vgl. z.B. BROCKWELL/DAVIS[1991, S. 362]):

$$(4.32) \quad \text{Prob} \left(\frac{\nu \hat{f}^D(\omega_k)}{\chi_{\nu, 1-\alpha/2}^2} \leq f(\omega_k) \leq \frac{\nu \hat{f}^D(\omega_k)}{\chi_{\nu, \alpha/2}^2} \right) = 1 - \alpha.$$

Durch Logarithmieren von (4.32) ergeben sich die Grenzen des Konfidenzintervalls für den Spektralschätzer $\log \hat{f}^D(\omega_k)$ als:

$$(4.33) \quad \left(\log \hat{f}^D(\omega_k) + \log \nu - \log \chi_{\nu, 1-\alpha/2}^2, \log \hat{f}^D(\omega_k) + \log \nu - \log \chi_{\nu, \alpha/2}^2 \right).$$

In der logarithmierten Form (4.33) besitzen die Konfidenzintervalle, unabhängig von der Fourierfrequenz $\omega_k \neq 0, \pm 0.5$, die gleiche Länge, was für die Darstellung nach (4.32) nicht gilt. Dies ist mit ein Grund, bei graphischen Darstellungen des geschätzten Spektrums eine (halb)logarithmierte Form zu bevorzugen.

Bei praktischen Anwendungen wird das Periodogramm häufig mit Hilfe einer schnellen Fouriertransformation (FFT) für $N' > N$ berechnet, wobei N' eine Potenz zur Basis 2 ist und die mittelwertbereinigten Originaldaten mit $N' - N$ Nullen ergänzt werden (Zero-Padding). Für die Schätzung ist allerdings nur die Zahl der Fourierfrequenzen maßgeblich, die in die Mittelung eingehen, so daß in diesem Fall die äquivalente Bandbreite korrigiert werden muß zu (vgl. SCHLITZGEN/STREITBERG[1984, S. 303]):

$$(4.34) \quad \nu = \frac{N'}{N} \frac{2}{\sum_{r=-m_N}^{m_N} g_r^2}.$$

4.3 Indirekte Spektralschätzer

Indirekte Spektralschätzer basieren auf einer geeignet gewichteten Autokovarianzfunktion. Sie repräsentieren das traditionelle Standardinstrument der moderneren Spektralanalyse. Dementsprechend umfangreich und differenziert erfolgte ihre Aufarbeitung in der Literatur. In dem hier gesetzten Rahmen beschränkt sich die Darstellung zunächst auf die Angabe elementarer Eigenschaften von allgemeinen indirekten Spektralschätzern, anschließend werden einige Spektralschätzer aus der Klasse der indirekten Spektralschätzer mit Lagfenster in Skalenparameterform vorgestellt. Für diese Klasse indirekter Spektralschätzer konnte PARZEN[1957] spezifische asymptotische Aussagen ableiten.

Def. 4.3.1 (*Indirekte-Spektralschätzer*)

Ein Spektralschätzer der Form

$$\begin{aligned}
 \hat{f}^{In}(\omega) &= \sum_{\tau=-T}^T k_{T,\tau} \hat{\gamma}_\tau e^{-i2\pi\omega\tau} \\
 (4.35) \qquad &= \int_{-0.5}^{0.5} I_N(\lambda) K_T(\omega - \lambda) d\lambda
 \end{aligned}$$

mit $k_{T,\tau} = 0$ für $\tau > T$, $T \leq (N - 1)$ sowie

$$K_T(\lambda) = \sum_{\tau=-T}^T k_{T,\tau} e^{-i2\pi\omega\tau}$$

heißt *indirekter Spektralschätzer mit dem Lagfenster $\{k_{T,\tau}\}$, dem Periodogrammfenster $K_T(\lambda)$ und dem Stützpunkt (Truncation-Point) T .*

■

Wie das Faltungsintegral (4.35) zeigt, sind indirekte Spektralschätzer im Vergleich zu direkten Spektralschätzern kein grundsätzlich anderer Ansatz der Spektralschätzung: Das Ergebnis der indirekten Spektralschätzung ergibt sich wieder als ein gewichtetes Mittel des Periodogramms. Die Gewichtungsfunktion ist allerdings diesmal das (kontinuierliche) Periodogrammfenster² $K_T(\lambda)$, das sich als Fouriertransformierte des Lagfensters $\{k_{T,\tau}\}$ ergibt.

²Der Begriff Periodogrammfenster wird hier in Abgrenzung zu dem Begriff Spektralfenster aus der Definition 4.2.2 gebraucht. In der Literatur wird auch häufig der Begriff Spektralfenster für die Fouriertransformierte des Lagfensters benutzt.

Analog wie bei den direkten Spektralschätzern ergeben sich konsistente Spektralschätzungen, wenn der Stützpunkt T in Abhängigkeit vom Stichprobenumfang N so festgelegt wird, daß gilt (vgl. z.B. PRIESTLEY[1981, S. 433]):

$$T \rightarrow \infty \quad \text{und} \quad T/N \rightarrow 0 \quad \text{für} \quad N \rightarrow \infty.$$

Viele der für die direkten Spektralschätzer getroffenen Aussagen lassen sich mit geringen Modifikationen übertragen, indem indirekte Spektralschätzer durch direkte Spektralschätzer approximiert werden. Für die Approximation durch direkte Spektralschätzer ist es einfacher das Faltungsintegral (4.35) in äquivalenter Form darzustellen als:

$$(4.36) \quad \hat{f}^{In}(\omega) = \int_{-0.5}^{0.5} K_T(\lambda) I(\omega - \lambda) d\lambda.$$

Approximiert man nun das Integral (4.36) durch eine Riemann-Summe mit $\Delta\lambda = 1/N$, dann gilt:

$$(4.37) \quad \hat{f}^{In}(\omega) \approx \sum_r \frac{1}{N} K_T\left(\frac{r}{N}\right) I\left(\omega - \frac{r}{N}\right),$$

d.h. die entsprechende approximative diskrete Gewichtsfunktion ergibt sich als:

$$g_r = \frac{1}{N} K_T\left(\frac{r}{N}\right).$$

Mit Hilfe der Approximation durch direkte Spektralschätzer lassen sich u.a. die folgenden Aussagen ableiten (vgl. z.B. SCHLITGEN/STREITBERG[1984, S. 314ff]):

Satz 4.3.1 (*Asymptotische Eigenschaften indirekter Spektralschätzer*)

Sei $\hat{f}^{In}(\omega)$ ein indirekter Spektralschätzer mit dem Lagfenster $\{k_{T,\tau}\}$ und dem Periodogrammfenster $K_T(\lambda)$. Dann gilt an den Fourierfrequenzen $\omega_k \neq 0.0, \pm 0.5$:

$$(4.38) \quad \lim_{N \rightarrow \infty} E [\hat{f}^{In}(\omega_k)] = f(\omega_k) \int_{-0.5}^{0.5} K_T(\lambda) d\lambda$$

$$(4.39) \quad \lim_{N \rightarrow \infty} \text{Var} [\hat{f}^{In}(\omega_k)] = f^2(\omega_k) \frac{1}{N} \int_{-0.5}^{0.5} K_T^2(\lambda) d\lambda.$$

■

Die äquivalente Bandbreite für indirekte Spektralschätzer läßt sich definieren als:

Def. 4.3.2 (*Äquivalente Bandbreite indirekter Schätzer*)

Sei $\hat{f}^{In}(\omega)$ ein indirekter Spektralschätzer mit dem Periodogrammfenster $K_T(\lambda)$. Dann heißt

$$(4.40) \quad \text{EBW} = \frac{1}{\int_{-0.5}^{0.5} K_T^2(\lambda) d\lambda}$$

äquivalente Bandbreite von $\hat{f}^{In}(\omega)$.

■

Selbstverständlich gilt auch hier wieder die Unschärferelation der Spektralschätzung,

$$\frac{1}{\int_{-0.5}^{0.5} K_T^2(\lambda) d\lambda} f^2(\omega) \frac{1}{N} \int_{-0.5}^{0.5} K_T^2(\lambda) d\lambda = f^2(\omega) \frac{1}{N},$$

d.h. das Produkt aus Bandweite und Varianz ist bei gegebenem Stichprobenumfang N konstant.

Für die Analyse des Bias indirekter Spektralschätzer ist die Betrachtung des Erwartungswertes bei endlichem Stichprobenumfang von Interesse (vgl. z.B. SCHLITTEGEN/STREITBERG[1984, S. 305]):

Satz 4.3.2 (*Erwartungswert indirekter Spektralschätzer*)

Sei $\hat{f}^{In}(\omega)$ ein indirekter Spektralschätzer mit dem Periodogrammfenster $K_T(\lambda)$. Dann gilt an den Fourierfrequenzen $\omega_k \neq 0.0 \pm 0.5$:

$$(4.41) \quad E[\hat{f}^{In}(\omega_k)] = \int_{-0.5}^{0.5} f(\lambda) S_N(\omega_k - \lambda) d\lambda,$$

mit:

$$(4.42) \quad S_N(\omega_k) = \int_{-0.5}^{0.5} F_N(\lambda) K_T(\omega_k - \lambda) d\lambda.$$

■

Das Spektralfenster indirekter Spektralschätzer ergibt sich also als Faltung des Fejerkerns $F_N(\cdot)$ mit dem Periodogrammfenster $K_T(\cdot)$.

Die Problematik, ein geeignetes Lagfenster $\{k_{T,\tau}\}$ zu ermitteln, so daß das korrespondierende Periodogrammfenster und damit auch das Spektralfenster „optimale“ Eigenschaften besitzt, ist in der Literatur sehr häufig aufgegriffen worden und hat, gegebenenfalls unter Verwendung unterschiedlicher Optimalitätskriterien, zu einer sehr großen Zahl von im Detail verschiedenen Antworten geführt (vgl. z.B. PRIESTLEY[1981, S. 437ff], NEWTON[1988, S. 180ff]). Im weiteren beschränkt sich die Darstellung deshalb auf die Klasse der indirekten Spektralschätzer in Skalenparameterform, was einfachere und speziellere Aussagen zuläßt.

Für die Klasse der indirekten Spektralschätzer mit Lagfenstern in Skalenparameterform entwickelte PARZEN[1957] eine spezifische asymptotische Theorie für Erwartungswert, Varianz und mittleren quadratischen Fehler. Da so gut wie alle relevanten indirekten Spektralschätzer zu dieser Klasse gehören, werden die wesentlichen Aussagen hier kurz skizziert. Die Ausführungen hier folgen im wesentlichen ANDERSON[1971, S. 522ff], NEWTON[1988, S. 183ff] und PRIESTLEY [1981, S. 458ff].

Def. 4.3.3 (*Lagfenster in Skalenparameterform*)

Ein Lagfenster $\{k_{T,\tau}\}$, das sich darstellen läßt als:

$$(4.43) \quad k_{T,\tau} = \mathcal{L}\left(\frac{\tau}{T}\right) = \mathcal{L}(u) \quad \text{mit } T \leq (N - 1)$$

heißt Lagfenster in Skalenparameterform mit dem Skalenparameter T . Die Funktion $\mathcal{L}(u)$, wird als stetige Funktion interpretiert, für die gelten soll:

$$(i) \quad \mathcal{L}(0) = 1$$

$$(ii) \quad \mathcal{L}(u) = \mathcal{L}(-u)$$

$$(iii) \quad \int_{-\infty}^{\infty} \mathcal{L}^2(u) du < \infty.$$

Die Funktion $\mathcal{L}(u)$ heißt Lagfenstergenerator.

■

Der Stützpunkt T wird hier als Skalenparameter interpretiert, der analog wie z.B. die Standardabweichung in der Dichtefunktion der Normalverteilung, den Lagfenstergenerator weitet bzw. zusammenzieht. Für Lagfenstergeneratoren vom sog. „gestützten“ Typ gilt $\mathcal{L}(u) = 0$ für $|u| > 1$.

Für Spektralschätzer auf der Grundlage von Lagfenstern in Skalenparameterform ist es möglich, asymptotische Aussagen über Varianz und Bias der Schätzungen zu machen, die an Kenngrößen des zugehörigen Lagfenstergenerators und an ein Maß für die Glätte der zugrundeliegenden Spektraldichte geknüpft sind.

Ein Lagfenstergenerator wird durch seinen charakteristischen Exponenten und seinen charakteristischen Wert beschrieben:

Def. 4.3.4 (*Charakteristischer Exponent und Wert*)

Der charakteristische Exponent q eines Lagfenstergenerators $\mathcal{L}(u)$ ist der größte ganzzahlige Wert $q > 0$, so daß

$$\lim_{u \rightarrow 0} \frac{1 - \mathcal{L}(u)}{|u|^q} = \mathcal{L}^{(q)}$$

existiert und von Null verschieden ist.

Wenn der Grenzwert für alle ganzen Zahlen q endlich ist, dann wird der charakteristische Exponent gleich unendlich gesetzt.

■

$\mathcal{L}(u)$ läßt sich in der Nähe von Null approximieren durch:

$$\mathcal{L}(u) \approx 1 - \mathcal{L}^{(q)}|u|^q.$$

Der charakteristische Exponent und der charakteristische Wert beschreiben die „Bandweite“ eines Lagfenstergenerators. Je größer q , um so langsamer fällt die Funktion $\mathcal{L}(u)$ für $\mathcal{L}^{(q)} > 0$ in der Umgebung von Null ab. Für gegebenes q fällt $\mathcal{L}(u)$ um so stärker ab, je größer $\mathcal{L}^{(q)}$ ist.

Eine in Hinblick auf die gewünschten asymptotischen Aussagen erforderliche Charakterisierung der Glätte einer Spektraldichte geschieht mit der folgenden Definition:

Def. 4.3.5 (*p-Differenzierbarkeit*)

Eine Spektraldichte $f(\omega)$ heißt *p-differenzierbar* mit *p beschränkten und stetigen Ableitungen*, wenn gilt:

$$\sum_{\tau=-\infty}^{\infty} |\tau|^p |\gamma_\tau| < \infty.$$

Die Funktion

$$f^{[p]}(\omega) = \sum_{\tau=-\infty}^{\infty} |\tau|^p |\gamma_\tau| e^{-i2\pi\omega\tau}$$

heißt verallgemeinerte *p-te* Ableitung von $f(\omega)$.

■

Allgemein gilt für $p > 0$:

$$\sum_{\tau=-\infty}^{\infty} |\tau|^p |\gamma_\tau| < \infty \implies \sum_{\tau=-\infty}^{\infty} |\gamma_\tau| < \infty.$$

Mit den obigen Definitionen läßt sich unter anderem der folgende Satz für den asymptotischen mittleren quadratischen Fehler von indirekten Spektralschätzern in skalenparametrischer Form angeben (vgl. z.B. NEWTON[1986, S. 190ff]):

Satz 4.3.3 (*Asymptotischer MSE indirekter Spektralschätzer*)

Sei $(X_1, \dots, X_N)'$ eine Stichprobe eines stochastischen Prozesses mit der Spektraldichte $f(\omega)$. $f(\omega)$ sei p -differenzierbar für ein $p \geq 1$, $\mathcal{L}(u)$ sei ein Lagfenstergenerator mit einem charakteristischen Exponenten $q > 0$ mit $p \geq q$ und dem charakteristischen Wert $\mathcal{L}^{(q)}$. Für den indirekten Spektralschätzer

$$\hat{f}^{In}(\omega) = \sum_{\tau=-T}^T \mathcal{L}\left(\frac{\tau}{T}\right) \hat{\gamma}_\tau e^{-i2\pi\omega\tau} \quad T \leq (N-1)$$

werde T so gewählt, daß $T \rightarrow \infty$ und $T/N \rightarrow 0$ für $N \rightarrow \infty$. Dann gilt (mit weiteren Regularitätsbedingungen):

$$\begin{aligned} \text{MSE}(\hat{f}^{In}(\omega)) &= \text{Var}[\hat{f}^{In}(\omega)] + [E[\hat{f}^{In}(\omega)] - f(\omega)]^2 \\ (4.44) \quad &\doteq \frac{T}{N} f^2(\omega) \int_{-\infty}^{\infty} \mathcal{L}^2(u) du + \frac{1}{T^{2q}} \left(\mathcal{L}^{(q)} f^{[q]}(\omega) \right)^2. \end{aligned}$$

Eine gleiche Konvergenzrate gegen Null für beide Terme erhält man, wenn T gewählt wird als $T = cN^{1/(2q+1)}$, wobei c eine beliebige Konstante ist. In diesem Fall gilt:

$$(4.45) \quad \lim_{N \rightarrow \infty} N^{2q/(2q+1)} \text{MSE}(\hat{f}^{In}(\omega)) = c f^2(\omega) \int_{-\infty}^{\infty} \mathcal{L}^2(u) du + \frac{1}{c^{2q}} \left(\mathcal{L}^{(q)} f^{[q]}(\omega) \right)^2$$

■

Zu den Annahmen und Aussagen von Satz 4.3.3 sind einige Erläuterungen erforderlich:

- (i) Die Bedingung $p \geq q$ soll sicherstellen, daß $\mathcal{L}(u)$ zumindest genauso schnell gegen Null konvergiert, wie die Autokovarianzfunktion.
- (ii) Das gegenläufige Verhalten von Varianz und Bias kommt dadurch zum Ausdruck, daß der Skalenparameter T im Zähler der Varianzkomponente und im Nenner des quadrierten Bias erscheint.

- (iii) Mit $T = cN^{1/(2q+1)}$ konvergieren beide Terme des MSE gleichmäßig mit der Rate $N^{-2q/2q+1}$ gegen Null. Dies ist ein Hinweis darauf, daß Lagfenstergeneratoren mit einem nicht zu kleinen charakteristischen Exponenten vorzuziehen sind. Es kann selbstverständlich nicht erwartet werden, daß für einen gegebenen *endlichen Stichprobenumfang* N durch Wahl eines Lagfenstergenerators mit großem charakteristischen Exponenten q die bei parametrischen Schätzungen mögliche Konvergenzrate von N^{-1} angenähert werden kann. Die hauptsächlich angewandten Lagfenstergeneratoren besitzen einen charakteristischen Exponenten $q = 2$, was einer Konvergenzrate von $N^{-4/5}$ entspricht.
- (iv) Betrachtet man Gleichung (4.45), dann beeinflußt bei gegebener Konstante c und gegebener Spektraldichte $f(\omega)$ der gewählte Lagfenstergenerator die Varianzkomponente des MSE durch den Faktor $\int_{-\infty}^{\infty} \mathcal{L}^2(u)du$ und die Biaskomponente durch den quadrierten charakteristischen Wert $[\mathcal{L}^{(q)}]^2$. Bei gleichem q ist damit ein Lagfenstergenerator $\mathcal{L}(\cdot)$ einem anderen Lagfenstergenerator $\mathcal{L}_*(\cdot)$ vorzuziehen, wenn gilt:

$$(4.46) \quad \int_{-\infty}^{\infty} \mathcal{L}^2(u)du < \int_{-\infty}^{\infty} \mathcal{L}_*^2(u)du \quad \text{und} \quad [\mathcal{L}^{(q)}]^2 < [\mathcal{L}_*^{(q)}]^2.$$

In Tabelle 4.1 sind einige Lagfenstergeneratoren mit ihren wichtigsten Kenngrößen dargestellt. Neben den genannten asymptotischen Beurteilungskriterien gibt es ein weiteres wichtiges Kriterium, bei dem geprüft wird, ob die Lagfenstergeneratoren tatsächlich immer positive Spektralschätzungen liefern. Bei Lagfenstergeneratoren, deren zugehöriges Periodogrammfenster in bestimmten Bereichen negative Werte annehmen kann (z.B. Tukey-Hamming, Tukey-Hanning), ist dies nicht immer gewährleistet.

Vom Bartlett-Fenster wird abgeraten, da es mit einem charakteristischen Exponenten von $q = 1$ im allgemeinen schlechtere Konvergenzeigenschaften erwarten läßt, als die Lagfenstergeneratoren mit $q = 2$.

Für die Lagfenstergeneratoren mit gleichem charakteristischen Exponenten $q = 2$ ist ein Vergleich anhand ihrer Varianz- und Biasfaktoren, $\int_{-\infty}^{\infty} \mathcal{L}^2(u)du$ und $[\mathcal{L}^{(q)}]^2$, anhand des Kriteriums (4.46) möglich. Als wesentliche Aussage ergibt sich, daß keiner der Lagfenstergeneratoren mit einem $q = 2$ einem anderen sowohl bezüglich des Varianzfaktors als auch bezüglich des Biasfaktors überlegen ist. Berücksichtigt man allerdings das Kriterium, daß die resultierenden Spektralschätzungen garantiert positiv sein sollen, so

Lagfenstergeneratoren			
Lagfenstergeneratoren vom gestutzten Typ, d.h. $\mathcal{L}(u) = 0$ für $ u > 1$			
	q	$\int_{-\infty}^{\infty} \mathcal{L}^2(u) du$	$[\mathcal{L}^{(q)}]^2$
Bartlett $\mathcal{L}(u) = 1 - u $	1	2/3	1
Tukey-Hamming $\mathcal{L}(u) = 0.54 + 0.46 \cos \pi u$	2	0.795	5.15
Tukey-Hanning $\mathcal{L}(u) = 0.5 + 0.5 \cos \pi u$	2	0.75	6.09
Parzen $\mathcal{L}(u) = \begin{cases} 1 - 6u^2 + 6 u ^3 & u \leq 0.5 \\ 2(1 - u)^3 & 0.5 < u \leq 1 \end{cases}$	2	0.539	36
Lagfenstergenerator vom ungestutzten Typ			
	q	$\int_{-\infty}^{\infty} \mathcal{L}^2(u) du$	$[\mathcal{L}^{(q)}]^2$
Bartlett-Priestley $\mathcal{L}(u) = \frac{3}{(\pi u)^2} \left(\frac{\sin \pi u}{\pi u} - \cos \pi u \right)$	2	1.2	0.97

Tabelle 4.1: Lagfenstergeneratoren: Varianz- und Biaskomponente

Lagfenstergeneratoren		
	EBW	EDF
Bartlett	$3/(2T)$	$3N/T$
Tukey-Hamming	$4/(3T)$	$2.672N/T$
Tukey-Hanning	$1.26/T$	$2.52N/T$
Parzen	$1.85/T$	$3.71N/T$
Bartlett-Priestley	$5/(6T)$	$1.67N/T$

Tabelle 4.2: Lagfenstergeneratoren: EBW und EDF

können das Tukey-Hamming-Fenster und das Tukey-Hanning-Fenster aus der engeren Wahl ausgeschlossen werden.

Konfidenzintervalle für indirekte Spektralschätzer $\hat{f}^{In}(\omega_k)$ bzw. $\log \hat{f}^{In}(\omega_k)$ ergeben sich analog wie bei den direkten Spektralschätzern (vgl. (4.32) und (4.33)). Die äquivalente Anzahl von Freiheitsgraden (EDF) ergibt sich für indirekte Spektralschätzer mit Lagfenstern in Skalenparameterform als (vgl. z.B. PRIESTLEY[1981, S. 467]):

$$(4.47) \quad \nu = \frac{2N}{T \int_{-\infty}^{\infty} \mathcal{L}^2(u) du}.$$

In Tabelle 4.2 sind äquivalente Bandbreite (EBW) und äquivalente Anzahl der Freiheitsgrade (EDF) für die Lagfenstergeneratoren aus Tabelle (4.1) zusammengefaßt.

Wie sich gezeigt hat, ist es aufgrund von Entscheidungskriterien, die sich hauptsächlich auf asymptotisch gültige Aussagen stützen, kaum möglich, die grundsätzliche Überlegenheit eines bestimmten Fensters zu belegen. Dies unterstreicht vom theoretischen Standpunkt, die durch praktische Beispiele belegte Ansicht von JENKINS/WATTS[1968, S. 272ff], daß das Problem der Wahl eines speziellen Lagfensters im Vergleich zu dem Problem einen geeigneten Wert für den Stützpunkt T festzulegen, von untergeordneter Bedeutung ist. Für die weitere Überprüfung, inwieweit für indirekte Spektralschätzer ein geeigneter Stützpunkt nach objektiven Kriterien aus den Daten ermittelt werden kann, wird sich die Arbeit auf den Parzen-Schätzer beschränken.

Kapitel 5

Diskrepanzmaße für Spektraldichtefunktionen

Bei Simulationsversuchen, sowohl mit nichtparametrischen als auch mit parametrischen Spektralschätzern, tritt das Problem auf, daß die zur Ergänzung der graphischen Darstellungen herangezogenen üblichen mathematisch-statistischen Fehlerkriterien (Diskrepanzen, Pseudo-Distanzen) unplausible Aussagen liefern. Ein Vergleich zweier nichtparametrischer Schätzer an Hand dieser Fehlerkriterien ergibt z.B., daß die visuell wahrnehmbaren Unterschiede zwischen den alternativen Schätzverfahren nur unzureichend quantitativ wiedergegeben werden. Bei den autoregressiven Spektralschätzern sind diese Fehlerkriterien nicht dazu geeignet, eine aus spektralanalytischer Sicht akzeptable Rangfolge für Schätzer mit völlig unterschiedlichem Auflösungsvermögen festzulegen.

Da ein aussagefähiger quantitativer Vergleich von Spektraldichtefunktionen einen hohen methodischen Stellenwert besitzt, erfolgt in diesem Abschnitt zunächst eine Evaluierung der geläufigsten Fehlerkriterien: Ein Vergleich zweier nichtparametrischer Spektralschätzer anhand der Realisation eines AR(2)-Prozesses soll die Grundproblematik verdeutlichen. Um die Wahl eines geeigneten Fehlerkriteriums möglichst allgemein und unabhängig von einem Schätzproblem analysieren zu können, erfolgt dann eine Diskussion des Problems anhand theoretischer Spektraldichtefunktionen. Bei der Auswahl der theoretischen Spektraldichtefunktionen werden allerdings Eigenschaften von verschiedenen autoregressiven Spektralschätzern bei kleinen Stichprobenumfängen berücksichtigt. Die vorgegebenen theoretischen Spektraldichten sind damit typisch für die Probleme, wie sie bei der Evaluierung hochauflösender Spektralschätzer auftreten können.

Ferner kann eine Differenzierung zwischen Problemen, die in den Fehlerkriterien selbst begründet sind und Problemen, die ihre Ursache in den numerischen Integrationsverfahren haben, nur anhand eines Vergleichs von theoretischen Spektraldichten erfolgen: Da hierzu die Fehlerkriterien als numerische Funktionen implementiert werden müssen, ist es erforderlich, daß die zu vergleichenden Funktionen in ihrer theoretischen Form bekannt sind.

Nachdem mit Hilfe der Beispiele verdeutlicht wurde, daß die gängigen Fehlerkriterien im allgemeinen keine sinnvollen spektralanalytischen Aussagen liefern, wird als genereller Lösungsansatz die Klasse der global korrigierten Diskrepanzmaße empfohlen. Zur Evaluierung der im Rahmen dieser Arbeit durchgeführten Simulationen wird aus der Klasse der global korrigierten Diskrepanzmaße ein Maß herangezogen, das besonderes geeignet erscheint, die visuell wahrnehmbaren Unterschiede zwischen zwei Spektren zu beschreiben.

5.1 Evaluierung klassischer Diskrepanzmaße

Grundlage eines quantitativen Vergleichs von Spektraldichten sind Diskrepanzmaße $d(f, g)$ die angeben, wie gut eine Spektraldichte $f = f(\omega)$ durch eine andere Spektraldichte $g = g(\omega)$ approximiert wird. Der Begriff Diskrepanzmaß wird in Anlehnung an LINHART/ZUCCHINI[1986, S. 11-12] mit der Absicht verwandt, die hier verwendeten Fehlerkriterien von den in der Analysis gebräuchlichen Distanzmaßen abzugrenzen, die stärkere metrische Eigenschaften besitzen.

Handelt es sich bei der approximierenden Spektraldichte g um eine Schätzung $\hat{f} = \hat{f}(\omega)$ der Spektraldichte f , so sind geeignete Diskrepanzmaße in mehrfacher Hinsicht von Bedeutung. Sie sind erforderlich, um theoretische Aussagen über die Güte des Schätzers \hat{f} abzuleiten. Weiterhin kann mit Hilfe von Simulationen die Beurteilung der Güte einer Spektralschätzung $\hat{f}(\omega)$ z.B. dadurch erfolgen, daß für ein geeignetes Diskrepanzmaß $d(\hat{f}, f)$ der Erwartungswert von $d(\hat{f}, f)$ als Mittelwert aus einer hinreichend großen Zahl von Realisationen des gewählten Diskrepanzmaßes $d(\hat{f}, f)$ geschätzt wird. Darüberhinaus sind Diskrepanzmaße $d(\hat{f}, f)$ der Ausgangspunkt, um Modellselektionskriterien abzuleiten, mit deren Hilfe es möglich ist, ansonsten subjektiv vorzugebende Größen für Spektralschätzverfahren anhand vorgegebener Daten zu ermitteln (vgl. z.B. HURVICH[1985]).

Die Eigenschaften von Diskrepanzmaßen für schwach stationäre stochastische Prozesse

im Zeit- oder Frequenzbereich werden seit langem in der Literatur, häufig auch unter Berücksichtigung von informationstheoretischen Maßen, diskutiert. Ohne Anspruch auf Vollständigkeit sei z.B. auf die folgenden Arbeiten verwiesen: KAILATH[1967], ITAKURA/SAITO[1970], GRAY/MARKEL[1976], GERSCH ET AL.[1979], KAZAKOS/PAPANTONI-KAZAKOS[1980], GRAY ET AL.[1980], SUGIMOTO/WADA[1988], EPHRAIM/LEVARI/GRAY[1988], SOONG/SONDHI[1988], BASSEVILLE[1989], JONES/BYRNE[1990], PARZEN[1982, 1983, 1992]. Eine ausführliche Diskussion dieses aktuellen Forschungsgebietes kann im Rahmen dieser Arbeit nicht erfolgen, die Darstellung beschränkt sich deshalb auf eine Evaluierung der bekanntesten Kriterien.

Sinnvolle Diskrepanzmaße $d(g, f)$ für den Vergleich theoretischer Spektraldichten lassen sich durch die folgenden Bedingungen charakterisieren:

Def. 5.1.1 (*Diskrepanzmaße*)

Seien f, g Spektraldichtefunktionen. Dann gilt für ein Diskrepanzmaß $d(f, g)$:

$$(5.1) \quad \begin{aligned} (i) \quad & d(f, g) \geq 0 \\ (ii) \quad & d(f, g) = 0 \Leftrightarrow f = g \end{aligned}$$

■

Als Diskrepanz- bzw. Fehlermaße für zwei Spektraldichten $f(\omega)$ und $g(\omega)$ auf dem Intervall $[a, b]$ mit $-0.5 \leq a < b \leq 0.5$ können u.a. die folgenden Größen herangezogen werden (vgl. z.B. BASSEVILLE[1989]):

(Integrierter absoluter Fehler, IAE)

$$(5.2) \quad d_1(f, g) = \int_a^b |\log f(\omega) - \log g(\omega)| d\omega,$$

(Integrierter quadratischer Fehler, ISE)

$$(5.3) \quad d_2(f, g) = \int_a^b [\log f(\omega) - \log g(\omega)]^2 d\omega,$$

(Maximaler absoluter Fehler, MAE)

$$(5.4) \quad d_\infty(f, g) = \max_\omega |\log f(\omega) - \log g(\omega)|.$$

(Kullback-Leibler-Informations-Kriterium, KLIC)

$$(5.5) \quad d_{KL}(f, g) = \text{KLIC}(f, g) = \frac{1}{2} \int_a^b \left[\frac{f(\omega)}{g(\omega)} - \ln \frac{f(\omega)}{g(\omega)} - 1 \right] d\omega.$$

Das Kullback-Leibler-Informationskriterium war in Abschnitt 2.6 diskutiert und abgeleitet worden. In der ingenieurwissenschaftlichen Literatur erscheint das Kullback-Leibler-Informationskriterium, multipliziert mit dem Faktor zwei, sehr häufig unter dem Begriff Itakura-Saito-Distanz (vgl. z.B. ITAKURA/SAITO[1970], BASSEVILLE[1989]). Bei der Erörterung von theoretischen Schätzeigenschaften wird u.a. auch das folgende Diskrepanzmaß herangezogen (vgl. z.B. SHIBATA[1981]):

(Integrierter relativer quadratischer Fehler, ISER)

$$(5.6) \quad d_{2R}(f, g) = \int_a^b \left[\frac{f(\omega) - g(\omega)}{f(\omega)} \right]^2 d\omega.$$

Für den integrierten absoluten Fehler und den integrierten quadratischen Fehler geben KOOPERBERG/STONE[1991] ein Beispiel aus dem Bereich der gewöhnlichen nichtparametrischen Dichteschätzung, das die mangelhafte Wiedergabe von visuell auffälligen Unterschieden zwischen Dichtefunktionen bei Anwendung dieser Fehlerkriterien belegt.

Zur Berechnung der in (5.2), (5.3), (5.5) und (5.6) auftretenden Integrale werden diese üblicherweise durch Riemannsummen approximiert (vgl. z.B. HURVICH[1985], MARKEL/GRAY[1976, S. 238]), so daß z.B. der integrierte quadratische Fehler berechnet wird als:

$$(5.7) \quad d_2(g, f) \approx \frac{b-a}{N_\omega} \sum_{j=1}^{N_\omega} [\log g(\omega_j) - \log f(\omega_j)]^2,$$

wobei N_ω geeignet festzulegen ist.

Zunächst werden zwei nichtparametrische Schätzungen für eine Realisation des AR(2)-Prozesses

$$(1 - 0.2e^{i2\pi 0.1}L)(1 - 0.2e^{-i2\pi 0.1}L)X_t = 1 - 0.32361X_{t-1} + 0.04X_{t-2} = \epsilon_t \quad \epsilon_t \sim \text{WN}(0, 1)$$

graphisch mit dem wahren Spektrum verglichen (vgl. Abbildung 5.1). Die Güte der beiden Approximationen wird anhand der Diskrepanzen (5.2), (5.3), (5.4), (5.5), (5.6) in Tabelle 5.1 bewertet. Wie der graphische Vergleich zeigt, ist die Approximation des wahren Spektrums durch den Daniellschätzer ($B = 57$) bzw. durch den Parzenschätzer ($T = 7$) in qualitativer Hinsicht höchst unterschiedlich: Während der Parzenschätzer eine glatte Approximation des wahren Spektrums liefert, ist der Verlauf des Daniellschätzers höchst unruhig, im Frequenzbereich $0.35 < \omega < 0.5$ weist

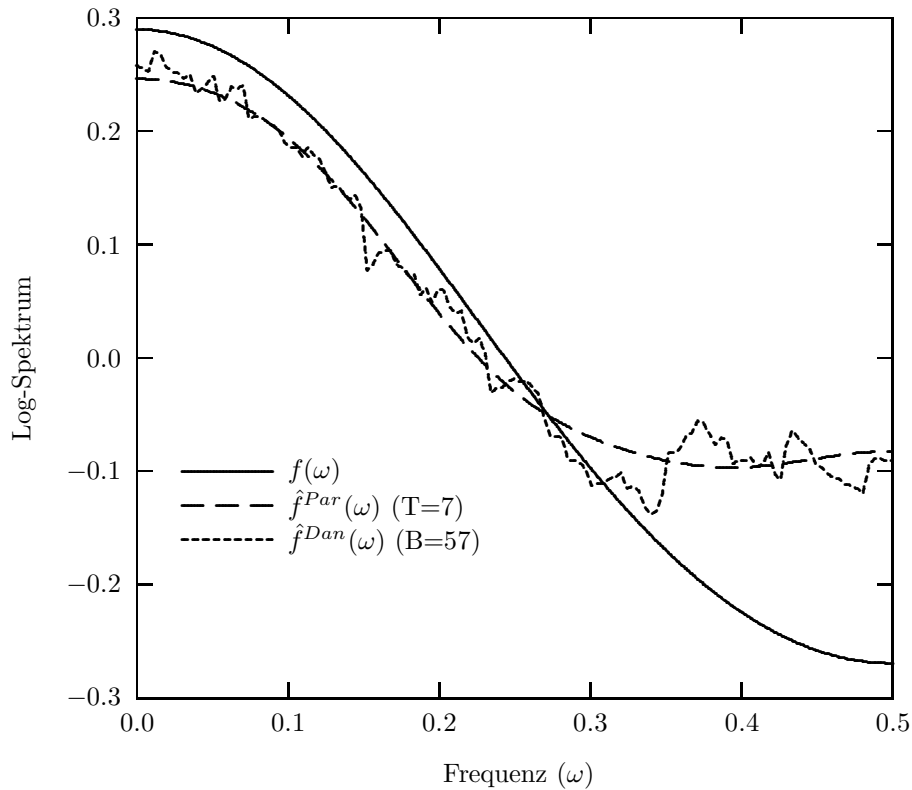


Abb. 5.1: Vergleich: Direkter Daniellschätzer ($B=57$) und Parzenschätzer ($T=7$)

der Daniellschätzer sogar zwei mehr oder weniger ausgeprägte breite Peaks auf und vermittelt dadurch den Eindruck, daß der datenerzeugende Prozeß relevante hochfrequente Schwingungskomponenten besitzt. Im Frequenzbereich $0.3 < \omega < 0.5$ besitzen beide Schätzer einen erheblichen Bias. Aus spektralanalytischer Sicht ist die glatte Approximation durch den Parzenschätzer der rauhen und unruhigen Approximation durch den Daniellschätzer vorzuziehen. Wie Tabelle 5.1 zu entnehmen ist, ergibt sich anhand der Diskrepanzen allerdings durchgängig eine gegenteilige Aussage: Der Parzenschätzer weist bezüglich aller Kriterien den größeren Fehler auf. Damit ist für ein Beispiel gezeigt, daß die gängigen Diskrepanzen bei einem Vergleich qualitativ verschiedener Schätzungen Aussagen liefern können, die mit einer graphischen Interpretation der Schätzergebnisse nicht zu vereinbaren sind.

Bei der Bewertung der Fehlerintegrale war jeweils eine Approximation durch Riemannsummen (vgl. Gleichung (5.7)) verwandt worden. Somit kann nicht ausgeschlossen werden, daß die von der graphischen Interpretation abweichenden Ergebnisse der Diskrepanzen nicht lediglich auf numerische Fehler zurückzuführen sind. Im folgenden soll

$N_\omega = 129$	Parzen ($T = 7$)	Daniell ($B = 57$)
IAE	0.03471	0.03260
ISE	0.00389	0.00379
MAE	0.18703	0.18523
KLIC	0.00933	0.00910
ISER	0.02883	0.02808

Tabelle 5.1: Diskrepanzvergleich: Daniell- und Parzenschätzung

deshalb anhand von Beispielen mit theoretischen Spektraldichten und unter Berücksichtigung von möglichen numerischen Fehlern gezeigt werden, daß im allgemeinen keines der angegebenen Kriterien ein zufriedenstellendes Diskrepanzmaß für Spektraldichtefunktionen ist. Hierzu wird zunächst eine Referenzspektraldichte f_0 mit vier weiteren approximierenden Spektraldichten f_1, \dots, f_4 paarweise anhand der oben vorgestellten Kriterien verglichen. Für jedes Kriterium wird eine Rangfolge der am besten approximierenden Spektraldichten ermittelt. Zur Beurteilung der ermittelten Rangfolgen wird eine Referenzrangfolge vorgegeben, die anhand eines graphischen Vergleiches nach spektralanalytischen Gesichtspunkten festgelegt wird.

Alle Spektraldichten f_0, \dots, f_4 werden anhand reeller AR(8)-Prozesse $\{X_{0,t}\}, \dots, \{X_{4,t}\}$ mit

$$\phi^{(k)}(L)X_{k,t} = \left(1 + \sum_{j=1}^8 \phi_{k,j}L^j\right)X_{k,t} = \epsilon_{k,t} \quad \epsilon_{k,t} \sim \text{WN}(0, 1) \quad k = 0, \dots, 4$$

konstruiert, wobei für die charakteristischen Polynome der Filter $\phi^{(k)}(L)$,

$$\phi^{(k)}(z) = \prod_{j=1}^8 (1 - z_{k,j}z^{-1}), \quad k = 0, \dots, 4,$$

unterschiedliche Wurzeln $z_{k,j}$ vorgegeben werden (in Klammern sind kennzeichnende Kurzbezeichnungen beigefügt) :

(Referenz)

$$(5.8) \quad \phi^{(0)}(z) = (1 - 0.999e^{i2\pi 0.125} z^{-1})(1 - 0.999e^{-i2\pi 0.125} z^{-1}) \\ (1 - 0.999e^{i2\pi 0.135} z^{-1})(1 - 0.999e^{-i2\pi 0.135} z^{-1}) \\ (1 - 0.999e^{i2\pi 0.145} z^{-1})(1 - 0.999e^{-i2\pi 0.145} z^{-1}) \\ (1 - 0.999e^{i2\pi 0.155} z^{-1})(1 - 0.999e^{-i2\pi 0.155} z^{-1})$$

(Sub-Peak)

$$(5.9) \quad \phi^{(1)}(z) = (1 - 0.975e^{i2\pi 0.125} z^{-1})(1 - 0.975e^{-i2\pi 0.125} z^{-1}) \\ (1 - 0.965e^{i2\pi 0.135} z^{-1})(1 - 0.965e^{-i2\pi 0.135} z^{-1}) \\ (1 - 0.965e^{i2\pi 0.145} z^{-1})(1 - 0.965e^{-i2\pi 0.145} z^{-1}) \\ (1 - 0.975e^{i2\pi 0.155} z^{-1})(1 - 0.975e^{-i2\pi 0.155} z^{-1})$$

(Cross-Peak)

$$(5.10) \quad \phi^{(2)}(z) = (1 - 0.985e^{i2\pi 0.130} z^{-1})(1 - 0.985e^{-i2\pi 0.130} z^{-1}) \\ (1 - 0.984e^{i2\pi 0.137} z^{-1})(1 - 0.984e^{-i2\pi 0.137} z^{-1}) \\ (1 - 0.984e^{i2\pi 0.143} z^{-1})(1 - 0.984e^{-i2\pi 0.143} z^{-1}) \\ (1 - 0.985e^{i2\pi 0.152} z^{-1})(1 - 0.985e^{-i2\pi 0.152} z^{-1})$$

(Bias-Two-Peak)

$$(5.11) \quad \phi^{(3)}(z) = (1 - 0.970e^{i2\pi 0.125} z^{-1})(1 - 0.970e^{-i2\pi 0.125} z^{-1}) \\ (1 - 0.999e^{i2\pi 0.137} z^{-1})(1 - 0.999e^{-i2\pi 0.137} z^{-1}) \\ (1 - 0.999e^{i2\pi 0.147} z^{-1})(1 - 0.999e^{-i2\pi 0.147} z^{-1}) \\ (1 - 0.970e^{i2\pi 0.155} z^{-1})(1 - 0.970e^{-i2\pi 0.155} z^{-1})$$

(Bias-Four-Peak)

$$(5.12) \quad \phi^{(4)}(z) = (1 - 0.999e^{i2\pi 0.125} z^{-1})(1 - 0.999e^{-i2\pi 0.125} z^{-1}) \\ (1 - 0.999e^{i2\pi 0.137} z^{-1})(1 - 0.999e^{-i2\pi 0.137} z^{-1}) \\ (1 - 0.999e^{i2\pi 0.149} z^{-1})(1 - 0.999e^{-i2\pi 0.149} z^{-1}) \\ (1 - 0.999e^{i2\pi 0.161} z^{-1})(1 - 0.999e^{-i2\pi 0.161} z^{-1}).$$

Die Spektraldichten f_0, \dots, f_4 werden berechnet als:

$$(5.13) \quad f_k(\omega) = \frac{\sigma_\epsilon^2}{|\sum_{j=0}^p \phi_{k,j} e^{-i2\pi\omega j}|^2} \quad \phi_{k,0} = 1, \quad \omega \in [0.0, 0.5].$$

Da die Spektraldichten über AR(8)-Prozesse festgelegt werden, ist zusätzlich ein Vergleich im Zeitbereich möglich.

Zunächst wird die Referenzrangfolge anhand der Abbildungen 5.2 (a) - (d) bestimmt, in denen jeweils die Referenzspektraldichte f_0 zusammen mit einer der approximierenden Spektraldichten f_k , ($k = 1, \dots, 4$) dargestellt ist. Die Spektraldichten wurden so gewählt, daß sie an den Flanken weitgehend übereinstimmen. Wesentliche Unterschiede befinden sich im Peak-Bereich, der in etwa das Frequenzband $0.12 \leq \omega \leq 0.16$ umfaßt.

Hauptkriterium zur Festlegung der Referenzrangfolge ist die Anzahl der übereinstimmenden Peaks, wobei kleinere Abweichungen bezüglich der Lokalisation der Peaks als nicht wesentlich eingestuft werden. Unter dieser Prämisse kann den Abbildungen 5.2 (a) - (d) entnommen werden, daß die Approximation der Spektraldichte f_0 durch die Spektraldichten f_k mit zunehmenden Index $k = 1, 2, 3, 4$ immer besser wird: Die Spektraldichte f_1 (Sub-Peak) besitzt lediglich einen breiten Peak, dessen Niveau weit unterhalb der entsprechenden Werte der Referenzspektraldichte liegt. Die Spektraldichte f_2 (Cross-Peak) erreicht mit einem breiten Peak zwar in etwa das Niveau der Referenzspektraldichte, visuell (als auch numerisch) ist jedoch nur ein Peak feststellbar. Bei der Spektraldichte f_3 (Bias-Two-Peak) sind zwei Peaks klar unterscheidbar, allerdings sind diese im Vergleich zur Referenzspektraldichte leicht nach rechts versetzt. Die Spektraldichte f_4 (Bias-Four-Peak) entspricht der Form nach weitgehend der Referenzspektraldichte, allerdings sind drei der vier Peaks leicht nach rechts unten verschoben. Somit ergibt sich anhand der graphischen Darstellungen aus spektralanalytischer Sicht die in Tabelle 5.2 dargestellte Referenzrangfolge für die Approximationsgüte des Prozesses $\{X_{0,t}\}$ bzw. der Spektraldichte f_0 durch die Prozesse $\{X_{k,t}\}$ bzw. die Spektraldichten f_k ($k = 1, \dots, 4$).

Um numerische Fehler weitgehend auszuschließen, wurden für $N_\omega = 128, 256, 512, 1024$ Kontrollrechnungen mit Hilfe der genaueren Gauss-Legendre-Quadratur durchgeführt. Hierzu wurde die Routine `intquad1` aus GAUSS 3.0 verwandt, wobei das Intervall $[0.0, 0.5]$ in $(N_\omega - 1)$ gleich lange Teilintervalle zerlegt wurde und für jedes Teilintervall eine Gauss-Legendre-Quadratur vierter Ordnung berechnet wurde. Verfahren

Prozeß, Spektraldichte		Referenzrang
Sub-Peak	$\{X_{1,t}\}, f_1$	(4)
Cross-Peak	$\{X_{2,t}\}, f_2$	(3)
Two-Peak	$\{X_{3,t}\}, f_3$	(2)
Bias-Four-Peak	$\{X_{4,t}\}, f_4$	(1)

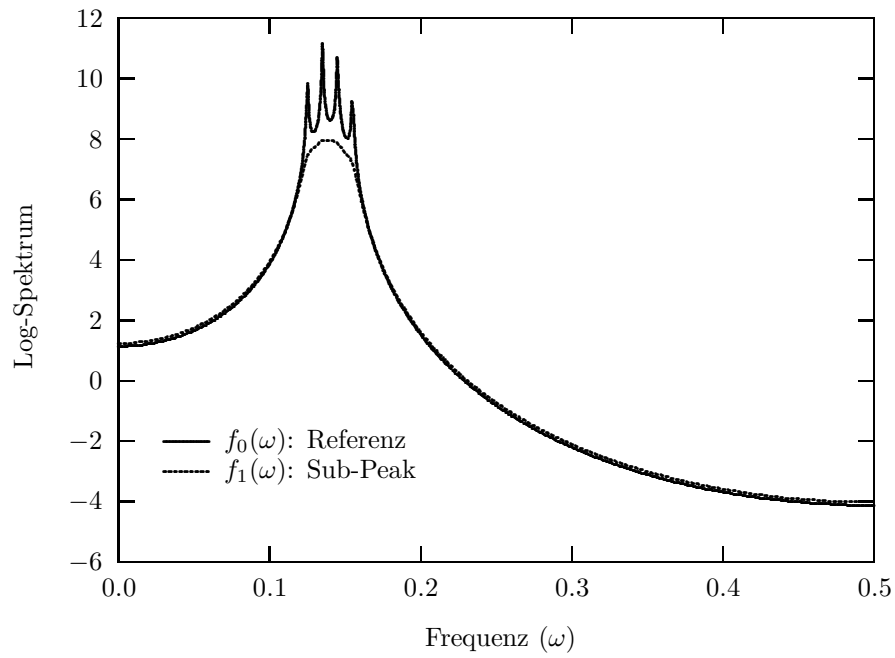
Tabelle 5.2: Referenzrangfolge der AR(8)-Prozesse/Spektraldichten

der numerischen Integration mit Berücksichtigung der Gauss-Legendre-Quadratur sind z.B. in PRESS ET AL. [1992, S. 129ff] dargestellt. Sowohl für den integrierten absoluten Fehler (IAE) als auch für den integrierten quadratischen Fehler (ISE) konnten für $N_\omega = 512$ auch bei den hier zugrundegelegten Spektraldichten mit ausgeprägten Peaks hinreichend genaue Ergebnisse erzielt werden; bei dem Kullback-Leibler-Informationskriterium (KLIC), in das auch die nicht logarithmierten Spektren eingehen, mußte die Anzahl der Frequenzpunkte auf $N_\omega = 1024$ erhöht werden (vgl. Tabelle 5.4). Der integrierte relative quadratische Fehler (ISER) erwies sich als numerisch instabil und wurde deshalb nicht weiter berücksichtigt.

Bereits eine Betrachtung im Zeitbereich, wobei lediglich die aufsummierten, quadrierten Koeffizientenabstände $\sum_{j=1}^8 (\phi_{k,j} - \phi_{0,j})^2$ als Diskrepanzmaß herangezogen werden, ergibt Abweichungen von der als spektralanalytisch sinnvoll ermittelten Referenzrangfolge (vgl. Tabelle 5.3): Während der Bias-Four-Peak-Prozeß $\{X_{4,t}\}$ und der Sub-Peak-Prozeß $\{X_{1,t}\}$ korrekt als bester bzw. schlechtester approximierender Prozeß eingestuft werden, wird für den eingipfeligen Cross-Peak-Prozeß $\{X_{2,t}\}$ ein deutlich geringerer quadrierter Koeffizientenabstand ausgewiesen als für den Bias-Two-Peak-Prozeß $\{X_{3,t}\}$.

Noch weitaus weniger akzeptabel sind die Ergebnisse der Fehlerkriterien $d_1(f_0, f_k)$ (IAE), $d_2(f_0, f_k)$ (ISE) und $d_\infty(f_0, f_k)$ (MAE) im Spektralbereich (vgl. Tabelle 5.4): Die Spektraldichte f_4 des Bias-Four-Peak-Prozesses wird sowohl nach dem integrierten absoluten Fehler (IAE) als auch nach dem integrierten quadratischen Fehler (ISE) mit jeweils Rang (4) als schlechteste Approximation eingestuft, die Bewertung nach dem maximalen absoluten Fehler ergibt den zweitschlechtesten Rang (3). Ein besseres Ergebnis erzielt das Kullback-Leibler-Informationskriterium (KLIC), lediglich die Approximation durch das Cross-Peak-Spektrum f_2 wird besser als die Approximation

(a) Referenz- und Sub-Peak-Spektrum



(b) Referenz- und Cross-Peak-Spektrum

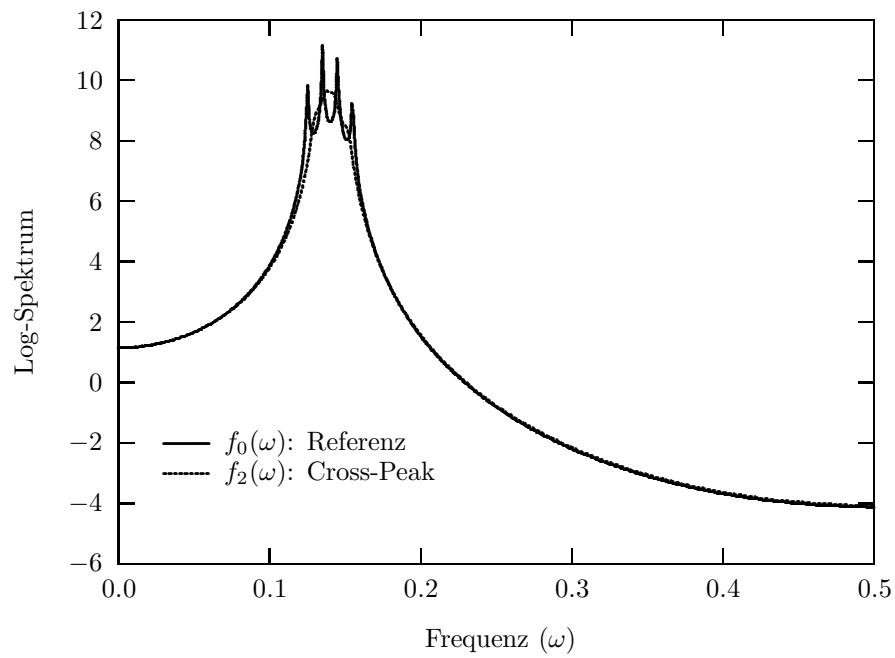
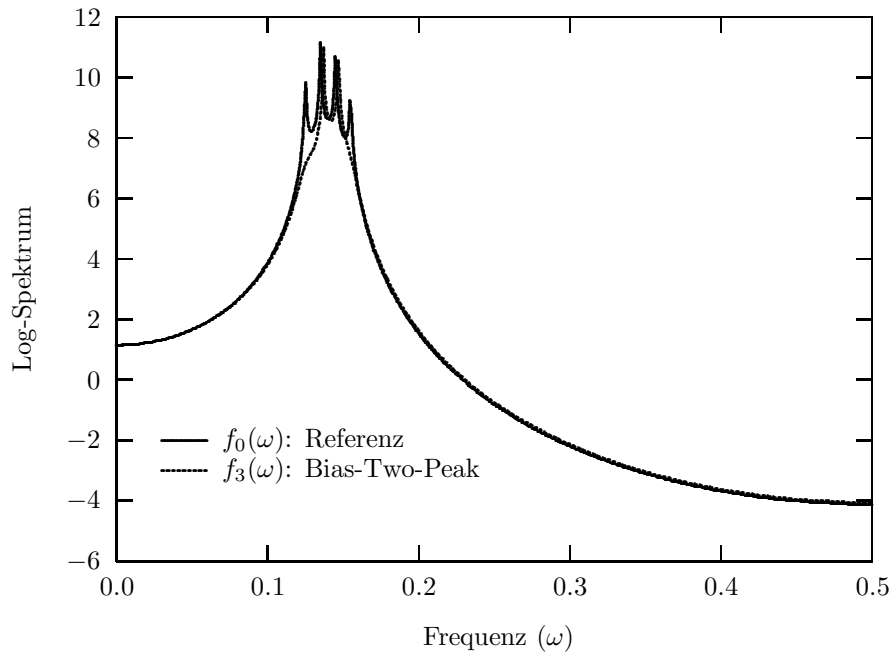


Abb. 5.2: Vergleich von Spektraldichten, siehe Angaben in Tabelle (5.4).

(c) Referenz- und Bias-Two-Peak-Spektrum



(d) Referenzspektrum und Bias-Four-Peak-Spektrum

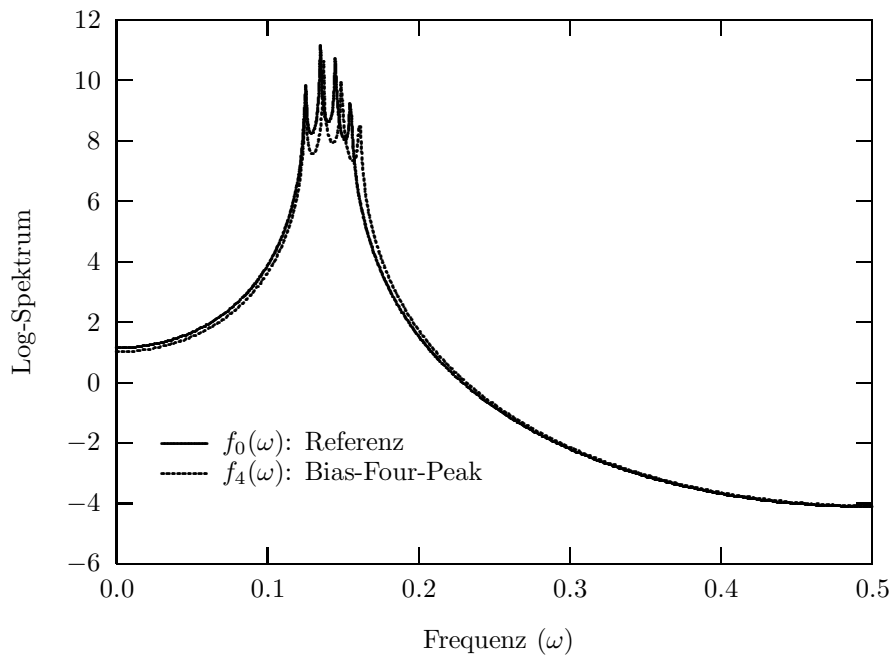


Abb. 5.2: (Fortsetzung) Vergleich von Spektraldichten, siehe Angaben in Tabelle (5.4).

Diskrepanzmaß - Zeitbereich			
		$\sum_{j=1}^8 (\phi_{k,j} - \phi_{0,j})^2$	Rang
Sub-Peak	$\{X_{1,t}\}$	30.08554	(4)
Cross-Peak	$\{X_{2,t}\}$	8.71211	(2)
Two-Peak	$\{X_{3,t}\}$	11.37324	(3)
Bias-Four-Peak	$\{X_{4,t}\}$	3.77941	(1)

Tabelle 5.3: Diskrepanzvergleich, AR(8)-Prozessen im Zeitbereich

durch das Bias-Four-Peak-Spektrum f_4 eingestuft.

Nach allen vier Kriterien wird die Spektraldichte f_2 des Cross-Peak-Prozesses als die beste approximierende Spektraldichte eingestuft.

Der erhebliche Unterschied zwischen der spektralanalytisch sinnvollen Referenzrangfolge und den mit Hilfe der Diskrepanzmaße ermittelten Rangfolgen läßt sich dadurch erklären, daß alle Diskrepanzmaße grundsätzlich durch lokale paarweise Punktvergleiche definiert sind, während bei der vergleichenden visuellen Analyse die globale Ähnlichkeitsstruktur der Spektraldichten erfaßt wird, wobei kleine lokale Abweichungen so gut wie keine Rolle spielen. Die Diskrepanzmaße können einerseits als zu exakt in dem Sinne bezeichnet werden, daß unbedeutende lokale Abweichungen ein viel zu großes Gewicht beigemessen wird. So ist die Verschiebung eines Peaks um einen geringen Betrag $\Delta\omega$ entlang der Frequenzachse spektralanalytisch akzeptabler, als diesen Peak gar nicht auszuweisen. Andererseits sind Diskrepanzmaße gleichzeitig insofern zu ungenau, als wichtige Information über die gemeinsame globale Struktur zweier Spektren zuwenig berücksichtigt wird. Es ist z.B. von Bedeutung ob zwei Spektraldichten die gleiche Anzahl von Peaks besitzen. Der Kern des Problems besteht somit darin, daß Diskrepanzmaße nur eine Teilmenge der verfügbaren Information für den Vergleich der Spektraldichten nutzen und daß diese Teilinformation spektralanalytisch nur bedingt aussagefähig ist.

BANDEMER/GOTTWALD[1993, S. 213ff] verweisen mit weiteren Literaturangaben auf ähnlich gelagerte Probleme bei spektralanalytischen Anwendungen in der Qualitätskontrolle, wo ein Referenzspektrum mit einem gemessenen Kontrollspektrum zu vergleichen ist. Sie stellen einen Lösungsansatz vor, bei dem die Peaks eines Spektrums mit Hilfe von unscharfen Mengen (fuzzy sets) abgebildet werden. Der Grundansatz hierzu

Diskrepanzmaße - Frequenzbereich					
		IAE ($N_\omega = 512$)			
		(Riemann)		(Gauss-Legendre)	
		$d_1(f_0, f_k)$	Rang	$d_1(f_0, f_k)$	Rang
Sub-Peak	f_1	0.08430	(3)	0.08423	(3)
Cross-Peak	f_2	0.05252	(1)	0.05268	(1)
Bias-Two-Peak	f_3	0.06253	(2)	0.06227	(2)
Bias-Four-Peak	f_4	0.09699	(4)	0.09671	(4)
		ISE ($N_\omega = 512$)			
		(Riemann)		(Gauss-Legendre)	
		$d_2(f_0, f_k)$	Rang	$d_2(f_0, f_k)$	Rang
Sub-Peak	f_1	0.06432	(3)	0.06339	(3)
Cross-Peak	f_2	0.03507	(1)	0.03531	(1)
Bias-Two-Peak	f_3	0.04785	(2)	0.04632	(2)
Bias-Four-Peak	f_4	0.07426	(4)	0.07419	(4)
		MAE ($N_\omega = 512$)			
		$d_\infty(f_0, f_k)$	Rang		
Sub-Peak	f_1	3.19085	(4)		
Cross-Peak	f_2	2.48719	(1)		
Two-Peak	f_3	2.52999	(2)		
Bias-Four-Peak	f_4	2.72017	(3)		
		KLIC ($N_\omega = 1024$)			
		(Riemann)		(Gauss-Legendre)	
		$d_{KL}(f_0, f_k)$	Rang	$d_{KL}(f_0, f_k)$	Rang
Sub-Peak	f_1	2.19682	(4)	2.23915	(4)
Cross-Peak	f_2	0.81018	(1)	0.80624	(1)
Bias-Two-Peak	f_3	1.12922	(3)	1.11911	(3)
Bias-Four-Peak	f_4	1.01950	(2)	1.04696	(2)

Tabelle 5.4: Vergleich von AR(8)-Prozessen im Frequenzbereich mit verschiedenen Diskrepanzmaßen

geht auf BLAFFERT[1984] zurück, der vor dem Hintergrund spektrometrischer Anwendungen in der Chemie ein Mustererkennungsverfahren zum Vergleich von gemessenen Spektren mit Referenzspektren aus einer Spektraldatenbank entwickelt hat.

Diese Beispiele machen deutlich, daß die in diesem Abschnitt angesprochenen Probleme bei praktischen Anwendungen besonders dann relevant sind, wenn Standardverfahren zur Auswertung von Massendaten verwendet werden. Da die Darstellung des theoretischen Hintergrundes von unscharfen Mengen den hier gesetzten Rahmen bei weitem überschreiten würde, wird diesem Ansatz nicht gefolgt.

5.2 Global korrigierte Diskrepanzmaße

Der folgende Lösungsvorschlag hält grundsätzlich an den Diskrepanzmaßen fest, wobei aber durch eine Gewichtung mit geeigneten globalen Maßen versucht wird, die zu strenge lokale Exaktheit der Diskrepanzmaße zu korrigieren.

Hierzu werden zunächst Maße benötigt, die die Ähnlichkeit von Spektraldichten beschreiben. Um die Vorgehensweise in Analogie zu den Diskrepanzmaßen zu entwickeln, werden Dissimilaritätsmaße (Unähnlichkeitsmaße) eingeführt. Für Dissimilaritätsmaße sollen die Eigenschaften der Definition 5.1.1 in abgeschwächter Form gelten:

Def. 5.2.1 (*Dissimilaritätsmaße*)

Seien f und g Spektraldichtefunktionen. Dann gilt für ein Dissimilaritätsmaß $ds(f, g)$:

$$(5.14) \quad \begin{aligned} (i) \quad & ds(f, g) \geq 0 \\ (ii) \quad & f = g \Rightarrow ds(f, g) = 0 \end{aligned}$$

■

Es existieren zahlreiche Möglichkeiten, sinnvolle Dissimilaritätsmaße für Spektraldichtefunktionen zu definieren. Da hier Dissimilaritätsmaße in Ergänzung bzw. als Korrektur zu den lokal orientierten Diskrepanzmaßen herangezogen werden sollen, liegt es nahe, globale Dissimilaritätsmaße zu betrachten oder genauer, Baustein für diese Dissimilaritätsmaße sollen deskriptive Kenngrößen sein, die einer logarithmierten Spektraldichtefunktionen in einem Intervall $[a, b]$ mit $-0.5 \leq a < b \leq 0.5$ eine (nicht-negative) reelle Zahl zuordnen. Globale Dissimilaritätsmaße vergleichen also bereits aggregierte

Informationen der einzelnen Spektraldichten. In Frage kommen z.B. die Anzahl der Peaks oder auch Streuungsmaße wie Varianz bzw. Variationskoeffizient. Die Betrachtung hier beschränkt sich auf Dissimilaritätsmaße, die auf den ersten und zweiten Differenzen der diskretisierten Spektraldichten beruhen und somit Informationen über das Steigungs- bzw. Krümmungsverhalten zum Vergleich der Spektraldichten mit heranziehen.

Hierzu werden die Spektraldichten diskret an N_ω äquidistanten Frequenzpunkten ω_j im Intervall $[a, b]$ mit $-0.5 \leq a < b \leq 0.5$ gemessen.

Die ersten und zweiten Differenzen einer diskretisierten, logarithmierten Spektraldichte $\log g(\omega_j)$ ergeben sich als:

$$\Delta \log g(\omega_j) = (1 - L) \log g(\omega_j) = \log g(\omega_j) - \log g(\omega_{j-1}) \quad j = 2, \dots, N_\omega$$

$$\Delta^2 \log g(\omega_j) = (1 - L)^2 \log g(\omega_j) = \log g(\omega_j) - 2 \log g(\omega_{j-1}) + \log g(\omega_{j-2}) \quad j = 3, \dots, N_\omega.$$

Es kann davon ausgegangen werden, daß diese Größen sich sehr gut für eine Quantifizierung der visuellen Charakteristika von Spektraldichten eignen. Dissimilaritätsmaße, die eine *globale* Aussage über das unterschiedliche Steigungs- bzw. Krümmungsverhalten der Spektraldichten machen, lassen sich z.B. formulieren als:

$$(5.15) \quad ds_{\Delta}(g, f) = \left(\frac{\sum_{j=2}^{N_\omega} |\Delta \log g(\omega_j)|}{\sum_{j=2}^{N_\omega} |\Delta \log f(\omega_j)|} - 1 \right)^2$$

und

$$(5.16) \quad ds_{\Delta^2}(g, f) = \left(\frac{\sum_{j=3}^{N_\omega} |\Delta^2 \log g(\omega_j)|}{\sum_{j=3}^{N_\omega} |\Delta^2 \log f(\omega_j)|} - 1 \right)^2.$$

Da die Dissimilaritäten hier im wesentlichen als Hilfskonstruktionen interpretiert werden, sollen die Spezialfälle, daß Zähler bzw. Nenner in (5.15) oder (5.16) einen Wert gleich Null annehmen können, nicht explizit betrachtet werden.

Dissimilaritäten - Frequenzbereich			
$N_\omega = 512$			
		$ds_\Delta(f_0, f_k)$	Rang
Sub-Peak	f_1	0.24358	(4)
Cross-Peak	f_2	0.15900	(3)
Bias-Two-Peak	f_3	0.04672	(2)
Bias-Four-Peak	f_4	0.00013	(1)
		$ds_{\Delta^2}(f_0, f_k)$	Rang
Sub-Peak	f_1	0.90441	(4)
Cross-Peak	f_2	0.80716	(3)
Bias-Two-Peak	f_3	0.17534	(2)
Bias-Four-Peak	f_4	0.00220	(1)

Tabelle 5.5: Vergleich von AR(8)-Prozessen im Frequenzbereich mit Dissimilaritäten

In Tabelle 5.5 sind für den Vergleich der Spektraldichten f_k ($k = 1, \dots, 4$) mit der Referenzspektraldichte f_0 die Dissimilaritätsmaße $ds_\Delta(f_0, f_k)$ und $ds_{\Delta^2}(f_0, f_k)$ sowie die Rangfolgen der Approximationsgüte wiedergegeben. Ganz im Gegensatz zu den lokal orientierten Diskrepanzmaßen der Tabelle 5.4 ist die von den beiden global orientierten Dissimilaritätsmaßen ermittelte Rangfolge mit der spektralanalytisch sinnvollen Referenzrangfolge identisch.

Allerdings haben Dissimilaritätsmaße einen Nachteil gegenüber den Diskrepanzmaßen: Aus $ds(f, g) = 0$ folgt nicht, daß gilt: $f = g$, vielmehr wird es für eine gegebene Spektraldichte f im allgemeinen eine ganze Klasse $\{g_k, k = 1, 2, \dots\}$ von Spektraldichten geben, die durch die Beziehung $ds(f, g_k) = 0$ festgelegt ist. Verwendet man z.B. die Anzahl der Peaks von Spektraldichten zur Konstruktion eines Dissimilaritätsmaßes, so besitzen alle Spektren mit gleicher Peakanzahl ein Dissimilaritätsmaß von Null, ohne Rücksicht darauf, an welchen Frequenzpunkten sich diese Peaks befinden. Damit sind Dissimilaritätsmaße als Fehlermaße nur eingeschränkt geeignet. Es erscheint deshalb zweckmäßig, an den strengeren Diskrepanzmaßen mit ihren Grundeigenschaften (vgl. Def. 5.1.1) festzuhalten, aber mit Hilfe der Dissimilaritätsmaße Korrekturen zur Berücksichtigung von globalen Informationen vorzunehmen.

Hierzu werden die Dissimilaritätsmaße in (multiplikativ) korrigierende Dissimilaritäts-

maße umformuliert:

Def. 5.2.2 (*Korrigierende Dissimilaritätsmaße*)

Seien f und g Spektraldichtefunktionen. Dann soll für ein (multiplikativ) korrigierendes Dissimilaritätsmaß $\text{dsc}(f, g)$ gelten:

$$(5.17) \quad \begin{aligned} (i) \quad & \text{dsc}(f, g) \geq 1 \\ (ii) \quad & f = g \Rightarrow \text{dsc}(f, g) = 1 \end{aligned}$$

■

Die Dissimilaritätsmaße $\text{ds}_\Delta(f, g)$ und $\text{ds}_{\Delta^2}(f, g)$ können in ihren korrigierenden Versionen formuliert werden als:

$$(5.18) \quad \text{dsc}_\Delta(f, g) = \text{ge_one} \left(\sum_{j=2}^{N_\omega} |\Delta \log f(\omega_j)|, \sum_{j=2}^{N_\omega} |\Delta \log g(\omega_j)| \right)$$

$$(5.19) \quad \text{dsc}_{\Delta^2}(f, g) = \text{ge_one} \left(\sum_{j=3}^{N_\omega} |\Delta^2 \log f(\omega_j)|, \sum_{j=3}^{N_\omega} |\Delta^2 \log g(\omega_j)| \right)$$

wobei die Funktion $\text{ge_one}(x_1, x_2)$ mit $x_1 \geq 0$ und $x_2 \geq 0$ definiert werden soll als:

$$\text{ge_one}(x_1, x_2) = \begin{cases} x_1/x_2 & x_1 \geq x_2, x_2 \neq 0 \\ x_2/x_1 & x_2 > x_1, x_1 \neq 0 \\ 1 & x_1 = 0 \text{ oder } x_2 = 0. \end{cases}$$

Bei der Konstruktion der korrigierenden Dissimilaritäten ist sicherzustellen, daß keines der beiden Argumente der Funktion $\text{ge_one}(x_1, x_2)$ negativ wird. Für den Fall, daß eines der Argumente gleich Null ist und eine sinnvolle Berechnung des korrigierenden Dissimilaritätsmaßes nicht durchgeführt werden kann, wird dieses gleich Eins gesetzt. Beliebige Diskrepanzmaße $d(f, g)$ können nun mit beliebig vielen korrigierenden Dissimilaritätsmaßen $\text{dsc}_1(f, g), \dots, \text{dsc}_M(f, g)$ gewichtet werden, so daß sich ein korrigiertes Diskrepanzmaß $\text{dc}(f, g)$ z.B. definieren läßt als:

Def. 5.2.3 (*Korrigierte Diskrepanzmaße*)

Seien f und g Spektraldichtefunktionen, sei $d(f, g)$ ein Diskrepanzmaß und seien

$$\text{dsc}_1(f, g), \dots, \text{dsc}_M(f, g)$$

(multiplikativ) korrigierende Dissimilaritätsmaße. Dann heißt

$$(5.20) \quad \text{dc}(f, g) = d(f, g) \prod_{m=1}^M \text{dsc}_m(f, g)$$

mit den Eigenschaften:

$$(5.21) \quad \begin{aligned} (i) \quad & \text{dc}(f, g) \geq 0 \\ (ii) \quad & \text{dc}(f, g) = 0 \Leftrightarrow f = g \end{aligned}$$

(multiplikativ) korrigiertes Diskrepanzmaß. ■

Die korrigierten Diskrepanzmaße können selbstverständlich noch allgemeiner formuliert werden, wenn z.B. jeder Faktor in (5.20) noch mit einem Gewichtungsfaktor potenziert wird,

$$(5.22) \quad \text{dc}(f, g) = d(f, g)^{w_0} \prod_{m=1}^M \text{dsc}_m^{w_m}(f, g)$$

oder wenn zugelassen wird, daß das Diskrepanzmaß $d(f, g)$ und die korrigierenden Dissimilaritätsmaße $\text{dsc}_m(f, g)$ sich auf unterschiedliche Intervalle $[a_m, b_m]$ beziehen:

$$(5.23) \quad \text{dc}(f, g) = d(f, g)_{[a_0, b_0]} \prod_{m=1}^M \text{dsc}_{m, [a_m, b_m]}(f, g)$$

oder indem eine Kombination aus beiden Erweiterungen wie folgt gebildet wird:

$$(5.24) \quad \text{dc}(f, g) = d(f, g)_{[a_0, b_0]}^{w_0} \prod_{m=1}^M \text{dsc}_{m, [a_m, b_m]}^{w_m}(f, g).$$

Eine Vorgabe von Gewichtungsfaktoren erlaubt z.B. eine Feinsteuerung bezüglich der noch akzeptablen Abweichung der Peaks von ihren korrekten Positionen. Ein Bezug auf verschiedene Frequenzbänder $[a_m, b_m]$ ermöglicht es, qualitativ unterschiedliche Informationen aus verschiedenen Frequenzbändern zu verknüpfen.

Die letztgenannten Möglichkeiten verdeutlichen die Flexibilität des Grundansatzes bezüglich der Anpassung an spezielle Fragestellungen. Liegen solche speziellen Fragestellungen nicht vor, ist es sinnvoll, sich auf allgemein plausible Maße zu beschränken, so daß hier nur die folgenden korrigierten Diskrepanzmaße betrachtet werden, die auf dem korrigierten integrierten quadratischem Fehler (ISE) bzw. dem korrigierten Kullback-Leibler-Informationskriterium (KLIC) beruhen:

$$(5.25) \quad \text{dc}_{2,\Delta}(f, g) = d_2(f, g) \text{dsc}_\Delta(f, g),$$

$$(5.26) \quad \text{dc}_{2,\Delta^2}(f, g) = d_2(f, g) \text{dsc}_{\Delta^2}(f, g),$$

$$(5.27) \quad \text{dc}_{2,\Delta,\Delta^2}(f, g) = d_2(f, g) \text{dsc}_\Delta(f, g) \text{dsc}_{\Delta^2}(f, g),$$

$$(5.28) \quad \text{dc}_{KL,\Delta}(f, g) = d_{KL}(f, g) \text{dsc}_\Delta(f, g),$$

$$(5.29) \quad \text{dc}_{KL,\Delta^2}(f, g) = d_{KL}(f, g) \text{dsc}_{\Delta^2}(f, g),$$

$$(5.30) \quad \text{dc}_{KL,\Delta,\Delta^2}(f, g) = d_{KL}(f, g) \text{dsc}_\Delta(f, g) \text{dsc}_{\Delta^2}(f, g).$$

In Tabelle (5.6) und (5.7) sind für das Standardbeispiel des Vergleichs von f_0 mit f_k ($k = 1, \dots, 4$) die korrigierten Diskrepanzmaße sowie die zugehörigen Rangfolgen eingetragen. Sowohl bei dem integrierten quadratischen Fehlers $d_2(f, g)$ als auch bei dem Kullback-Leibler-Informationskriterium $d_{KL}(f, g)$ ist eine Korrektur nur mit Hilfe der korrigierenden Dissimilaritäten $\text{dsc}_\Delta(f_0, f_k)$ nicht ausreichend, um die vorgegebene Referenzrangfolge zu erhalten. Bei beiden Diskrepanzmaßen ergibt eine Korrektur mit den Dissimilaritäten $\text{dsc}_{\Delta^2}(f_0, f_k)$ bereits vollständige Übereinstimmung mit der Referenzrangfolge. Eine Doppelkorrektur sowohl mit $\text{dsc}_\Delta(f_0, f_k)$ als auch mit $\text{dsc}_{\Delta^2}(f_0, f_k)$ ergibt zusätzlich eine Verbesserung der relativen Größen der korrigierten Diskrepanzmaße.

Vor allem aufgrund der einfacheren numerischen Berechnung wird zur Auswertung der Simulationen in dieser Arbeit das doppelt korrigierte Diskrepanzmaß $dc_{2,\Delta,\Delta^2}(g, f)$ verwendet.

Selbstverständlich können für andere Anwendungsgebiete andere Gütekriterien besser geeignet sein, mit BASSEVILLE[1989] läßt sich die Problematik, ein geeignetes Diskrepanzmaß für eine spezielle Fragestellung zu finden, wie folgt abschließend zusammenfassen:

”It is quite difficult to give strong and definite recommendations about the choice of a distance measure for a given particular application. The fact that, even in the field of speech processing where this question has been investigated probably for the longest time, new variants of spectral distance measures still emerge each year, shows that this problem of choice is quite complex.”

ISE, korrigiert - Frequenzbereich			
$N_\omega = 512$			
		$dc_{2,\Delta}(f_0, f_k)$	Rang
Sub-Peak	f_1	0.12037	(4)
Cross-Peak	f_2	0.05530	(1)
Bias-Two-Peak	f_3	0.05788	(2)
Bias-Four-Peak	f_4	0.07609	(3)
		$dc_{2,\Delta^2}(f_0, f_k)$	Rang
Sub-Peak	f_1	1.31279	(4)
Cross-Peak	f_2	0.34525	(3)
Bias-Two-Peak	f_3	0.08232	(2)
Bias-Four-Peak	f_4	0.07792	(1)
		$dc_{2,\Delta,\Delta^2}(f_0, f_k)$	Rang
Sub-Peak	f_1	2.45674	(4)
Cross-Peak	f_2	0.54436	(3)
Bias-Two-Peak	f_3	0.09957	(2)
Bias-Four-Peak	f_4	0.07984	(1)

Tabelle 5.6: Vergleich von AR(8)-Prozessen im Frequenzbereich mit dem korrigierten integrierten quadratischen Fehler (ISE)

KLIC, korrigiert - Frequenzbereich			
$N_\omega = 1024$			
		$dc_{KL,\Delta}(f_0, f_k)$	Rang
Sub-Peak	f_1	4.33761	(4)
Cross-Peak	f_2	1.34750	(2)
Bias-Two-Peak	f_3	1.44061	(3)
Bias-Four-Peak	f_4	1.03147	(1)
		$dc_{KL,\Delta^2}(f_0, f_k)$	Rang
Sub-Peak	f_1	65.50155	(4)
Cross-Peak	f_2	11.44839	(3)
Bias-Two-Peak	f_3	2.29260	(2)
Bias-Four-Peak	f_4	1.03187	(1)
		$dc_{KL,\Delta,\Delta^2}(f_0, f_k)$	Rang
Sub-Peak	f_1	129.33259	(4)
Cross-Peak	f_2	19.04104	(3)
Bias-Two-Peak	f_3	2.92479	(2)
Bias-Four-Peak	f_4	1.04399	(1)

Tabelle 5.7: Vergleich von AR(8)-Prozessen im Frequenzbereich mit dem korrigierten Kullback-Leibler-Informationskriterium (KLIC).

Kapitel 6

Automatische Schätzung von Spektren

In diesem Kapitel werden Modellselektionskriterien sowohl für parametrische, autoregressive Spektralschätzungen als auch für nichtparametrische, direkte und indirekte Spektralschätzer vorgestellt und diskutiert.

Anschließend erfolgt im Rahmen einer Simulationsstudie mit verschiedenen datenerzeugenden Prozessen eine Evaluierung der Modellselektionskriterien für verschiedene Problemstellungen. Eine abschließende Evaluierung wird dann anhand jeweils einer Zeitreihe aus den Gebieten Ökonomie, Endokrinologie und Astronomie vorgenommen.

6.1 Modellselektionskriterien für autoregressive Spektralschätzungen

Für $AR(p)$ -Modelle sind seit langem Modellselektionskriterien in Gebrauch, die es erlauben, die Ordnung p an Hand der Daten zu bestimmen. Eine Übersicht über geläufige Kriterien geben z.B. ANDĚL[1982] oder SHIBATA[1985, 1989]. Aktuelle Erweiterungen und Modifikationen werden z.B. von HURVICH/TSAI[1989], SHIBATA[1989], HURVICH/SHUMWAY/TSAI[1990], HURVICH[1992] und HURVICH/TSAI[1993] diskutiert. Angesichts der großen Zahl von Modellselektionskriterien beschränkt sich die Darstellung hier auf einige relevante Kriterien, die jedoch typische Eigenschaften von verschiedenen Klassen von Modellselektionskriterien repräsentieren. Auf die teilweise sehr aufwendige mathematische Herleitung der Kriterien bzw. der Optimalitätseigen-

schaften wird verzichtet. Dagegen werden die unterschiedlichen modelltheoretischen Annahmen und Herangehensweisen bei der Herleitung von Optimalitätskriterien hervorgehoben.

Approximiert man den datenerzeugenden Prozeß einer Stichprobe (X_1, \dots, X_N) durch ein AR(p)-Modell,

$$X_t + \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} = \epsilon_t, \quad \{\epsilon_t\} \sim \text{WN}(0, \sigma^2)$$

und ermittelt für verschiedene Werte von p mit $0 \leq p \leq p_{max}$ Schätzungen der Modellparameter $(\hat{\phi}_1, \dots, \hat{\phi}_p, \hat{\sigma}^2)$, so können als Modellselektionskriterien für die Ordnung p u.a. die folgenden Größen herangezogen werden:

(Final Prediction Error, FPE)

$$(6.1) \quad \text{FPE}(p) = \hat{\sigma}^2 \frac{N+p}{N-p}$$

(Akaike Informationskriterium, AIC)

$$(6.2) \quad \text{AIC}(p) = N \ln \hat{\sigma}^2 + 2(p+1)$$

(Korrigiertes Akaike Informationskriterium, AICC)

$$(6.3) \quad \text{AICC}(p) = N \ln \hat{\sigma}^2 + \frac{2N(p+1)}{N-p-2}$$

$$(6.4) \quad = \text{AIC} + \frac{2(p+1)(p+2)}{N-p-2}$$

(Schwarz Informationskriterium)

$$(6.5) \quad \text{SIC}(p) = N \ln \hat{\sigma}^2 + p \ln N$$

(Hannan-Quinn-Kriterium)

$$(6.6) \quad \text{HQ}(p) = N \ln \hat{\sigma}^2 + 2cp \ln \ln N \quad c \geq 1.$$

Es ist jeweils diejenige Modellordnung p zu wählen, für die das ausgewählte Modellselektionskriterium minimal wird.

Das FPE-Kriterium wurde von AKAIKE[1969a, 1969b] entwickelt. Die Grundidee des FPE-Kriteriums beruht darauf, eine Schätzung des Einschnitt-Prognosefehlers für eine weitere unabhängige Realisation des datenerzeugenden Prozesses zu erhalten. Das AIC (vgl. AKAIKE[1974]) basiert auf dem Zusammenhang zwischen Maximum-Likelihood-Schätzung und der Minimierung der Kullback-Leibler-Information (vgl. Kapitel 2.6): Der Minimierung des AIC entspricht approximativ eine Minimierung der Kullback-Leibler-Information. Das AICC wurde von HURVICH/TSAI[1989] vorgeschlagen. Es beruht auf einer Bias-Korrektur des AIC-Kriteriums für kleine Stichprobenumfänge, die von SUGUIRA[1978] entwickelt wurde. Die Eigenschaften der Modellselektionskriterien bei kleinen Stichprobenumfängen und die Relevanz von Korrekturtermen wird weiter unten ausführlicher diskutiert. SCHWARZ[1978] entwickelte im Rahmen eines bayesianischen Ansatzes das nach ihm benannte SIC. Das gleiche Kriterium wird häufig auch als BIC (Bayes Informationskriterium) bezeichnet und entspricht im wesentlichen einem weiteren Kriterium, das von AKAIKE[1978a, 1979] als bayesianische Modifikation des AIC abgeleitet wurde. Das HQ-Kriterium stammt von Hannan/Quinn[1979]. In ihren Simulationsbeispielen setzen sie den Parameter $c = 1$, hier wird im weiteren ebenso verfahren.

Alle Kriterien werden in unterschiedlichen Varianten angewandt, die sich durch die Addition bzw. Subtraktion von Konstanten und/oder durch die Division bzw. Multiplikation mit dem Stichprobenumfang N voneinander unterscheiden. Da die Lage des Minimums durch diese Transformationen nicht verändert wird, sind all diese Varianten hinsichtlich der Ermittlung der optimalen Modellordnung äquivalent. Das FPE-Kriterium kann durch die Transformation

$$(6.7) \quad N \ln \text{FPE}(p) = N \ln \hat{\sigma}^2 + N \ln \frac{N+p}{N-p},$$

in eine Funktion überführt werden, die für den gleichen Wert p minimal wird wie das ursprüngliche FPE-Kriterium. Aus der Darstellung (6.7) kann abgeleitet werden, daß für $p \ll N$ gilt (vgl. z.B. PRIESTLEY[1981, S. 374]):

$$\text{AIC}(p) \approx N \ln \text{FPE}(p),$$

d.h. FPE-Kriterium und AIC sind asymptotisch äquivalent. Da sich aus der Darstellung (6.4) die asymptotische Äquivalenz von AIC und AICC ergibt, können das FPE-Kriterium, das AIC und das AICC als asymptotisch äquivalent betrachtet werden. Um

bei Graphiken soweit wie möglich einheitliche Größenordnungen verwenden zu können, wird das FPE-Kriterium in der transformierten Form gemäß (6.7) verwandt.

Unter Berücksichtigung der Darstellung (6.7) für das FPE-Kriterium ergibt sich, daß alle vorgestellten Modellselektionskriterien die gleiche Struktur besitzen. Sie bestehen im wesentlichen aus zwei Summanden: Der Term $\ln \hat{\sigma}^2$ mißt die Güte der Anpassung des AR(p)-Prozesses an die Daten. Bei festem N wird mit zunehmender Ordnung p die Anpassung an die Daten immer besser, $\ln \hat{\sigma}^2$ ist eine nicht-zunehmende Funktion der Modellordnung p . Der zweite Term enthält jeweils eine Funktion, die in Abhängigkeit von der Modellordnung p zunimmt. Dieser Term kann als Straffunktion für zusätzliche Parameter interpretiert werden.

Damit repräsentieren alle Modellselektionskriterien das gleiche grundlegende Prinzip der Modellauswahl, wonach einfache Modelle komplexen Modellen vorzuziehen sind („Prinzip der Sparsamkeit“, „Occams razor“). Es existiert allerdings kein einheitlicher Ansatz, dieses Prinzip für konkrete Fragestellungen quantitativ umzusetzen. Die Kriterien besitzen unterschiedliche Sensitivität der Straffunktionen bezüglich zusätzlicher Modellparameter. Daraus ergibt sich unmittelbar die Frage nach einer „optimalen“ Straffunktion bzw. einem „optimalen“ Modellselektionskriterium. Im Rahmen der Schätzung autoregressiver Prozesse bzw. der autoregressiven Spektralschätzung lassen sich drei Konzepte der Optimalität von Modellselektionskriterien unterscheiden, denen unterschiedliche Sichtweisen und modelltheoretische Annahmen zugrunde liegen. Diese drei Konzepte lassen sich kurz durch die Begriffe Ordnungskonsistenz, Effizienz sowie Diskrepanz-Optimalität charakterisieren. Im Rahmen der autoregressiven Spektralschätzung wäre es durchaus vertretbar, sich ausschließlich auf das Konzept der Diskrepanz-Optimalität im Frequenzbereich zu beschränken (vgl. Kapitel 5). Das Konzept der Diskrepanz-Optimalität schließt die Verwendung von „inkonsistenten“ oder „ineffizienten“ Modellselektionskriterien jedoch nicht aus, was Unklarheiten und Mißverständnisse hervorrufen kann. Aus diesem Grund werden alle drei Konzepte kurz skizziert.

Das mit am häufigsten angeführte Optimalitätskonzept ist die Ordnungskonsistenz. Diesem Konzept liegt die Vorstellung zugrunde, daß es sich bei dem datenerzeugenden Prozeß um einen AR-Prozeß mit endlicher Modellordnung p_0 handelt. Ein Modellselektionskriterium ist (stark) ordnungskonsistent, wenn für $N \rightarrow \infty$ diese „wahre“ Modellordnung mit Wahrscheinlichkeit Eins gewählt wird. Nach einem Satz von Shi-

bata[1976] überschätzt das AIC asymptotisch die wahre Modellordnung p_0 eines datenerzeugenden $AR(p_0)$ -Prozesses in dem Sinne, daß die Wahrscheinlichkeit, daß das AIC Modellordnungen $\hat{p} > p_0$ ermittelt, asymptotisch von Null verschieden ist. Von den angeführten Modellselektionskriterien sind das FPE-Kriterium, AIC und AICC ordnungsinkonsistent, während das SIC sowie das HQ-Kriterium ordnungskonsistent sind (vgl. z.B. HURVICH/TSAI[1989]).

Die praktische und theoretische Relevanz des Konzeptes der Ordnungskonsistenz ist allerdings zweifelhaft. Als unmittelbarer Einwand kann vorgebracht werden, daß bei praktischen Anwendungen die Annahme eines datenerzeugenden $AR(p_0)$ -Prozesses zu restriktiv ist, vielmehr dient ein $AR(p_0)$ -Modell lediglich als einfaches approximierendes Modell für einen unbekanntem datenerzeugenden Prozeß. Eine „wahre“ Modellordnung ist in diesem allgemeineren Rahmen nicht sinnvoll ermittelbar. Es kann nur festgestellt werden, ob die durch ein Modellselektionskriterium ermittelte Modellordnung im wesentlichen den praktischen Erfordernissen entspricht. Die Ergebnisse bei praktischen Anwendungen sind widersprüchlich. STEINBERG/GASSER/FRANKE[1985] bestätigen im Rahmen von EEG-Studien, daß das AIC, gemessen an empirischen neurophysiologischen Standards, „zu große“ Modellordnungen liefert. Dagegen verweist KAYSHAP[1980] im Kontext von Studien zur Sprachdatenverarbeitung darauf, daß das AIC bei Stichprobenumfängen von $N = 200$ bis $N = 300$ gewöhnlich Modellordnungen zwischen $\hat{p} = 2$ und $\hat{p} = 5$ schätzt. Audiophysologisch befriedigende Ergebnisse können aber erst bei Modellordnungen zwischen $p = 8$ und $p = 14$ erzielt werden. Im Rahmen geophysikalischer Untersuchungen berichten z.B. LANDERS/LACOSS[1977], daß u.a. AIC und FPE-Kriterium Modellordnungen ermitteln, die unter den Ordnungen liegen, die nach physikalischen Kriterien als sinnvoll erachtet werden. Auch wenn man sich auf datenerzeugende $AR(p_0)$ -Prozesse beschränkt, kann aus dem asymptotischen Ergebnis von SHIBATA[1976] bei endlichen Stichprobenumfängen nicht gefolgert werden, daß das AIC dazu tendiert, die wahre Modellordnung zu überschätzen. Trivialerweise läßt sich für jeden endlichen Stichprobenumfang N ein datenerzeugender $AR(p_0)$ -Prozeß mit $p_0 > N$ angeben, so daß die „wahre“ Modellordnung zwangsläufig unterschätzt werden muß. Dementsprechend lehnen SAKAMOTO/ISHIGURO/KITAGAWA[1986, S. 85] für endliche Stichprobenumfänge das Konzept einer „wahren“ Modellordnung als bedeutungslos ab. Diese Ansicht kann auch im Rahmen der Simulationsstudie im Abschnitt 6.3 bestätigt werden. Noch grundsätzlicher kritisiert DAHLHAUS[1987] das Konzept der Ordnungskonsistenz:

„For me it doesn't make sense to consider the distance of orders (which is done by proving convergence in probability), since this distance of orders doesn't say very much about the "nearness" of the processes themselves (an AR(2) process with moderate roots may be much "nearer" to, e.g., an AR(1000) process, than to an AR(3) process with roots close to the unit circle).“

Damit wird hervorgehoben, daß auch innerhalb der Klasse von AR-Prozessen mit endlicher Ordnung p_0 die Modellordnung p_0 keine sinnvollen metrischen Eigenschaften besitzt: AR-Prozesse gleicher Ordnung können sehr unterschiedliche Autokovarianzfunktionen bzw. Spektren besitzen während AR-Prozesse mit unterschiedlicher Modellordnung andererseits sehr ähnliche Autokovarianzfunktionen bzw. Spektren haben können. Das Konzept der Ordnungskonsistenz scheint im allgemeinen nicht zur Beurteilung von Modellselektionskriterien geeignet zu sein.

Das Konzept der asymptotischen Effizienz zur Beurteilung von Modellselektionskriterien ist auf SHIBATA[1980] zurückzuführen. Er untersucht die Problemstellung, daß ein datenerzeugender Gauss'scher AR-Prozeß von unendlicher Ordnung möglichst gut durch einen AR-Prozeß von endlicher Ordnung approximiert werden soll. Wie bei dem FPE-Kriterium ist bei ihm der mittlere quadratische Einschnitt-Prognosefehler für eine weitere unabhängige Realisation des datenerzeugenden Prozesses maßgeblich. Er leitet im Rahmen dieser Problemstellung eine untere Schranke für den mittleren quadratischen Einschnitt-Prognosefehler ab. Ein Modellselektionskriterium ist asymptotisch effizient, wenn es für $N \rightarrow \infty$ gegen diese untere Schranke des mittleren quadratischen Einschnitt-Prognosefehlers in Wahrscheinlichkeit konvergiert. SHIBATA[1980] zeigt, daß das AIC und FPE-Kriterium asymptotisch effizient sind. Ebenso ist das AICC asymptotisch effizient (vgl. HURVICH/TSAI[1989]). Darüberhinaus zeigt SHIBATA[1980], daß Modellselektionskriterien die ordnungskonsistent sind, nicht gleichzeitig asymptotisch effizient sein können. Ordnungskonsistenz kann nur durch Verzicht auf asymptotische Effizienz erreicht werden. Demnach sind SIC und das HQ-Kriterium nicht asymptotisch effizient. Es muß aber berücksichtigt werden, daß das Konzept der asymptotischen Effizienz den mittleren quadratischen Einschnitt-Prognosefehler als Optimierungskriterium verwendet und damit auf einem sehr speziellen Fehlermaß beruht. Es ist nicht auszuschließen, daß für nicht prognoseorientierte Modellierung mit den ineffizienten Kriterien wie SIC oder dem HQ-Kriterium bessere Ergebnisse erzielt werden können wie

mit den effizienten Kriterien FPE, AIC bzw. AICC. Darüberhinaus ist zu beachten, daß selbst wenn das Ziel der Modellbildung die Erstellung einer Prognose ist, keine Einigkeit darüber besteht, welches Maß für den Prognosefehler am besten geeignet ist (vgl. z.B. AHLBERG[1992], CHATFIELD[1992]).

Das für die autoregressive Spektralschätzung wichtigste Optimalitätskonzept ist die Diskrepanz-Optimalität, wobei die Eignung eines Modellselektionskriteriums unmittelbar im Frequenzbereich mit Hilfe von Diskrepanzmaßen $d(f, \hat{f})$ bewertet wird. Bezüglich der Optimalität der Modellselektionskriterien liegen im wesentlichen zwei Ergebnisse für klassische Diskrepanzmaße vor. SHIBATA[1981] zeigt unter der Grundannahme eines datenerzeugenden Gauss'schen AR-Prozesses mit unendlicher Ordnung, daß bei Anwendung des AIC zur Bestimmung der Modellordnung eines approximierenden $AR(p)$ -Prozesses, asymptotisch der integrierte relative quadratische Fehler (ISER) in Wahrscheinlichkeit gegen eine untere Schranke konvergiert. HURVICH/TSAI[1989] unterstellen einen Gauss'schen AR-Prozeß mit endlicher Ordnung als datenerzeugenden Prozeß und zeigen, daß die Minimierung des AICC approximativ der Minimierung der Kullback-Leibler-Information (KLIC) entspricht. Da das FPE-Kriterium, AIC und AICC asymptotisch äquivalente Modellselektionskriterien sind, kann davon ausgegangen werden, daß die jeweils für AIC bzw. AICC abgeleiteten Aussagen bei großen Stichprobenumfängen für alle drei Kriterien gelten. Wie bereits in Kapitel 5 gezeigt worden war, sind sowohl der integrierte relative quadratische Fehler (ISER) als auch die Kullback-Leibler-Information (KLIC) als Diskrepanzmaße für Spektraldichtefunktionen problematisch: Für beide Maße können Beispiele angegeben werden, die zeigen, daß ihre Anwendung zu Aussagen führen, die spektralanalytisch nicht sinnvoll sind. Eine ausführlichere kritische Diskussion war bereits in Kapitel 5 erfolgt. In Kapitel 5 war auch darauf hingewiesen worden, daß für unterschiedliche Problemstellungen unterschiedliche Diskrepanzmaße geeignet sein können. Auch wenn für das SIC und das HQ-Kriterium keine Optimalitätsbeweise vorliegen, ist damit nicht ausgeschlossen, daß diese Kriterien optimal bezüglich eines weniger gebräuchlichen Diskrepanzmaßes sein können.

Die bisherige Diskussion der Optimalität von Modellselektionskriterien läßt sich damit zunächst wie folgt zusammenfassen: Das Konzept der Ordnungskonsistenz ist zur Beurteilung von Modellselektionskriterien im allgemeinen nicht geeignet. Weder das Konzept der Effizienz noch das Konzept der Diskrepanz-Optimalität liefert klare theo-

retische Anhaltspunkte dafür, daß allgemein ein bestimmtes Kriterium den anderen vorgezogen werden kann. Im Rahmen dieser Arbeit werden die Modellselektionskriterien für die autoregressive Spektralschätzung unter Verwendung der global korrigierten Diskrepanz $dc_{2,\Delta,\Delta^2}(f, \hat{f}_p)$ (vgl. Abschnitt 5) mit Hilfe von Simulationsstudien evaluiert werden.

Ein weiteres Problemfeld bei der Beurteilung von Modellselektionskriterien ergibt sich, wenn nur relativ kleine Stichprobenumfänge vorliegen (vgl. HURVICH/TSAI[1989]). Im vorangegangenen Abschnitt war davon ausgegangen worden, daß eine optimale Modellordnung p in einem Bereich $0 \leq p \leq p_{max}$ gesucht wird, wobei der Parameter p_{max} geeignet vorgegeben werden muß. Hierzu gibt es verschiedene Vorschläge. Urych/Bishop[1975] raten aufgrund von Simulationsstudien, mit denen sie Eigenschaften des AIC und des FPE-Kriteriums überprüften, zu einer oberen Grenze in Abhängigkeit vom Stichprobenumfang mit $p_{max} \approx N/3$. SAKAMOTO/ISHIGURO/KITAGAWA[1986, S. 83] empfehlen im allgemeinen bei Anwendung des AIC den Parameter p_{max} in Abhängigkeit vom Stichprobenumfang N als $p_{max} \approx 2\sqrt{N}$ festzulegen. Als oberste Grenze geben sie $p_{max} \approx N/2$ an. Der Grund liegt darin, daß bei gegebenem Stichprobenumfang N und zunehmender Modellordnung p das AIC sowie auch andere Modellselektionskriterien ab $p \approx N/2$ fallende Funktionen in Abhängigkeit von p sind und damit erheblich von dem idealtypischen, in etwa U-förmigen bzw. V-förmigen¹ Verlauf abweichen können, der für Modellselektionskriterien im allgemeinen unterstellt wird. Das Problem sowie eine Lösung unter Verwendung von korrigierten Modellselektionskriterien soll unter Verwendung des AR(4)-Prozesses

$$(6.8) \quad X_t + 0.18287X_{t-1} + 0.51338X_{t-2} + 0.14813X_{t-3} + 0.65610X_{t-4} = \epsilon_t,$$

$$\{\epsilon_t\} \sim \text{WN}(0, 1).$$

veranschaulicht werden. Die Wurzeln des charakteristischen Polynoms des AR(4)-Prozesses (6.8) lauten:

$$z_{1/2} = 0.9e^{\pm i2\pi 0.15} \quad z_{3/4} = 0.9e^{\pm i2\pi 0.35}.$$

¹In Abhängigkeit von dem datenerzeugenden Prozeß sind auch andere Kurvenverläufe möglich. Diese Fälle sind für die Diskussion hier nicht relevant.

In Abb. 6.1 (a) ist das theoretische Spektrum des AR(4)-Prozesses (6.8) wiedergegeben. In Abb. 6.1 (b) sind Schätzungen für $E[\text{dc}_{2,\Delta,\Delta^2}(f, \hat{f}_p)]$ mit $p = 1, 2, \dots, 30$ wiedergegeben. Hierzu wurden 100 Realisationen des AR(4)-Prozesses (6.8) mit einem Stichprobenumfang von $N = 32$ erzeugt, wobei jeweils $N_E = 128$ Einschwingwerte berücksichtigt wurden. Für jede Realisation wurden mit Hilfe des Burg-Schätzers AR-Prozesse mit den Ordnungen $p = 1, 2, \dots, 30$ angepaßt sowie die autoregressiven Spektralschätzungen $\hat{f}_p(\omega)$ und die global korrigierten Diskrepanzen $\text{dc}_{2,\Delta,\Delta^2}(f, \hat{f}_p)$ berechnet. Das theoretische Spektrum und die geschätzten Spektren wurden im Bereich $0 \leq \omega \leq 0.5$ an $N_\omega = 128$ äquidistanten Frequenzpunkten berechnet. Zur Schätzung von $E[\text{dc}_{2,\Delta,\Delta^2}(f, \hat{f}_p)]$ wurde für jede Ordnung p über die 100 Realisationen von $\text{dc}_{2,\Delta,\Delta^2}(f, \hat{f}_p)$ gemittelt. Das Ergebnis ist in Abbildung 6.1 (b) dargestellt und ergibt eine idealtypisch V-förmig verlaufende Diskrepanz mit einem einzigen globalen Minimum bei der Ordnung $p = 4$. Der Kurvenverlauf von Schätzungen des Erwartungswertes der global korrigierten Diskrepanzen $\text{dc}_{2,\Delta,\Delta^2}(f, \hat{f}_p)$ soll als grobe Referenz zur graphischen Beurteilung der analog zu berechnenden Kurvenverläufe der Modellsselektionskriterien dienen. Da bei den global korrigierten Diskrepanzen $\text{dc}_{2,\Delta,\Delta^2}(f, \hat{f}_p)$ sehr kleine und sehr große Werte auftreten, erfolgt ihre graphische Darstellung in logarithmierter Form.

Die Abbildungen 6.2 (a), 6.2 (c), 6.2 (e) und 6.2 (g) enthalten die Verlaufskurven für Schätzungen der Erwartungswerte von $\text{AIC}(p)$, $\text{SIC}(p)$, $\text{HQ}(p)$ bzw. $\text{FPE}(p)$ für $p = 1, 2, \dots, 30$. Die Schätzungen wurden analog wie für $E[\text{dc}_{2,\Delta,\Delta^2}(f, \hat{f}_p)]$ anhand derselben 100 Realisationen des AR(4)-Prozesses (6.8) berechnet. Der von SAKAMOTO/ISHIGURO/KITAGAWA[1986, S. 83] für das AIC beschriebene Effekt läßt sich auch bei den anderen drei unkorrigierten Modellsselektionskriterien beobachten. Alle unkorrigierten Kriterien besitzen ein lokales Minimum bei $p = 4$. Ab ca. $p = N/2 = 16$ sind alle Kriterien eine fallende Funktion in Abhängigkeit von der Modellordnung p , ein globales Minimum wird von allen unkorrigierten Kriterien bei der Modellordnung $p_{max} = 30$ erreicht. Damit weichen die Kurvenverläufe aller unkorrigierten Kriterien erheblich von der wünschenswerten idealtypischen Form ab (vgl. Abb. 6.1). HURVICH/TSAI[1989] demonstrierten diesen Effekt in ähnlicher Weise für verschiedene unkorrigierte Modellsselektionskriterien mit Hilfe eines AR(2)-Prozesses. Als Abhilfe schlugen sie das AICC vor, bei dem wie Abb. 6.2 (b) zeigt, dieser Effekt nicht auftritt. Als experimentelle Erweiterung des Ansatzes von HURVICH/TSAI[1989] werden hier ad-hoc Korrekturen auch für das SIC, das HQ- und das FPE-Kriterium eingeführt.

Hierzu wird zu den Ausgangskriterien einfach der Korrekturterm des AICC addiert (vgl. (6.4)). Die neuen korrigierten Kriterien ergeben sich damit als:

$$(6.9) \quad \text{SICC}(p) = \text{SIC}(p) + \frac{2(p+1)(p+2)}{N-p-2},$$

$$(6.10) \quad \text{HQC}(p) = \text{HQ}(p) + \frac{2(p+1)(p+2)}{N-p-2},$$

$$(6.11) \quad \text{FPEC}(p) = N \ln \text{FPE}(p) + \frac{2(p+1)(p+2)}{N-p-2}.$$

Ein Vergleich der Kurvenverläufe der korrigierten Modellselektionskriterien $\text{AICC}(p)$, $\text{SICC}(p)$, $\text{HQC}(p)$ und $\text{FPEC}(p)$ in den Abbildungen 6.2 (b), 6.2 (d), 6.2 (f) bzw. 6.2 (h) zeigt, daß die Addition des Korrekturterms bei allen Modellselektionskriterien den gleichen Effekt bewirkt: Alle korrigierten Modellselektionskriterien besitzen ein globales Minimum bei $p = 4$ und steigen danach mit zunehmender Modellordnung in etwa exponentiell an. Damit entspricht der Kurvenverlauf der korrigierten Kriterien in sehr guter Näherung der wünschenswerten idealtypischen Form. Bei den korrigierten Kriterien ist zu erwarten, daß die Schätzung der optimalen Modellordnung weitgehend unabhängig von dem Parameter p_{max} ist. Dies ist bei den unkorrigierten Kriterien nicht der Fall. Experimentell wurden noch weitere verschiedene Korrekturfaktoren ausprobiert. Ähnliche Ergebnisse wie mit dem Korrekturterm des AICC ergaben sich z.B. mit den Korrekturfaktoren $c_1 = p^{2.5}/N$ oder $c_2 = p^2/\ln N$. Auf eine graphische Darstellung dieser Varianten wird verzichtet. Bisher konnte allerdings lediglich die korrigierte Form AICC theoretisch als Biaskorrektur des AIC begründet werden (vgl. SUGUIRA[1978], HURVICH/TSAI[1989]). Die hier vorgestellten experimentellen Ergebnisse verweisen allerdings darauf, daß auch SIC, HQ- und FPE-Kriterium bei kleinen Stichprobenumfängen einen Bias aufweisen, der ähnlich wie im Falle des AIC einer theoretischen Analyse zugänglich sein sollte. Bei der Interpretation der Ergebnisse muß beachtet werden, daß die gemittelten Verlaufskurven lediglich die grundlegende Tendenz wiedergeben. Auch wenn die Verlaufskurven z.B. nur für $1 \leq p \leq 10$ berechnet wurden, war häufig bei einzelnen Realisationen kein klar ausgeprägtes Minimum bei $p = 4$ ersichtlich, wobei die unkorrigierten Modellselektionskriterien einen problematischeren Eindruck hinterließen als die korrigierten Kriterien.

Zusammenfassend kann festgestellt werden: Mit Hilfe der verschiedenen Konzepte der Optimalität ist keine klare theoretische Aussage darüber möglich, ob allgemein ein Modellselektionskriterium den anderen Kriterien vorzuziehen ist. Die Ergebnisse von HURVICH/TSAI[1989]) sowie die experimentellen Ergebnisse hier legen jedoch die Aussage nahe, daß korrigierte Modellselektionskriterien bei kleinen Stichprobenumfängen grundsätzlich vorzuziehen sind.

6.2 Modellselektionskriterien für nichtparametrische Spektralschätzungen

Modellselektionskriterien für nichtparametrische Spektralschätzungen werden fast genauso lange in der Literatur diskutiert wie Modellselektionsverfahren für autoregressive Schätzverfahren. Dennoch haben sie bei weitem nicht die gleiche Akzeptanz und Verbreitung finden können. Dies kann zum einen damit begründet werden, daß einige der ersten nichtparametrischen Modellselektionskriterien für Spektralschätzungen inhaltlich eng an das Konzept von Glättungsplines gebunden waren (vgl. z.B. COGBURN/DAVIS[1974], WHABA/WOLD[1975a, 1975b], BYRD/TAPIA/THOMPSON[1978], WHABA[1980]). Damit waren sie losgelöst von den üblichen Standardverfahren der nichtparametrischen Spektralschätzung. Darüberhinaus basieren diese Modellselektionskriterien meist auf dem Konzept der Kreuzvalidierung. Sofern nicht effiziente Algorithmen zur Verfügung stehen, kann dies mit erheblichem Rechenaufwand verbunden sein. Im weiteren wird zunächst das Konzept der Kreuzvalidierung vorgestellt. Anschließend werden zwei Kreuzvalidierungskriterien für direkte Spektralschätzer diskutiert und ihre Eigenschaften mit Hilfe einer kleinen Simulationsstudie beschrieben. Abschließend werden sog. strafende nichtparametrische Modellselektionskriterien vorgestellt, die für direkte und indirekte Spektralschätzungen gleichermaßen geeignet sind und geringen Rechenaufwand erfordern.

Das Grundkonzept von Kreuzvalidierungsverfahren besteht darin, eine Stichprobe in zwei Untergruppen zu teilen (vgl. z.B. STONE[1974]): Eine Untergruppe wird zur Anpassung verschiedener Modelle benutzt. Es wird dann dasjenige Modell gewählt, welches die Daten der zweiten Untergruppe am besten reproduzieren kann. Eine wichtige Verfeinerung ist das Konzept der "Leave-One-Out"-Kreuzvalidierung. Statt die Stichprobe in zwei Gruppen aufzuteilen, wird jeweils eine Beobachtung weggelassen und für die restliche Stichprobe ein Modell angepaßt. Mit dem angepaßten Modell wird der Prognosefehler für die ausgelassene Beobachtung ermittelt. Einen Kreuzvalidierungsindex für ein Modell erhält man, indem man über alle "Leave-One-Out"-Prognosefehler summiert. Es wird dasjenige Modell gewählt, für das der Kreuzvalidierungsindex ein Minimum annimmt.

Das Konzept der Kreuzvalidierung setzt voraus, daß die Beobachtungen Realisationen von unabhängigen Zufallsvariablen sind und kann damit nicht unmittelbar auf Zeit-

reihen angewandt werden. Kreuzvalidierungskonzepte zur Spektralschätzung basieren daher auf den zumindest asymptotisch unkorrelierten Werten des Periodogramms $I(\omega_j)$ (vgl. Satz 4.1.1). Auf die Darstellung sehr spezieller Kreuzvalidierungsansätze zur automatischen Spektralschätzung, wie z.B. von WHABA/WOLD[1975b] oder WHABA[1980] wird verzichtet. Die Darstellung hier beschränkt sich vielmehr auf Weiterentwicklungen von HURVICH[1985] und BELTRÃO/BLOOMFIELD[1987]. In ihren Arbeiten werden verschiedene Kreuzvalidierungskriterien vorgeschlagen, die auf "Leave-One-Out"-Periodogrammen basieren und unmittelbar für direkte Spektralschätzungen verwandt werden können. Die Darstellung hier beschränkt sich auf die einfachste Variante eines "Leave-One-Out"-Periodogrammes, die auch von BELTRÃO/BLOOMFIELD[1987] zugrundegelegt wird. In HURVICH[1985] sind weitere Variationen von "Leave-One-Out"-Periodogrammen beschrieben.

Für direkte "Leave-One-Out"-Spektralschätzungen,

$$\hat{f}_\nu^{-j}(\omega_k) = \sum_{r=-m_N}^{m_N} g_r I^{-j}(\omega_{k-r}),$$

basierend auf dem "Leave-One-Out"-Periodogramm,

(6.12)

$$I^{-j}(\omega_k) = \begin{cases} 0.5[I(\omega_{k-1}) + I(\omega_{k+1})] & j = k \\ I(\omega_k) & j \neq k \end{cases} \quad \omega_k = k/N, \omega_k \in (0.0, 0.5),$$

lassen sich u.a. die beiden folgenden Modellselektionskriterien angeben:

(Whaba-Hurvich-Kriterium, WH)

$$(6.13) \quad \text{WH}(\nu) = \frac{1}{n_\omega} \sum_{0.0 < \omega_k < 0.5} \left[\ln \hat{f}_\nu^{-j}(\omega_k) - [\ln I(\omega_k) + 0.57721] \right]^2$$

und

(Cross-Validated-Log-Likelihood-Kriterium, CVLL)

$$(6.14) \quad \text{CVLL}(\nu) = \frac{1}{n_\omega} \sum_{0.0 < \omega_k < 0.5} \left[\ln \hat{f}_\nu^{-j}(\omega_k) + \frac{I(\omega_k)}{\hat{f}_\nu^{-j}(\omega_k)} \right].$$

In (6.13) bzw. (6.14) bezeichnet n_ω die Anzahl der Fourierfrequenzen im Intervall $0.0 < \omega_k < 0.5$. Die Euler-Konstante 0.57721 in (6.13) dient zur Bias-Korrektur des

logarithmierten Periodogramms (vgl. z.B. WHABA[1980]). Die Kreuzvalidierungskriterien wurden in Abhängigkeit von ν dargestellt, um auszudrücken, daß als nichtparametrischer Modellselektionsparameter die äquivalente Zahl von Freiheitsgraden (EDF bzw. ν) herangezogen wird. Die äquivalente Zahl von Freiheitsgraden ν war für direkte Spektralschätzer bereits in Abschnitt 4.2 definiert worden als:

$$(6.15) \quad \nu = \frac{2}{\sum_{r=-m_N}^{m_N} g_r^2}.$$

Die äquivalente Bandbreite (EBW) hätte ebenso gut als Modellselektionsparameter herangezogen werden können. Da die äquivalente Zahl von Freiheitsgraden (EDF) aber bereits in der Literatur im Zusammenhang mit weiteren nichtparametrischen Modellselektionskriterien als Selektionsparameter herangezogen wurde (vgl. z.B. HURVICH/BELTRÃO[1990]), soll dies auch hier beibehalten werden.

Als optimale äquivalente Zahl von Freiheitsgraden wird diejenige Zahl ν gewählt, für die das gewählte Kreuzvalidierungskriterium ein Minimum annimmt. Die endgültige Spektralschätzung erfolgt dann unter Verwendung des üblichen Periodogramms.

Das WH-Kriterium beruht auf einer Verallgemeinerung des Kreuzvalidierungsansatzes von WHABA[1980] durch HURVICH[1985]. Es minimiert asymptotisch den Erwartungswert des integrierten quadratischen Fehlers (ISE). Das CVLL-Kriterium wurde von BELTRÃO/BLOOMFIELD[1987] vorgeschlagen und beruht auf der Approximation der Loglikelihoodfunktion eines Gauss'schen Prozesses im Frequenzbereich. Sie zeigten, daß dieses Kriterium asymptotisch den Erwartungswert des integrierten relativen Fehlers (ISER) minimiert. Das CVLL-Kriterium ist minimierungsäquivalent zur diskretisierten Kreuzentropie $S(\hat{f}_\nu^{-j}(\omega_k), I(\omega_k))$, so daß die Minimierung des CVLL-Kriteriums asymptotisch zu den gleichen Ergebnissen führen sollte, wie die Minimierung der Kullback-Leibler-Information (KLIC) (vgl. Abschnitt 2.6). Daß dies tatsächlich der Fall ist, zeigen HURVICH/BELTRÃO[1990]. Darüberhinaus benutzen sie den Zusammenhang des CVLL-Kriteriums mit der Kullback-Leibler-Information um ein weiteres Modellselektionskriterium abzuleiten, das nicht auf Kreuzvalidierung beruht. Hier auf wird weiter unten eingegangen.

Im Zusammenhang mit Kreuzvalidierungsverfahren sind generell folgende Probleme zu beachten (vgl. z.B. MARRON[1988]): Ein Kreuzvalidierungskriterium kann eine große Stichprobenvariabilität besitzen, d.h. es können für den gleichen datenerzeugenden Prozeß stark unterschiedliche Glättungsparameter ausgewiesen werden. Darüberhinaus

können langsame Konvergenzeigenschaften große Stichprobenumfänge erforderlich machen. Das Auftreten von multiplen lokalen Minima in Kreuzvalidierungskriterien wird in THOMPSON/KAY/TITTERINGTON[1989] und HALL/MARRON[1991] analysiert. Für den Einsatz der Kreuzvalidierung im Rahmen der Spektralschätzung bestätigen BELTRÃO/BLOOMFIELD[1987] das Problem der großen Stichprobenvariabilität sowie das Auftreten von multiplen lokalen Minima in der Kreuzvalidierungsfunktion $CVLL(\nu)$. Sie hatten für ihre Simulationstudie den direkten Daniellschätzer verwandt und einen Stichprobenumfang von $N = 128$ zugrundegelegt. Von HURVICH[1985] liegen keine Angaben zu diesen Problemgebieten vor.

Für bestimmte Anwendungsgebiete der Kreuzvalidierung, wie z.B. der nichtparametrischen Regression mit Glättungssplines, stehen effiziente Algorithmen zur Verfügung (vgl. z.B. FUNKE[1991]). Dies gilt nicht für die hier vorgestellten Kreuzvalidierungsverfahren im Frequenzbereich. Somit besteht ein weiteres Problem bei der Anwendung dieser Verfahren darin, daß ein sehr großer Rechenaufwand erforderlich sein kann. Für die Realisation eines reellwertigen stochastischen Prozesses mit Stichprobenumfang $N = 256$ ergeben sich z.B. 127 "Leave-One-Out"-Spektralschätzungen \hat{f}_ν^{-j} . Berechnet man z.B. die Kreuzvalidierungsfunktion für $N_\nu = 100$ verschiedene Werte, so müssen bereits für eine Realisation 12700 Spektralschätzungen berechnet werden. Soll ein Kreuzvalidierungskriterium mit Hilfe einer Simulation von 100 Wiederholungen untersucht werden, so müssen bereits über eine Million Spektren, nämlich $12700 \times 100 = 1270000$ berechnet werden. So verwundert es nicht, daß sich z.B. HURVICH[1985] in seiner Simulationsstudie auf die Werte $N_\nu = 24$ und $n_\omega = 49$ bei 40 Versuchswiederholungen beschränkt. Da wie oben bereits angegeben wurde, das Auftreten von lokalen Minima nicht ausgeschlossen werden kann, können auch keine schnellen Optimierungsalgorithmen wie z.B. der Golden-Section-Search-Algorithmus (vgl. z. B. PRESS ET AL.[1992, S. 397ff]) eingesetzt werden, was die Anzahl der Funktionsbewertungen N_ν erheblich reduzieren könnte. Vielmehr ist es erforderlich, für ein Intervall $[\nu_{min}, \nu_{max}]$ die Kreuzvalidierungskriterien für eine hinreichend große Zahl von Werten zu berechnen. Da ein Hauptanwendungsgebiet von automatischen Spektralschätzungen in der standardisierten Bearbeitung von Massendaten liegt, kommt dem Faktor Rechenaufwand erhebliche Bedeutung zu. Um genauere Vorstellungen von dem tatsächlich erforderlichen Zeitaufwand für die Berechnung der Kreuzvalidierungskriterien zu erhalten, wurden zunächst einzelne Realisationen des AR(4)-Prozesses (6.8) aus Abschnitt 6.1 herangezogen. Dabei wurden die folgenden Werte verwandt:

$N = 256$, $N_E = 128$, $n_\omega = 127$, $N_\nu = 75$. Der Zeitaufwand für die Berechnung der Kreuzvalidierungskriterien (6.13) und (6.14) an 75 Punkten variierte in Abhängigkeit von der gewählten Methode (Daniell-Schätzer, iterative Hamming-Filterung) und der spezifischen Implementierung zwischen ca. 15 und 30 Minuten. Die Berechnungen wurden mit GAUSS 3.0 auf einem 386-PC (33 MHz) durchgeführt. Für eine Simulation mit 100 Wiederholungen ergibt sich damit im günstigsten Fall, wenn für eine Realisation nur 15 Minuten Rechenzeit benötigt werden, eine Gesamtrechendauer von ca. 25 Stunden. Auch wenn sich mit gängigen, leistungsfähigeren Prozessoren die Rechenzeiten auf ein Drittel reduzieren lassen, bleibt der Aufwand beachtlich. Dies gilt erst recht für längere Zeitreihen. Während lange Zeitreihen früher nur in den Naturwissenschaften zu finden waren, so stehen heute auch für ökonomische Fragestellungen sehr lange Zeitreihen zur Verfügung. So hat z.B. die Firma Olsen & Associates aus Zürich für ökonomische Forschungszwecke den Hochfrequenzdatensatz HFDF93 zusammengestellt, bei dem einzelne Zeitreihen mehrere hunderttausend Beobachtungen umfassen. Selbst nach zeitlicher Aggregation können diese Reihen leicht mehrere tausend Werte umfassen. Grundsätzlich scheinen die "Leave-One-Out"-Kreuzvalidierungsverfahren im Frequenzbereich durch folgendes Dilemma gekennzeichnet zu sein: Aufgrund des Rechenaufwandes sind sie eher für kleine Stichprobenumfänge geeignet. Dem steht das Problem der Stichprobenvariabilität entgegen, was wiederum große Stichprobenumfänge erforderlich machen kann. Detaillierte Untersuchungen hierzu liegen nicht vor.

Ein weiteres Problem der "Leave-One-Out"-Kreuzvalidierungskriterien besteht darin, daß sie unmittelbar nur mit direkten Spektralschätzern angewandt werden können. HURVICH[1985] erweitert das Konzept auch auf autoregressive und indirekte Spektralschätzer indem er u.a. mit einem geeigneten "Leave-One-Out"-Periodogramm unter Verwendung der inversen Fourier-Transformation "Leave-One-Out"-Autokovarianzen definiert. Da wie gezeigt wurde, bereits "Leave-One-Out"-Kreuzvalidierung bei direkten Spektralschätzungen sehr aufwendig sein können, wird dieser Ansatz hier nicht weiter verfolgt. Wie weiter unten ausgeführt wird, existiert eine Klasse von Modellselektionskriterien, die ohne großen Rechenaufwand sowohl mit direkten als auch mit indirekten Spektralschätzern angewandt werden können.

Da für die Simulationsstudien in Abschnitt (6.3) keine Kreuzvalidierungskriterien verwandt werden, sollen das WH- und CVLL-Kriterium hier zumindest mit Hilfe einer kleinen Simulationstudie charakterisiert werden. Hierzu wurde auf den AR(4)-Prozeß (6.8)

aus Abschnitt 6.1 zurückgegriffen und 40 Realisationen der Länge $N = 256$ erzeugt, wobei jeweils 128 Einschwingwerte berücksichtigt wurden. Es wurden unter Verwendung der iterativen Hamming-Filterung direkte Spektralschätzungen berechnet. Das Periodogramm, von dem die "Leave-One-Out"-Periodogramme abgeleitet wurden, war jeweils mit Hilfe der schnellen Fouriertransformation berechnet worden. Bei Anwendung der iterativen Hamming-Filterung wird der Modellselektionsparameter ν lediglich indirekt über die Anzahl der Iterationen vorgegeben. Der Zusammenhang zwischen der Anzahl der iterativen Glättungen und den entsprechenden äquivalenten Zahlen von Freiheitsgraden ist in Abbildung 6.3 dargestellt. Einer einmaligen Glättung mit dem Hamming-Filter entspricht eine äquivalente Zahl von Freiheitsgraden von $\nu = 5.33$, bei 75 iterativen Glättungen beträgt $\nu = 43.45$.

Analog wie in Abschnitt 6.1 wurden Schätzungen für die Erwartungswerte von $WH(\nu)$, $CVLL(\nu)$ sowie der global korrigierten Diskrepanz $dc_{2,\Delta,\Delta^2}(f, \hat{f}_\nu)$ dadurch ermittelt, daß für jeden von insgesamt 75 verschiedenen Werten für ν über 40 Realisationen gemittelt wurde. Die Ergebnisse sind in Abb. 6.4 (a) - (c) dargestellt. Alle drei Kurvenverläufe besitzen im Bereich $5.33 \leq \nu \leq 43.45$ ein globales Minimum, weitere lokale Minima liegen nicht vor. Der Kurvenverlauf für die Referenzdiskrepanz besitzt ein Minimum für $\nu = 27.96$, was 31 iterativen Glättungen entspricht (vgl. Abb. 6.4 (a)). Bei dem WH-Kriterium liegt das Minimum bei $\nu = 15.95$ (10 iterative Glättungen). Für das CVLL-Kriterium ergibt sich für $\nu = 30.96$ (37 iterative Glättungen) ein Minimum. Die Minima der Kurvenverläufe für die Referenzdiskrepanz und dem CVLL-Kriterium liegen sehr nahe zusammen. Es kann angenommen werden, daß zumindest im Mittel mit dem CVLL-Kriterium Spektralschätzungen erzielt werden, die nahezu optimal bezüglich der global korrigierten Diskrepanz $dc_{2,\Delta,\Delta^2}(f, \hat{f}_\nu)$ sind. Bei automatischen Spektralschätzungen unter Verwendung des WH-Kriteriums ist dagegen eine erhebliche Unterglättung der Spektralschätzungen zu erwarten. Es ist allerdings zu beachten, daß nur ein spezieller datengenerierender AR(4)-Prozeß betrachtet worden war. Maße zur Evaluierung der Stichprobenvariabilität waren ebenfalls nicht berücksichtigt worden. Eine genauere Analyse von Kreuzvalidierungskriterien im Frequenzbereich ist erstrebenswert, kann aber im Rahmen dieser Arbeit nicht geleistet werden.

Anwendungen und Erweiterungen der Kreuzvalidierungskriterien im Frequenzbereich sind z.B. automatische Prognoseverfahren im Frequenzbereich (vgl. HURVICH[1987]), sowie Verfahren für die nichtparametrische Regression bei autokorrelierten Störtermen (vgl. HURVICH/ZEGER[1990]).

Um die rechenintensive Kreuzvalidierung zu vermeiden, entwickelten HURVICH/BELTRÃO[1990] ein Kriterium, das in HURVICH[1992] ausführlicher dargestellt wird. Es hat sich aber gezeigt, daß ihre Weiterentwicklung des CVLL-Kriteriums vom Grundaufbau her sehr viel Ähnlichkeit mit Kriterien besitzt, die bereits von HANNAN/RISANEN[1988] als nichtparametrische ad hoc Modellselektionskriterien vorgeschlagen wurden. Eine kurze heuristische Herleitung einer kreuzvalidierungsfreien Variante des CVLL-Kriteriums ergibt sich wie folgt (HURVICH/BELTRÃO[1990], HURVICH[1992]): Approximiert man im Erwartungswert der Kreuzentropie,

$$(6.16) \quad ES(f, \hat{f}_\nu) = E \int_{-0.5}^{0.5} \left[\ln \hat{f}_\nu(\omega) + \frac{f(\omega)}{\hat{f}_\nu(\omega)} \right] d\omega$$

das Integral bezüglich ω durch eine Riemannsumme, so erhält man

$$(6.17) \quad ES(f, \hat{f}_\nu) \approx E \frac{1}{N} \sum_{k=1}^N \left[\ln \hat{f}_\nu(\omega_k) + \frac{f(\omega_k)}{\hat{f}_\nu(\omega_k)} \right],$$

bzw.

$$(6.18) \quad ES(f, \hat{f}_\nu) \approx E \frac{1}{N} \sum_{k=1}^N \ln \hat{f}_\nu(\omega_k) + E \frac{1}{N} \sum_{k=1}^N \frac{f(\omega_k)}{\hat{f}_\nu(\omega_k)}.$$

Da sich die Verteilung eines direkten oder indirekten Spektralschätzers \hat{f}_ν approximativ angeben läßt als (vgl. Abschnitt 4.2, (4.30) und Abschnitt 4.3, (4.47))

$$\hat{f}_\nu(\omega_k) \sim (1/\nu) f_\nu(\omega_k) \chi_\nu^2,$$

folgt mit $E(1/\chi_\nu^2) = 1/(\nu - 2)$

$$(6.19) \quad E \frac{1}{N} \sum_{k=1}^N \frac{f(\omega_k)}{\hat{f}_\nu(\omega_k)} \approx \frac{1}{N} \sum_{k=1}^N \frac{\nu}{\nu - 2} = \frac{\nu}{\nu - 2}.$$

Daraus ergibt sich, daß das Kriterium,

$$\text{CVLL2}(\nu) = \frac{1}{N} \sum_{k=1}^N \ln \hat{f}_\nu(\omega_k) + \frac{\nu}{\nu - 2},$$

ein approximativ unverzerrter Schätzer von $ES(f, \hat{f}_\nu)$ ist. Damit entspricht die Minimierung des CVLL2-Kriteriums approximativ einer Minimierung der Kreuzentropie $S(f, \hat{f}_\nu)$ sowie der Kullback-Leibler-Information $\text{KLIC}(f, \hat{f}_\nu)$.

Ist die Zahl der Frequenzpunkte, an denen die Spektralschätzung berechnet wird, größer als die Zahl der Fourierfrequenzen, so ist zu beachten, daß die äquivalente Zahl von Freiheitsgraden mit den in Abschnitt 4.2 angegebenen Korrekturen zu berechnen ist.

HURVICH/BELTRÃO[1990] verwenden eine andere Variante des CVLL2-Kriteriums, die aber bezüglich der Minimierungseigenschaften äquivalent zu der hier vorgestellten Version ist. Bei HURVICH[1992] wird diese minimierungsäquivalente Variante als AICC bezeichnet, da sie, ebenso wie das im Zeitbereich definierte AICC, einer Minimierung der Kullback-Leibler-Information entspricht. Um diesem Sprachgebrauch entgegenzukommen, aber gleichzeitig den Unterschied der Modellselektionskriterien hervorzuheben, wird das CVLL2-Kriterium im weiteren als FAICC bezeichnet werden, d.h. es soll gelten:

$$(6.20) \quad \text{FAICC}(\nu) = \text{CVLL2}(\nu) = \frac{1}{N} \sum_{k=1}^N \ln \hat{f}_\nu(\omega_k) + \frac{\nu}{\nu - 2}.$$

Das FAICC kann auf interessante Art und Weise in Analogie zu den im Zeitbereich definierten Modellselektionskriterien AIC, AICC, SIC, HQ und FPE interpretiert werden. Nach der Formel von Kolmogorov gilt für die Innovationsvarianz σ_ϵ^2 eines schwach stationären stochastischen Prozesses $\{X_t\}$, für den eine kausale MA(∞)-Darstellung existiert (vgl. Abschnitt 3.3, (3.32)),

$$(6.21) \quad \sigma_\epsilon^2 = \exp\left[\int_{-0.5}^{0.5} \ln f(\omega) d\omega\right],$$

bzw.

$$(6.22) \quad \ln \sigma_\epsilon^2 = \int_{-0.5}^{0.5} \ln f(\omega) d\omega.$$

Mit einer Riemannsummen-Approximation und Ersatz der theoretischen Größen durch geschätzte Größen ergibt sich

$$(6.23) \quad \ln \hat{\sigma}_\epsilon^2 = \frac{1}{N} \sum_{k=1}^N \ln \hat{f}_\nu(\omega_k).$$

Damit kann der erste Term des FAICC genauso wie bei Kriterien im Zeitbereich als Term interpretiert werden, der die Anpassungsgüte des nichtparametrischen Modells

\hat{f}_ν an die Daten mißt. Der zweite Term des FAICC, $\nu/(\nu - 2)$ kann nun ebenfalls in Analogie zu den Kriterien im Zeitbereich als Straffunktion für zusätzliche „Parameter“ interpretiert werden, wobei allerdings folgende Unterschiede zu beachten sind: Eine sehr glatt verlaufende Spektralschätzung ist als Modell mit wenigen Parametern zu interpretieren, eine raue Spektralschätzung mit zahlreichen Peaks ist als Modell mit einer großen Zahl von Parametern zu interpretieren. Damit entspricht eine glatte Spektralschätzung mit einer großen Zahl äquivalenter Freiheitsgrade bzw. einer großen äquivalenten Bandbreite (EBW) einem parametersparsamen Modell. Umgekehrt entspricht eine raue Spektralschätzung mit einer kleinen Zahl von äquivalenten Freiheitsgraden bzw. einer kleinen äquivalenten Bandbreite einem Modell mit großzügiger Parametrisierung. Soll die Abhängigkeit des nichtparametrischen Modellselektionskriteriums $\text{FAICC}(\nu)$ von der äquivalenten Zahl der Freiheitsgrade ν ganz in Analogie zur Modellordnung p für autoregressive Spektralschätzer formuliert werden, ist als Argument eigentlich der Kehrwert von ν vorzuziehen. Kleine Werte von $1/\nu$ entsprechen einem parametersparsamen Modell, große Werte von $1/\nu$ entsprechen einem Modell mit vielen Parametern.

Mit der Interpretation von $1/\nu$ als nichtparametrischem Modellselektionsparameter liegt es nahe, in Analogie zu den entsprechenden Kriterien im Zeitbereich, die folgenden nichtparametrischen Modellselektionskriterien zu formulieren:

$$(6.24) \quad \text{FAIC}(\nu) = \frac{1}{N} \sum_{k=1}^N \ln \hat{f}_\nu(\omega_k) + \frac{2}{\nu}$$

$$(6.25) \quad \text{FSIC}(\nu) = \frac{1}{N} \sum_{k=1}^N \ln \hat{f}_\nu(\omega_k) + \frac{\ln N}{\nu}$$

$$(6.26) \quad \text{FHQ}(\nu) = \frac{1}{N} \sum_{k=1}^N \ln \hat{f}_\nu(\omega_k) + \frac{2c \ln \ln N}{\nu} \quad \text{mit } c \geq 1.$$

Nichtparametrische Modellselektionskriterien, die unmittelbar auf einer geeigneten Straffunktion in Abhängigkeit von einem Modellselektionsparameter basieren, sollen im weiteren als strafende nichtparametrische Modellselektionskriterien bezeichnet werden. Wenn der Zusammenhang klar ist, wird auch einfach von strafenden Kriterien gesprochen.

	EDF, ν	Anzahl der Iterationen
$dc_{2,\Delta,\Delta^2}(f, \hat{f}_\nu)$	27.96	31
FAIC(ν)	15.95	10
FAICC(ν)	18.84	14
FSIC(ν)	28.85	33
FHQ(ν)	22.49	20

Tabelle 6.1: Vergleich der globalen Minima

FAIC und FSIC wurden bereits von HANNAN/RISSANEN[1988] als ad hoc Kriterien vorgeschlagen. Für die Anwendungen des FHQ-Kriterium in dieser Arbeit wird, wie bei dem analogen Kriterium im Zeitbereich, $c = 1$ gesetzt. Die Konstruktion zahlreicher weiterer direkter Kriterien ist möglich. Es liegen allerdings keine theoretischen Aussagen darüber vor, welche Optimalitätseigenschaften die Kriterien FAIC, FSIC und FHQ besitzen. Im Rahmen dieser Arbeit erfolgt im wesentlichen eine experimentelle Evaluierung mit Hilfe von Simulationen.

Um zumindest einen groben Vergleich mit den Kreuzvalidierungskriterien zu ermöglichen, wurden mit dem gleichen AR(4)-Prozeß, der für die Erstellung der Abbildungen 6.4 zugrundegelegt wurde, weitere 40 Realisationen erzeugt und mit Hilfe der iterativen Hamming-Filterung für 75 Werte von ν direkte Spektralschätzungen berechnet. Analog wie für die Kreuzvalidierungskriterien wurden Schätzungen für die Erwartungswerte von $dc_{2,\Delta,\Delta^2}(f, \hat{f}_\nu)$, FAIC(ν), FAICC(ν), FSIC(ν), FHQ(ν) ermittelt. Die Kurvenverläufe sind in Abbildung 6.5 (a) bis (d) wiedergegeben, in Tabelle 6.1 ist die genaue Lage der globalen Minima zusammengefaßt. Der Kurvenverlauf für die Referenzdiskrepanz $dc_{2,\Delta,\Delta^2}(f, \hat{f}_\nu)$ unterschied sich so geringfügig von der bereits in Abb. 6.4 (a) dargestellten Funktion, daß auf eine Darstellung verzichtet wurde, insbesondere lag das Minimum wiederum bei $\nu = 27.96$ bzw. 31 iterativen Glättungen.

Die beste Übereinstimmung mit der Referenzdiskrepanz erzielt FSIC, das mit 33 iterativen Glättungen im Vergleich zur Referenzdiskrepanz mit 31 iterativen Glättungen, geringfügig überglättet. Mit den anderen Kriterien ergeben sich rauhere Spek-

tralschätzungen, bei Anwendung des FAIC sind die rauhesten Spektralschätzungen zu erwarten. Vergleicht man den Kurvenverlauf des CVLL-Kriteriums aus Abb. 6.4 (c) mit den Kurvenverläufen der Abbildungen 6.5 (a) bis (d) so fällt auf, daß entgegen den Erwartungen nicht der Kurvenverlauf des FAICC die beste Übereinstimmung zeigt, sondern der Kurvenverlauf des FSIC. Dies ist ein klarer Hinweis darauf, zwischen dem Kreuzvalidierungskriterium CVLL und dem strafendem Kriterium FAICC streng zu unterscheiden, obwohl beide in engem Zusammenhang stehen.

Als Vorteil der strafenden Kriterien im Vergleich zu den "Leave-One-Out"-Kreuzvalidierungskriterien ergibt sich zunächst, daß die strafenden Kriterien lediglich einen geringen Bruchteil der Rechenzeit benötigen, der für die Kreuzvalidierungskriterien erforderlich ist. Für das oben angegebene Rechenbeispiel mit $N = 256$, $n_\omega = 127$ und $N_\nu = 100$ sind für eine Realisation statt $N_\nu \times n_\omega = 12700$ Spektren lediglich $N_\nu = 100$ Spektren zu berechnen. Da die strafenden Kriterien kein "Leave-One-Out"-Periodogramm erfordern und die äquivalente Zahl von Freiheitsgraden auch für indirekte Spektralschätzer definiert ist (vgl. z.B. Tabelle 4.2), können diese Kriterien auch unmittelbar mit den indirekten Spektralschätzern bei sehr geringem Rechenaufwand eingesetzt werden. Darüberhinaus kann für spezielle Anwendungen der trade-off zwischen Anpassungsgüte und Straffunktion beliebig durch die Wahl verschiedener Straffunktionen vorgegeben werden.

Der Nachteil, daß für die ad hoc Kriterien keine theoretischen Ergebnisse bezüglich ihrer Optimalitätseigenschaften vorliegen, ist von untergeordneter Bedeutung. Wie sich bei der Diskussion der parametrischen Modellselektionskriterien im Zeitbereich gezeigt hatte, war es auch mit theoretischen Ergebnissen bezüglich der Optimalitätseigenschaften nicht möglich, einem bestimmten Kriterium allgemeinen Vorrang zu geben.

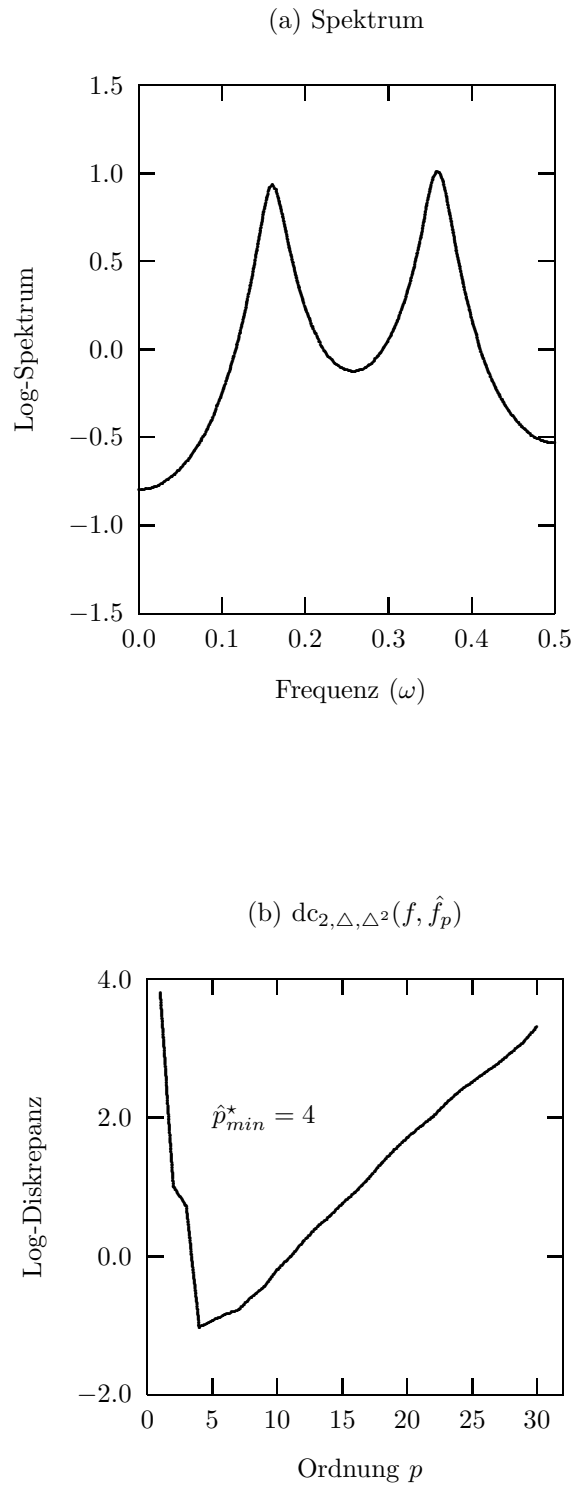


Abb. 6.1: (a) Spektrum des AR(4)-Prozesses (6.8), (b) Schätzung für $E[dc_{2,\Delta,\Delta^2}(f, \hat{f}_p)]$, gemittelt über 100 Realisationen des AR(4)-Prozesses (6.8) mit $N = 32$.

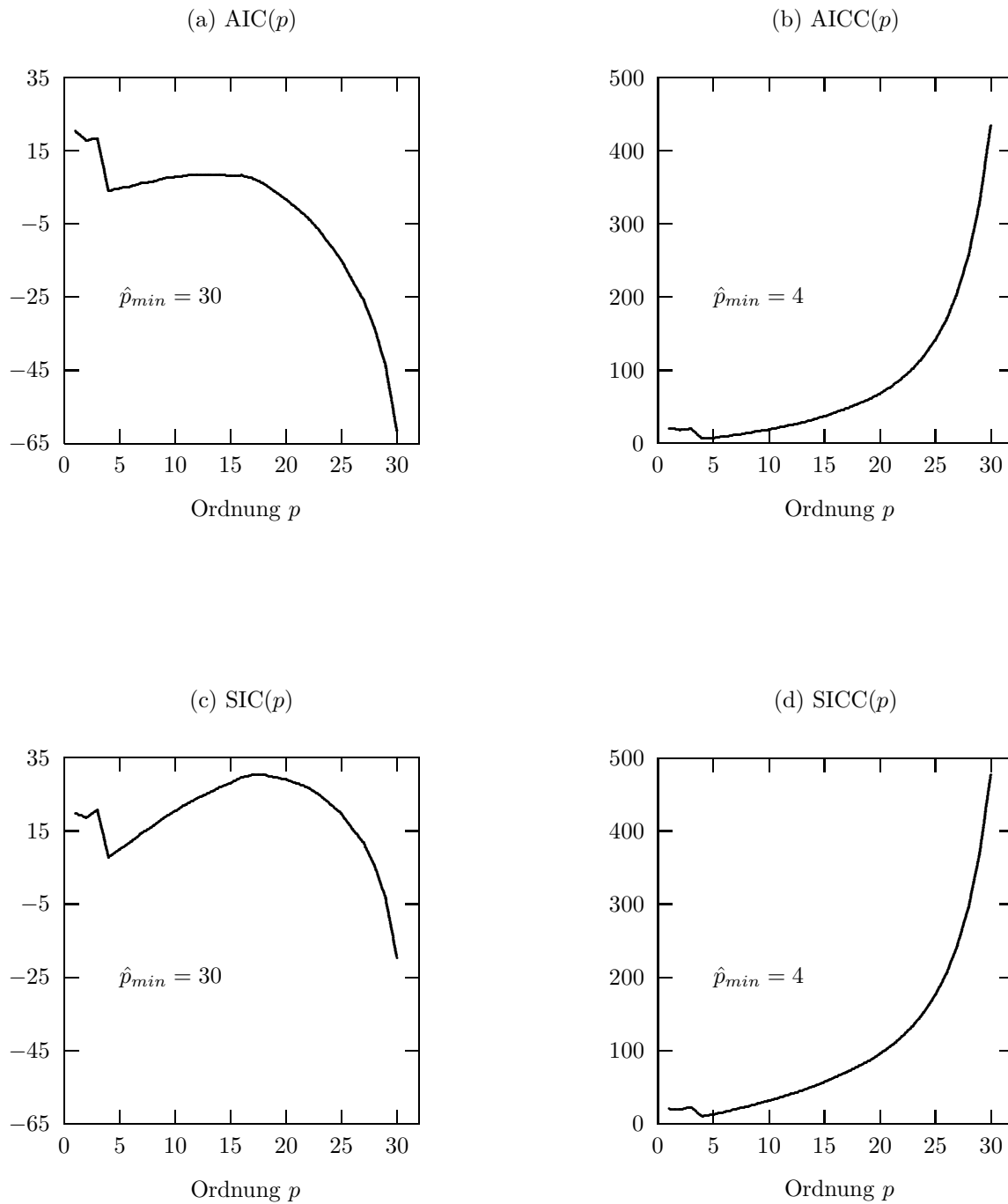


Abb. 6.2: Vergleich von unkorrigierten und korrigierten Modellselektionskriterien anhand von 100 Realisationen des AR(4)-Prozesses (6.8) mit $N = 32$, $N_E = 128$. Minimum der Referenzdiskrepanz: $\hat{p}_{min}^* = 4$.

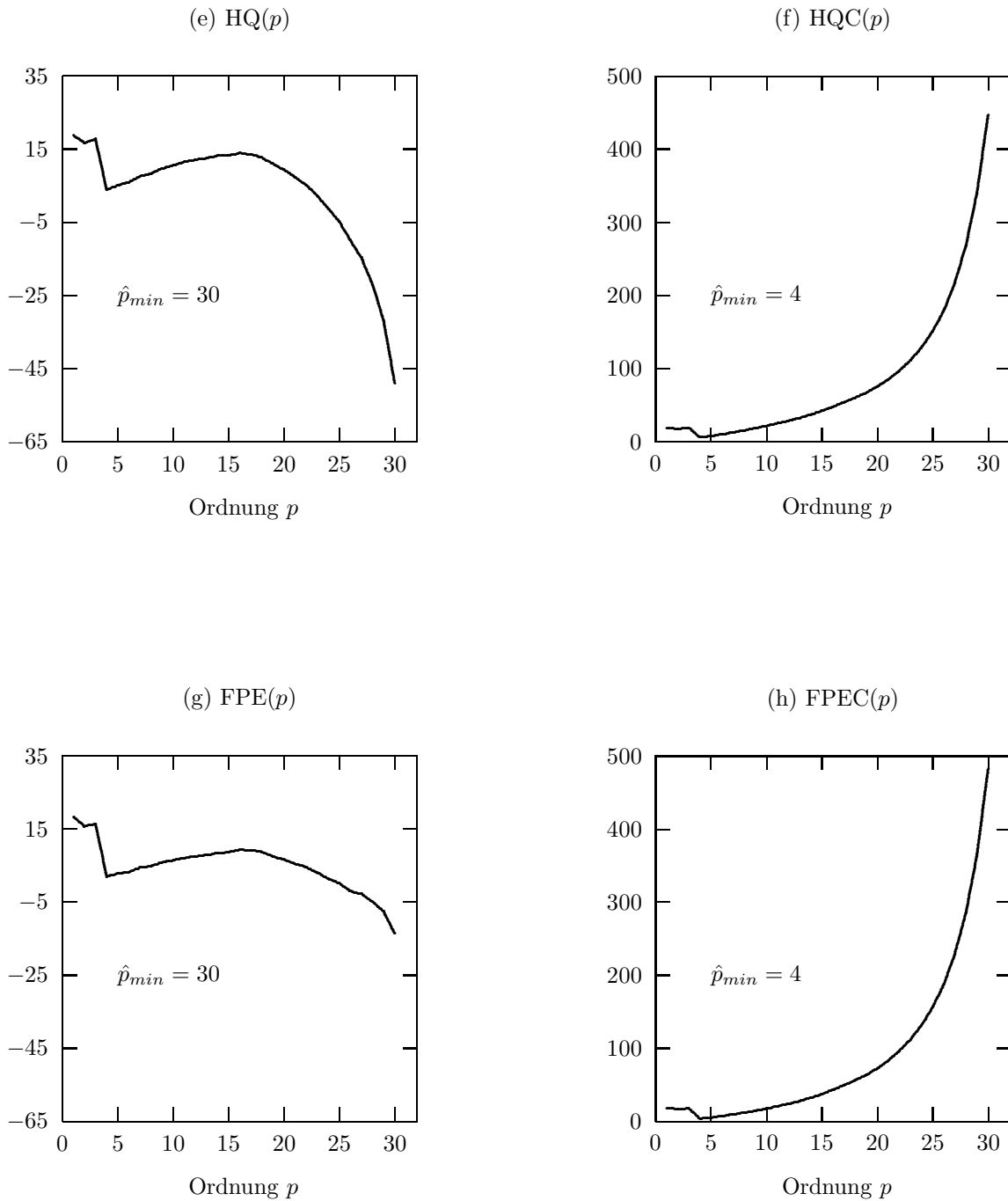


Abb. 6.2: (Fortsetzung) Vergleich von unkorrigierten und korrigierten Modellselektionskriterien anhand von 100 Realisationen des AR(4)-Prozesses (6.8) mit $N = 32$, $N_E = 128$. Minimum der Referenzdiskrepanz: $\hat{p}_{min} = 4$.

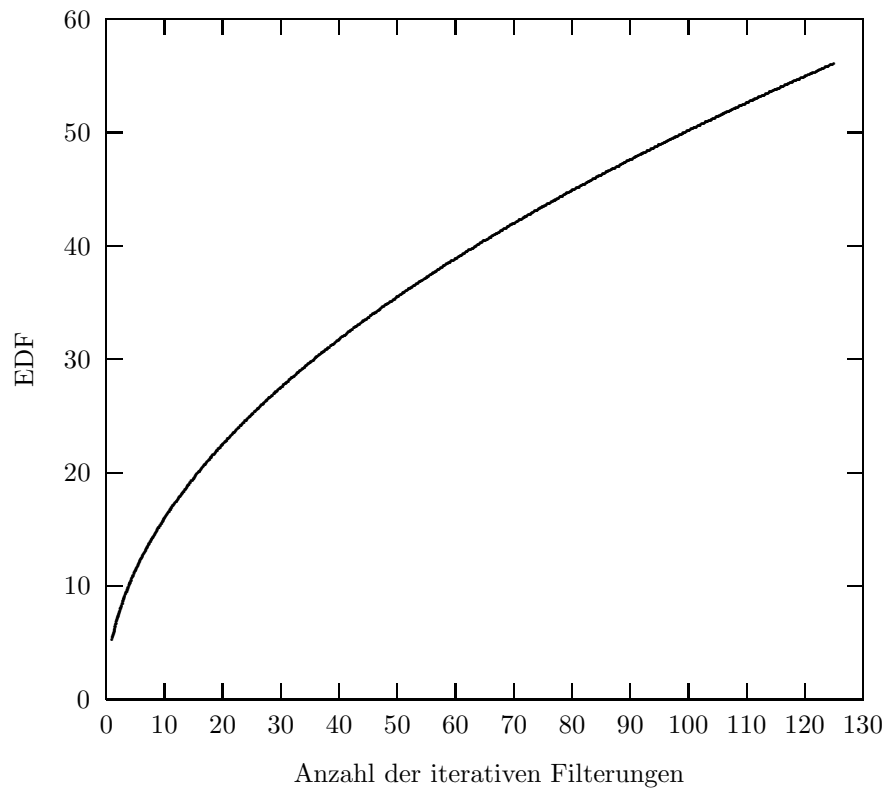


Abb. 6.3: Zahl der äquivalenten Freiheitsgrade (EDF) bei iterativer Anwendung des Hamming-Filters $\{0.25, 0.5, 0.25\}$

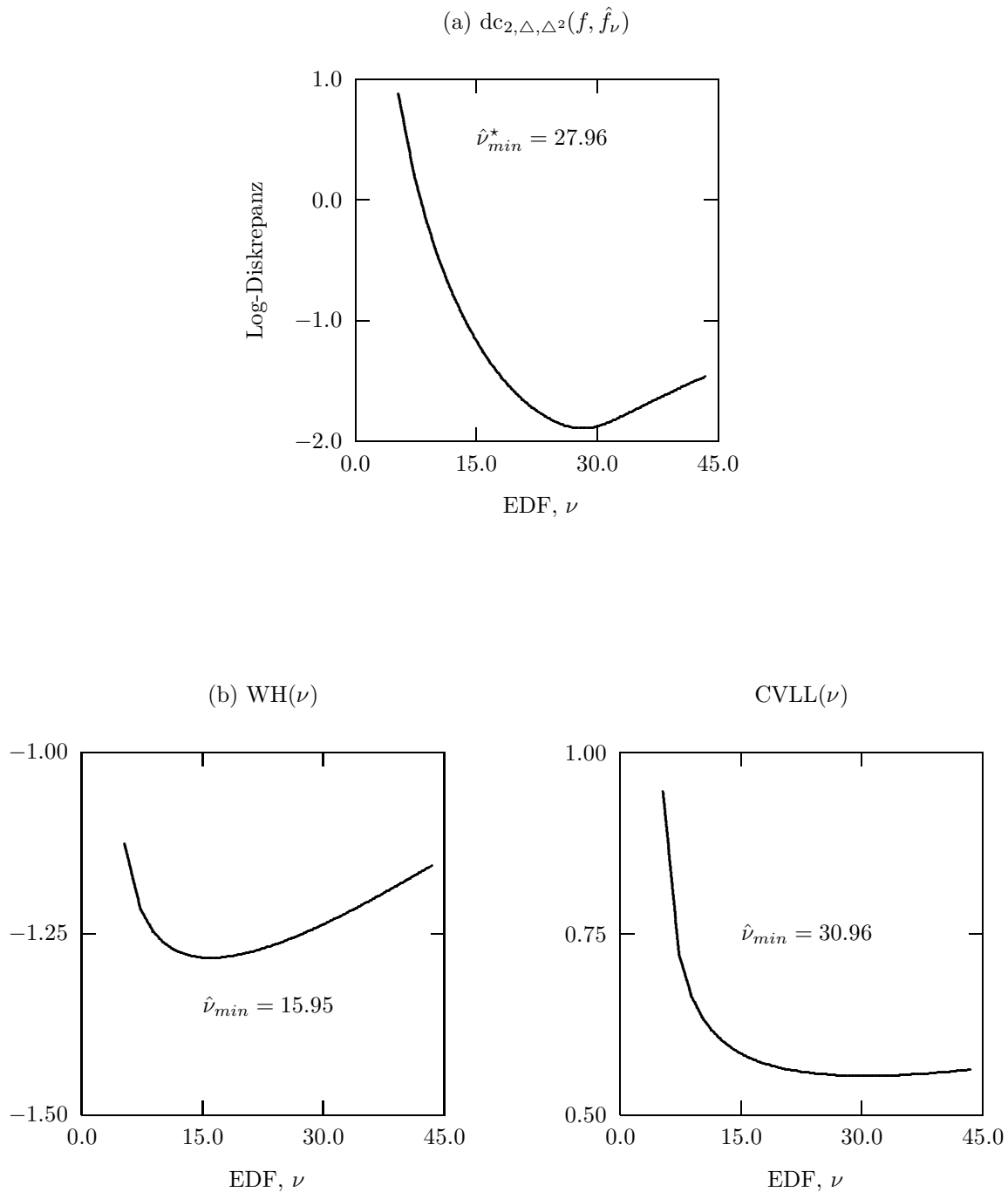


Abb. 6.4: Schätzungen für $E[dc_{2,\Delta,\Delta^2}(f, \hat{f}_\nu)]$, $E[WH(\nu)]$, $E[CVLL(\nu)]$, gemittelt über 40 Realisationen des AR(4)-Prozesses (6.8) mit $N = 256$, $N_E = 128$. Direkte Spektralschätzungen mit iterativer Hamming-Filterung.

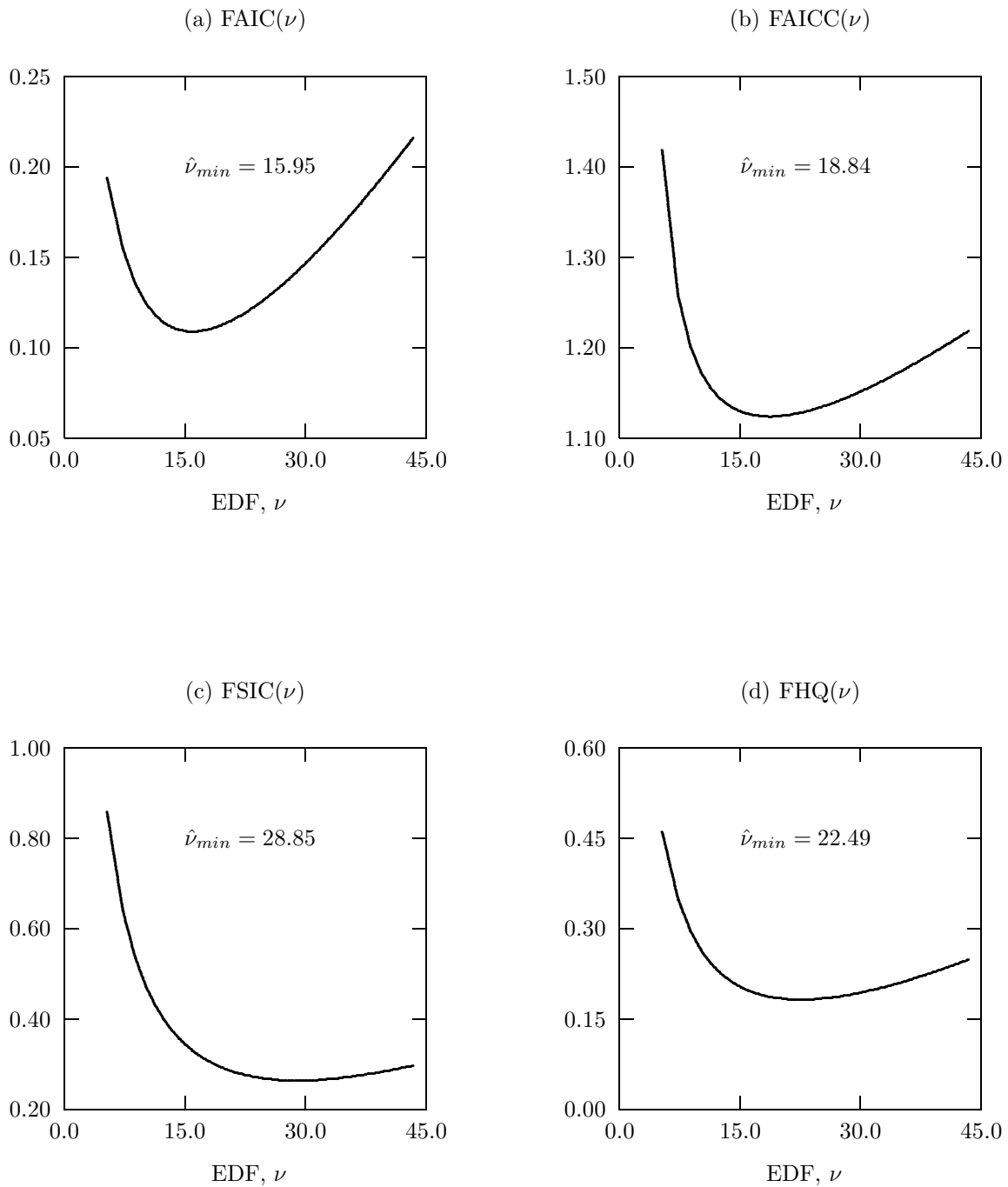


Abb. 6.5: Schätzungen für $E[\text{FAIC}(\nu)]$, $E[\text{FAICC}(\nu)]$, $E[\text{FSIC}(\nu)]$, $E[\text{FHQ}(\nu)]$ gemittelt über 40 Realisationen des AR(4)-Prozesses (6.8) mit $N = 256$, $N_E = 128$, $N_\omega = 129$. Direkte Spektralschätzungen mit iterativer Hamming-Filterung. Minimum der Referenzdiskrepanz: $\hat{\nu}_{min}^* = 27.96$.

6.3 Simulation

Anhand von drei autoregressiven Prozessen und einem MA-Prozeß erfolgt in diesem Abschnitt ein Vergleich von nichtparametrischen und parametrischen Modellselektionskriterien für autoregressive, direkte und indirekte Spektralschätzungen. Die charakteristischen Polynome der drei AR-Prozesse und des MA-Prozesses, für die geeignete Kurzbezeichnungen vergeben werden, lauten:

(Four-Peak)

$$\begin{aligned}
 (6.27) \quad \phi(z) &= (1 - 0.95e^{i2\pi(0.25/3)}z^{-1})(1 - 0.95e^{-i2\pi(0.25/3)}z^{-1}) \\
 &\quad (1 - 0.95e^{i2\pi(0.50/3)}z^{-1})(1 - 0.95e^{-i2\pi(0.50/3)}z^{-1}) \\
 &\quad (1 - 0.95e^{i2\pi(1.00/3)}z^{-1})(1 - 0.95e^{-i2\pi(1.00/3)}z^{-1}) \\
 &\quad (1 - 0.95e^{i2\pi(1.25/3)}z^{-1})(1 - 0.95e^{-i2\pi(1.25/3)}z^{-1}) \\
 &= 1 + 0.81451z^{-4} + 0.66342z^{-8},
 \end{aligned}$$

(Smooth)

$$\begin{aligned}
 (6.28) \quad \phi(z) &= (1 - 0.2e^{i2\pi 0.1}z^{-1})(1 - 0.2e^{-i2\pi 0.1}z^{-1}) \\
 &= 1 - 0.32361z^{-1} + 0.04z^{-2},
 \end{aligned}$$

(One-Peak)

$$\begin{aligned}
 (6.29) \quad \phi(z) &= (1 - 0.2e^{i2\pi 0.063}z^{-1})(1 - 0.2e^{-i2\pi 0.063}z^{-1}) \\
 &\quad (1 - 0.4e^{i2\pi 0.125}z^{-1})(1 - 0.4e^{-i2\pi 0.125}z^{-1}) \\
 &\quad (1 - 0.6e^{i2\pi 0.188}z^{-1})(1 - 0.6e^{-i2\pi 0.188}z^{-1}) \\
 &\quad (1 - 0.8e^{i2\pi 0.250}z^{-1})(1 - 0.8e^{-i2\pi 0.250}z^{-1}) \\
 &= 1.00000 - 1.39446z^{-1} + 1.83853z^{-2} - 1.49874z^{-3} + 0.95826z^{-4} \\
 &\quad - 0.42039z^{-5} + 0.12467z^{-6} - 0.02072z^{-7} + 0.00147z^{-8}
 \end{aligned}$$

(MA-Smooth)

$$(6.30) \quad \theta(z) = 1 + 0.95z^{-1}.$$

Die Spektren der vier Prozesse sowie die Lage der Wurzeln der charakteristischen Polynome gegenüber dem Einheitskreis sind in den Abbildungen 6.6 (a) und (b), 6.12 (a) und (b), 6.18 (a) und (b) sowie 6.24 (a) und (b) dargestellt. Das Spektrum des Four-Peak-Prozesses (6.27) enthält vier ausgeprägte Peaks, die symmetrisch um die Frequenz $\omega = 0.25$ verteilt sind. Wie Gleichung (6.27) bzw. Abb. 6.6 (a) zu entnehmen ist, wurde auf einen deutlichen Abstand der vier Peakfrequenzen geachtet, so daß auch direkte und indirekte Spektralschätzer sinnvolle Ergebnisse erzielen können. Der Smooth-Prozeß (6.28) ist ein AR(2)-Prozeß, dessen Spektrum monoton von einem Maximum bei $\omega = 0$ bis zu einem Minimum bei $\omega = 0.5$ abfällt. Der One-Peak-Prozeß (6.29) ist wiederum ein AR(8)-Prozeß, wobei aber drei von vier Wurzelpaaren so gewählt wurden, daß sie weit im inneren des Einheitskreises liegen. Entsprechend besitzt das Spektrum des One-Peak-Prozesses nur einen einzigen breiten Peak, die Koeffizienten des charakteristischen Polynoms fallen ab Lag drei in etwa exponentiell ab, die Koeffizienten zum Lag sieben und zum Lag acht sind bereits sehr nahe bei Null. Sowohl bei dem Smooth-Prozeß (6.28) als auch bei dem One-Peak-Prozeß (6.29) ist zu erwarten, daß die parametrischen Modellselektionskriterien Ordnungen \hat{p} ermitteln, die unterhalb der vorgegebenen „wahren“ Modellordnung p liegen. Das Spektrum des MA-Smooth-Prozesses (6.30) fällt ebenso wie das Spektrum des (AR-)Smooth-Prozesses monoton von einem Maximum bei $\omega = 0$ bis zu einem Minimum bei $\omega = 0.5$, jedoch ist das Steigungs- und Krümmungsverhalten der beiden Spektren sehr unterschiedlich.

Zur Evaluierung der Modellselektionskriterien wurde wie folgt vorgegangen: Zunächst wurden die Verfahren Forward-Linear-Prediction (FLP), iteratives Hamming und die Parzen-Schätzung als autoregressives, direktes bzw. indirektes Schätzverfahren ausgewählt. Das iterative Hamming-Verfahren soll im weiteren auch abgekürzt einfach als Hamming-Verfahren bezeichnet werden. Für die FLP-Methode wurden die Modellselektionskriterien AIC, AICC, SIC sowie das HQ-Kriterium herangezogen, wobei einheitlich für alle Prozesse als maximale Modellordnung $p_{max} = 20$ festgelegt wurde. Für den direkten und indirekten Schätzer wurden als nichtparametrische Modellselektionskriterien das FAICC sowie die ad hoc Kriterien FAIC, FSIC und FHQ gewählt. Die Bereiche für die äquivalente Zahl von Freiheitsgraden wurde in Abhängigkeit von den datenerzeugenden Prozessen unterschiedlich festgelegt. Die Grenzen lagen bei dem Four-Peak- und One-Peak-Prozeß ca. im Intervall ($5.0 < \nu < 45.0$), bei dem Smooth-Prozeß ca. im Intervall ($5.0 < \nu < 100.0$) und bei dem MA-Smooth-Prozeß ca. im Intervall ($5.0 < \nu < 60.0$). Das Hamming-Verfahren erwies sich insbesondere bei dem

Smooth-Prozeß als sehr rechenintensiv, da für eine äquivalente Zahl von Freiheitsgraden mit $\nu = 100$ ca. 400 iterative Glättungen erforderlich waren. Aus Zeitgründen wurde allerdings darauf verzichtet, mit weiteren direkten Schätzern zu experimentieren. Von jedem Prozeß wurden für den Stichprobenumfang $N = 128$ jeweils 100 unabhängige Realisationen erzeugt, wobei jeweils $N_E = 128$ Einschwingbeobachtungen berücksichtigt wurden.

Zur Beurteilung der automatischen Schätzverfahren wurden die folgenden Kriterien herangezogen:

Zum einen erfolgte für jeden datenerzeugenden Prozeß und jedes Schätzverfahren ein graphischer Vergleich der Schätzungen für $E[\text{AIC}(p)]$, $E[\text{AICC}(p)]$, $E[\text{SIC}(p)]$, $E[\text{HQ}(p)]$ bzw. $E[\text{FAIC}(\nu)]$, $E[\text{FAICC}(\nu)]$, $E[\text{FSIC}(\nu)]$, $E[\text{FHQ}(\nu)]$ mit den Schätzungen für die Erwartungswerte der Referenzdiskrepanzen, $E[\text{dc}_{2,\Delta,\Delta^2}(f, \hat{f}_p)]$ (FLP-Schätzung) bzw. $E[\text{dc}_{2,\Delta,\Delta^2}(f, \hat{f}_\nu)]$ (Hamming- und Parzen-Schätzung). Dieser Vergleich ermöglicht Aussagen darüber, ob ein Bias zwischen der Lage der globalen Minima der Kurvenverläufe für die Modellselektionskriterien und der Lage der globalen Minima der Kurvenverläufe für die Referenzdiskrepanzen besteht. Anhand eines Bias kann auch eine Tendenz zur Über- oder Unterglättung eines Modellselektionskriteriums festgestellt werden. Die Ergebnisse für die drei AR-Prozesse und den MA(1)-Prozeß sind in den folgenden Abbildungen dargestellt:

Four-Peak-Prozeß:	Abb. 6.7 (a) - (c), Abb. 6.8 (a) - (d), Abb. 6.9 (a) - (d), Abb. 6.10 (a) - (d),
Smooth-Prozeß:	Abb. 6.13 (a) - (c), Abb. 6.14 (a) - (d), Abb. 6.15 (a) - (d), Abb. 6.16 (a) - (d),
One-Peak-Prozeß:	Abb. 6.19 (a) - (c), Abb. 6.20 (a) - (d), Abb. 6.21 (a) - (d), Abb. 6.22 (a) - (d),
MA-Smooth-Prozeß:	Abb. 6.25 (a) - (c), Abb. 6.26 (a) - (d), Abb. 6.27 (a) - (d), Abb. 6.28 (a) - (d).

Um die graphische Interpretation zu erleichtern, sind die numerischen Werte der globalen Minima der gemittelten Kurvenverläufe in den Graphiken explizit mit angegeben worden.

Zum anderen erfolgte für jedes der Schätzverfahren sowie für die parametrischen bzw. nichtparametrischen Modellselektionskriterien eine weitere Evaluierung mit Hilfe der

global korrigierten Diskrepanz $dc_{2,\Delta,\Delta^2}(f, \hat{f})$. Hierzu wurde für jede Realisation, jedes Schätzverfahren und jedes Modellselektionskriterium die jeweils optimale Schätzung $\hat{f}_{\hat{p}_{min}}$ bzw. $\hat{f}_{\hat{v}_{min}}$ ermittelt und die global korrigierten Diskrepanzen $dc_{2,\Delta,\Delta^2}(f, \hat{f}_{\hat{p}_{min}})$ bzw. $dc_{2,\Delta,\Delta^2}(f, \hat{f}_{\hat{v}_{min}})$ berechnet. Die Ergebnisse sind mit Hilfe von Boxplots, bei denen jeweils das Minimum, 1. Quartil, Median, 3. Quartil und das Maximum graphisch wiedergegeben werden, in den Abbildungen 6.11 (a) - (c) (Four-Peak-Prozeß), 6.17 (a) - (c) (Smooth-Prozeß), 6.23 (a) - (c) (One-Peak-Prozeß) sowie 6.29 (a) - (c) (MA-Smooth-Prozeß) dargestellt. Damit ist eine unmittelbare Gegenüberstellung von parametrischen und nichtparametrischen Schätzverfahren möglich. Die eng gestrichelte Hilfslinie in diesen Graphiken halbiert jeweils die Ordinatenachse und soll den visuellen Vergleich erleichtern.

Um einen Gesamtüberblick der Simulationsergebnisse geben zu können, sind die Medianwerte aus diesen Boxplots in Tabelle 6.2 zusammengefaßt. Da die global korrigierten Diskrepanzen $dc_{2,\Delta,\Delta^2}(f, \hat{f})$ innerhalb eines relativ großen Wertebereichs lagen, war es für eine einheitliche Darstellung erforderlich, jeweils die logarithmierten Diskrepanzen wiederzugeben.

Betrachtet man zunächst die Ergebnisse für die Kurvenverläufe der geschätzten Erwartungswerte der verschiedenen parametrischen und nichtparametrischen Modellselektionskriterien so ergeben sich u.a. die folgenden Aussagen: Für die FLP-Schätzung bei dem Smooth- und dem One-Peak-Prozeß weisen die Referenzdiskrepanzen (vgl. Abb. 6.13 (a), 6.19 (a)) globale Minima bei $\hat{p}^* = 1$ bzw. $\hat{p}^* = 5$ auf. Damit ergeben sich für diese beiden Prozesse bei einem Stichprobenumfang von $N = 128$ jeweils optimale Modellordnungen, die unter den vorgegebenen „wahren“ Modellordnungen von $p = 2$ bzw. $p = 8$ liegen. Die globalen Minima der Kurvenverläufe der Schätzungen für $E[AIC(p)]$, $E[AICC(p)]$, $E[SIC(p)]$ und $E[HQ(p)]$ stimmen mit den für die jeweiligen Referenzdiskrepanzen ermittelten globalen Minima überein (vgl. Abb. 6.20 (a) - (d) und Abb. 6.14 (a) - (d)). In Übereinstimmung mit den Referenzdiskrepanzen wird also die „wahre“ Modellordnung von allen parametrischen Modellselektionskriterien für den Smooth- und den One-Peak-Prozeß unterschätzt. Dies belegt die in Abschnitt 6.1 vorgetragene Ansicht, daß bei endlichen Stichproben die „wahre“ Modellordnung eines AR-Prozesses mit endlicher Ordnung p von untergeordneter Bedeutung ist. Für den Four-Peak-Prozeß stimmt die von den Referenzdiskrepanzen ausgewiesene optimale Modellordnung $\hat{p}_{min}^* = 8$ mit der „wahren“ Modellordnung des datenerzeugenden Prozesses überein (vgl. Abb. 6.7 (a)), die geschätzten Erwartungswerte der Modell-

		(F)AIC	(F)AICC	(F)SIC	F(HQ)
Four-Peak	FLP	-1.57	-1.67	-1.74	-1.70
	Hamming	-1.27	-1.25	-0.95	-1.26
	Parzen	-1.32	-1.29	-0.99	-1.28
Smooth	FLP	-1.90	-2.04	-2.35	-2.25
	Hamming	-1.80	-1.86	-2.18	-2.07
	Parzen	-1.86	-1.96	-2.19	-2.11
One-Peak	FLP	-1.45	-1.52	-1.60	-1.63
	Hamming	-0.81	-1.02	-1.31	-1.12
	Parzen	-0.90	-1.06	-1.30	-1.15
MA-Smooth	FLP	-0.61	-0.73	-0.90	-0.84
	Hamming	-1.08	-1.21	-1.22	-1.30
	Parzen	-1.14	-1.23	-1.28	-1.36

Tabelle 6.2: Median von $\log dc_{2,\Delta,\Delta^2}(f, \hat{f}_{\hat{p}_{min}})$ bzw. $\log dc_{2,\Delta,\Delta^2}(f, \hat{f}_{\hat{v}_{min}})$, 100 Realisationen mit $N = 128$, $N_E = 128$, $N_\omega = 65$.

selektionskriterien $AIC(p)$, $AICC(p)$, $SIC(p)$, $HQ(p)$ ermitteln ebenfalls eine optimale Modellordnung von $\hat{p}_{min} = 8$. Bei dem MA-Smooth-Prozeß ergeben sich allerdings bei allen Modellselektionskriterien Abweichungen von der mit Hilfe der Referenzdiskrepanz ermittelten optimalen Modellordnung von $\hat{p}_{min}^* = 7$. Die geringste Abweichung ergibt sich für das HQ-Kriterium, das globale Minimum des Kurvenverlaufes der Schätzung für $E[HQ(p)]$ liegt bei $\hat{p}_{min} = 6$ (vgl. 6.25 (a) - (d)).

Damit sind die parametrischen Modellselektionskriterien, was die Übereinstimmung der geschätzten globalen Minima mit den geschätzten globalen Minima der Referenzdiskrepanzen betrifft, für die drei datenerzeugenden AR-Prozesse gleichermaßen optimal. Handelt es sich bei dem datenerzeugenden Prozeß dagegen nicht um einen AR-Prozeß, wie im Falle des MA-Smooth-Prozesses, dann können sich erhebliche Abweichungen ergeben.

Bei den nichtparametrischen Schätzverfahren läßt sich zunächst feststellen, daß sich das Hamming- und das Parzen-Schätzverfahren bezüglich der für die Referenzdiskrepanzen ermittelten Kurvenverläufe bei allen vier Prozessen lediglich geringfügig unterscheiden (vgl. Abb. 6.7 (b) und (c), Abb. 6.13 (b) und (c), Abb. 6.19 (b) und (c), Abb. 6.25 (b) und (c)).

Für die nichtparametrischen Modellselektionskriterien bzw. Schätzverfahren ergibt ein Vergleich der gemittelten Kurvenverläufe deutliche Unterschiede: Bei dem Smooth-Prozeß und dem One-Peak-Prozeß liegen für alle Kriterien die ausgewiesenen globalen Minima unter der anhand der jeweiligen Referenzdiskrepanzen ermittelten optimalen Zahl von äquivalenten Freiheitsgraden (vgl. Abb. 6.15 (a) - (d), Abb. 6.16 (a) - (d), Abb. 6.21 (a) - (d), Abb. 6.22 (a) - (d)). Die besten Ergebnisse erzielt bei diesen beiden Prozessen das ad hoc Kriterium FSIC. Diese Aussage kehrt sich für den Four-Peak-Prozeß um: Das FSIC liefert in diesem Fall die schlechtesten Ergebnisse, wobei die Lage des optimalen globalen Minimums überschätzt und eine zu große Zahl äquivalenter Freiheitsgrade ausgewiesen wird. Die besten Ergebnisse bezüglich der Übereinstimmung der Lage der globalen Minima werden bei dem Four-Peak-Prozeß mit dem FAIC (Parzen-Schätzung) bzw. FHQ (Hamming-Schätzung) erzielt (vgl. Abb. 6.9 (a) - (d), Abb. 6.10 (a) - (d)). Bei dem MA-Smooth-Prozeß ergeben sich sowohl Über- als auch Unterschätzungen der Lage des optimalen Minimums (vgl. Abb. 6.27 (a) - (d), Abb. 6.28 (a) - (d)), das beste Ergebnis erzielt das HQ-Kriterium in Kombination mit der Parzen-Schätzung, wobei die Lage des Minimums der (gemittelten) Referenzdiskrepanz exakt ermittelt wird.

Der Vergleich der Kurvenverläufe zur Beurteilung der Modellselektionskriterien kann damit wie folgt zusammengefaßt werden: Sowohl bei den parametrischen Modellselektionskriterien als auch bei den nichtparametrischen Modellselektionskriterien ergibt sich eine Abhängigkeit des optimalen Modellselektionskriteriums von dem datenerzeugenden Prozeß. Beschränkt man die Betrachtung auf die drei datenerzeugenden AR-Prozesse, dann lassen sich keine Unterschiede zwischen den parametrischen Modellselektionskriterien feststellen. Bei allen vier betrachteten datenerzeugenden Prozessen gab es mindestens ein nichtparametrisches ad hoc Kriterium, das bessere Ergebnisse erzielt, als das theoretisch abgeleitete Kriterium FAICC.

Die Beurteilung der Modellselektionskriterien mit Hilfe der global korrigierten Diskrepanzen $dc_{2,\Delta,\Delta^2}(f, \hat{f}_{\hat{p}_{min}})$ bzw. $dc_{2,\Delta,\Delta^2}(f, \hat{f}_{\hat{v}_{min}})$ ergibt u.a. die folgenden Aussagen:

Der Tabelle 6.2 kann zunächst entnommen werden, daß bezüglich der Medianwerte der

global korrigierten Diskrepanzen das SIC für alle vier datenerzeugenden Prozesse die besten Ergebnisse unter den parametrischen Modellselektionskriterien erzielt. Berücksichtigt man zusätzlich noch die Boxplots 6.11 (a), 6.17 (a) sowie 6.23 (a) und vergleicht die Stichprobenvariabilität der parametrischen Modellselektionskriterien anhand der Spannweite bzw. dem Interquartilabstand, so ergibt sich insgesamt eine deutliche Überlegenheit des SIC, insbesondere im Vergleich zum AIC und AICC. Allerdings besitzen bei dem MA-Smooth-Prozeß die nichtparametrischen Schätzverfahren bei allen verwendeten Modellselektionskriterien geringere Medianwerte der global korrigierten Diskrepanz.

Bei den nichtparametrischen Modellselektionskriterien läßt sich dagegen kein Kriterium finden, das für alle vier datenerzeugenden Prozesse gleichermaßen bessere Ergebnisse erzielt (vgl. Tabelle 6.2 sowie Abb. 6.11 (b) und (c), 6.17 (b) und (c), 6.23 (b) und (c), 6.29 (b) und (c)). Bei dem Four-Peak-Prozeß erzielt das ad hoc Kriterium FAIC bezüglich der Medianwerte der global korrigierten Diskrepanzen die besten Ergebnisse, ebenso kann den Boxplots entnommen werden, daß das FAIC hinsichtlich der Stichprobenvariabilität mindestens ebenso gut abschneidet wie die anderen nichtparametrischen Kriterien. Die schlechtesten Ergebnisse bei dem Four-Peak-Prozeß erzielt das FSIC. Für die glatt verlaufenden Spektren des Smooth- bzw. One-Peak-Prozesses kehren sich die Aussagen um: Die besten Ergebnisse erzielt das FSIC, die schlechtesten das FAIC. Besonders auffällig ist hierbei die große Stichprobenvariabilität des FAIC bei Anwendung der Hamming-Schätzung (vgl. Abb. 6.17 (b) und 6.23 (b)). Bei dem MA-Smooth-Prozeß werden mit dem HQ-Kriterium die besten Ergebnisse bezüglich der Medianwerte der global korrigierten Diskrepanz erzielt (vgl. Tabelle 6.2). Auch bei dem MA-Smooth-Prozeß fällt die große Stichprobenvariabilität des FAIC bei Anwendung der Hamming-Schätzung auf (vgl. Abb. 6.29 (b)). Ebenso wie bei der Analyse der gemittelten Kurvenverläufe der global korrigierten Diskrepanzen kann festgestellt werden, daß für jeden datenerzeugenden Prozeß mindestens ein nichtparametrisches ad hoc Kriterium existiert, das bessere Resultate erzielt als das theoretisch fundierte FAIC.

Die Simulationsergebnisse können damit wie folgt zusammengefaßt werden. Bei den parametrischen Modellselektionskriterien erzielt das SIC, unabhängig von dem datenerzeugenden Prozeß, die besten Ergebnisse. Allerdings zeigt das Beispiel des MA-Smooth-Prozesses, daß wenn es sich bei dem datenerzeugenden Prozeß nicht um einen autoregressiven Prozeß handelt, mit den nichtparametrischen Schätzverfahren bessere

Ergebnisse erzielt werden können.

Beschränkt man sich auf einen Vergleich der nichtparametrischen Modellselektionskriterien, so kann kein gleichmäßig bestes Kriterium ermittelt werden. In Abhängigkeit von dem datenerzeugenden Prozeß können entweder mit dem ad hoc Kriterium FAIC, dem ad hoc Kriterium FSIC oder dem ad hoc Kriterium FHQ die besten Ergebnisse erzielt werden. Damit setzt die Anwendung der nichtparametrischen Modellselektionskriterien a priori Information über das Spektrum des datenerzeugenden Prozesses voraus. Kann z.B. davon ausgegangen werden, daß es sich um ein relativ glatt verlaufendes Spektrum ohne sehr stark ausgeprägte Peaks handelt, dann ist bei Anwendung eines nichtparametrischen Schätzverfahrens das Kriterium FSIC oder FHQ empfehlenswert.

Bei der Interpretation der Ergebnisse ist allerdings zu beachten, daß Aussagen, die lediglich auf Simulationsergebnissen beruhen, keine Allgemeingültigkeit besitzen können. Es ist auch nochmals besonders darauf hinzuweisen, daß spezielle Fragestellungen die Anwendung besonderer Diskrepanzmaße erforderlich machen können (vgl. Kapitel 5). Insgesamt kann jedoch festgestellt werden, daß die optimale Auswahl eines automatischen Spektralschätzverfahrens a priori Information über den datenerzeugenden Prozeß notwendig macht. Dementsprechend ist bei praktischen Anwendungen darauf zu achten, möglichst viel a priori Information bei der Auswahl eines Schätzverfahrens bzw. Modellselektionskriteriums einfließen zu lassen.

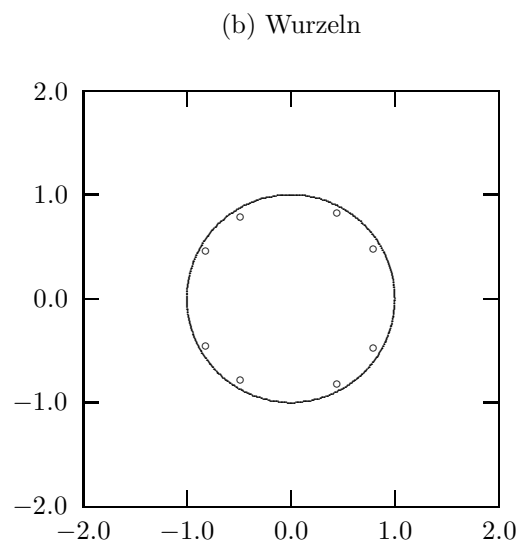
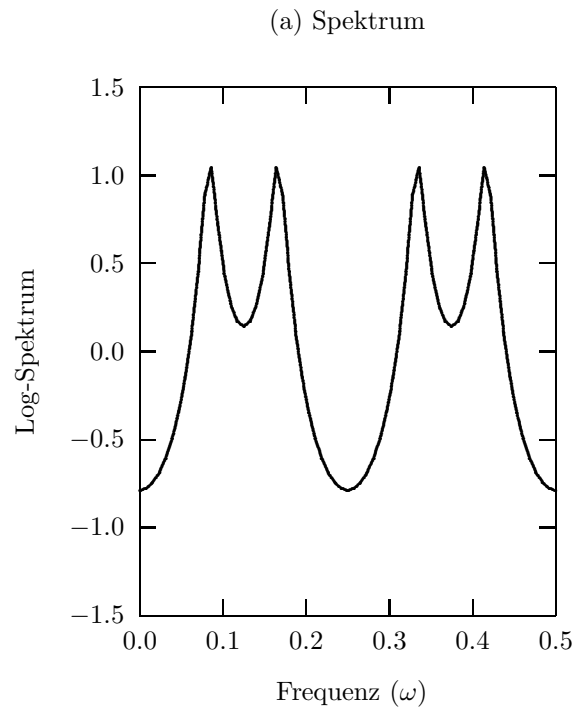


Abb. 6.6: (a) Theoretisches Spektrum des AR(8)-(Four-Peak)-Prozesses (6.27), (b) Lage der Wurzeln gegenüber dem Einheitskreis

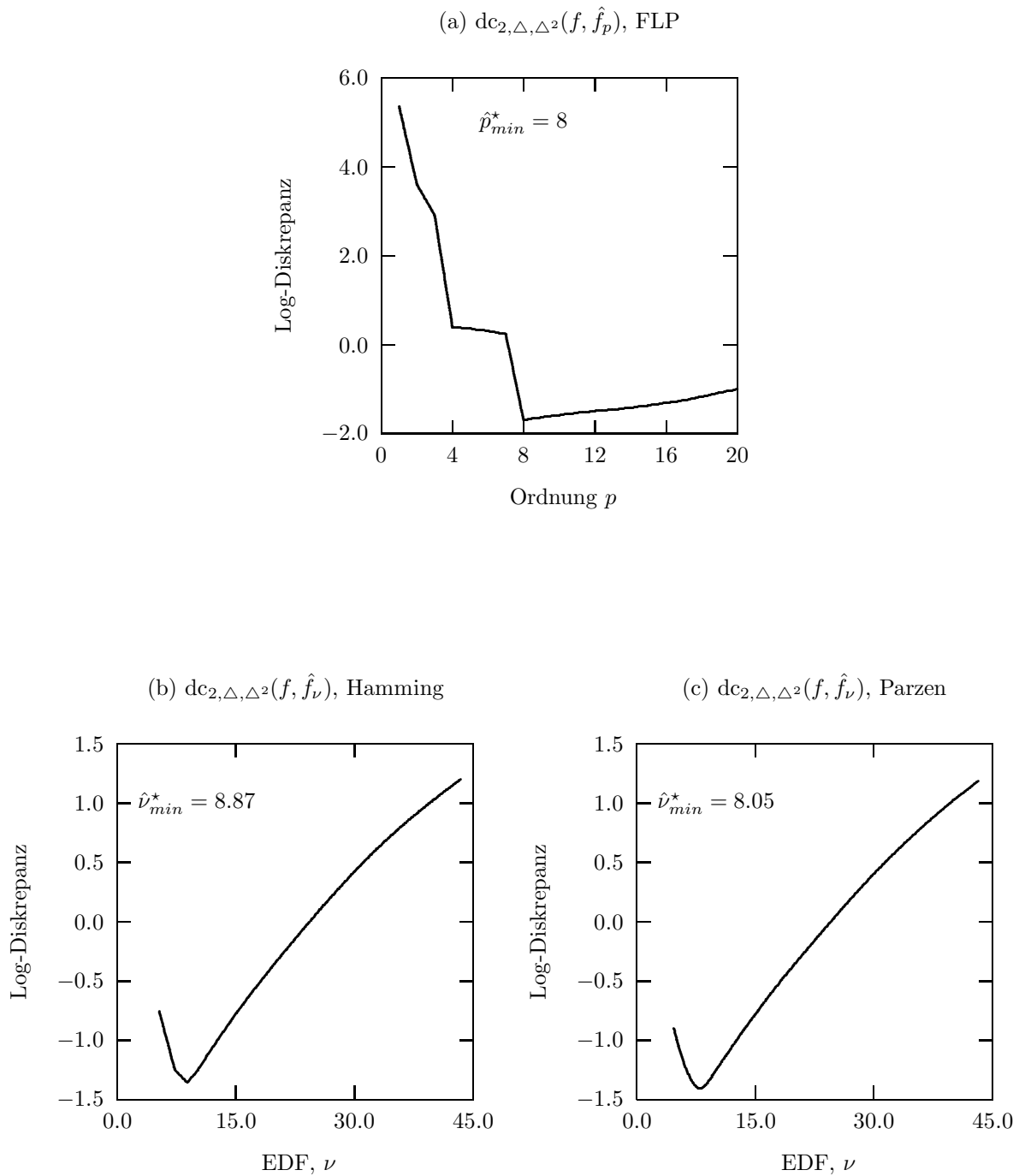


Abb. 6.7: Referenzdiskrepanzen, Schätzungen für $E[dc_{2,\Delta,\Delta^2}(f, \hat{f})]$, gemittelt über 100 Realisationen des AR(8)-(Four-Peak)-Prozesses (6.27) mit $N = 128$, $N_E = 128$, $N_\omega = 65$.

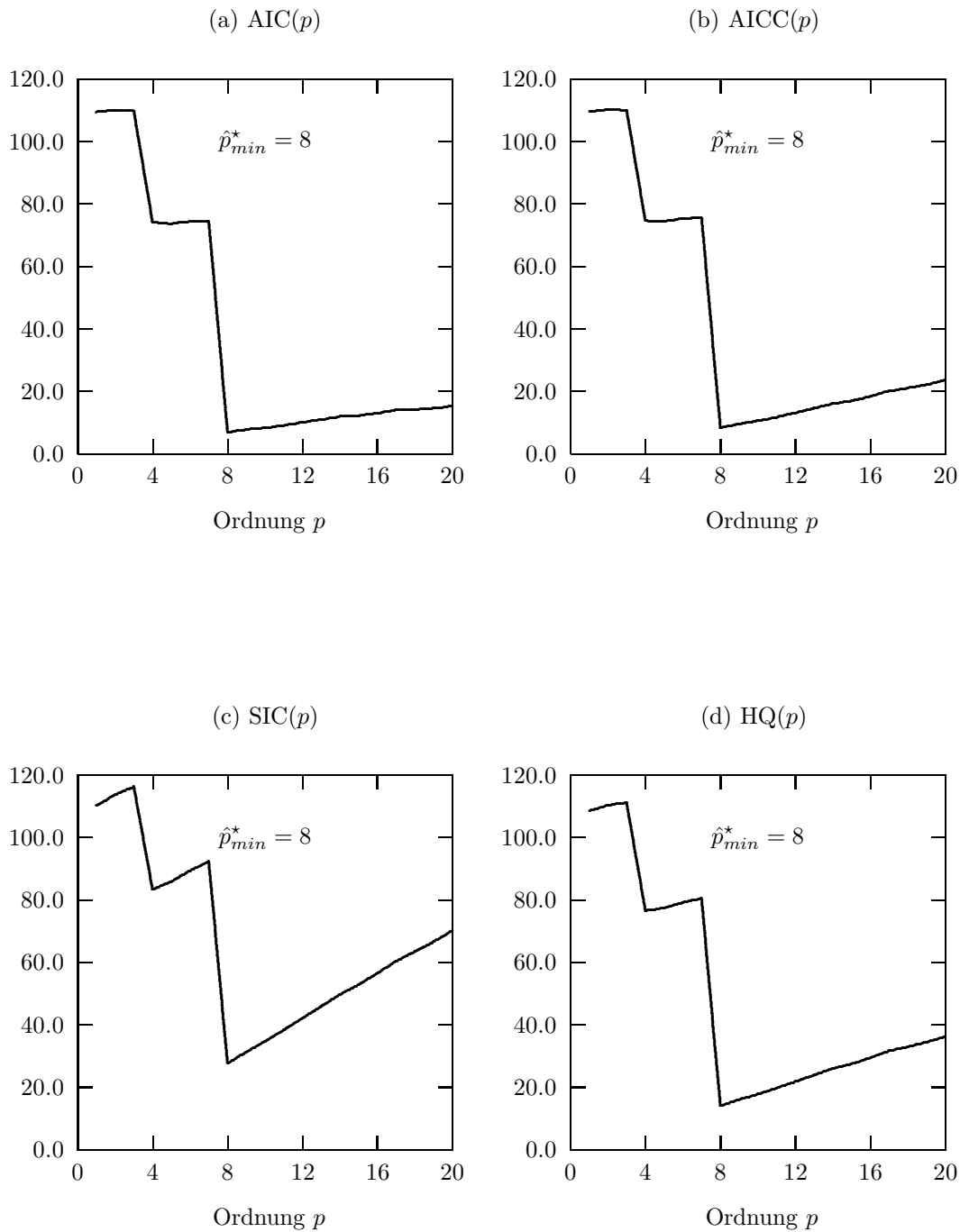


Abb. 6.8: Vergleich von Modellselektionskriterien, FLP-Methode, Schätzungen für $E[\text{AIC}(p)]$, $E[\text{AICC}(p)]$, $E[\text{SIC}(p)]$, $E[\text{HQ}(p)]$ gemittelt über 100 Realisationen des AR(8)-(Four-Peak)-Prozesses (6.27) mit $N = 128$, $N_E = 128$, $N_\omega = 65$. Minimum der Referenzdiscrepanz: $\hat{p}_{min}^* = 8$.

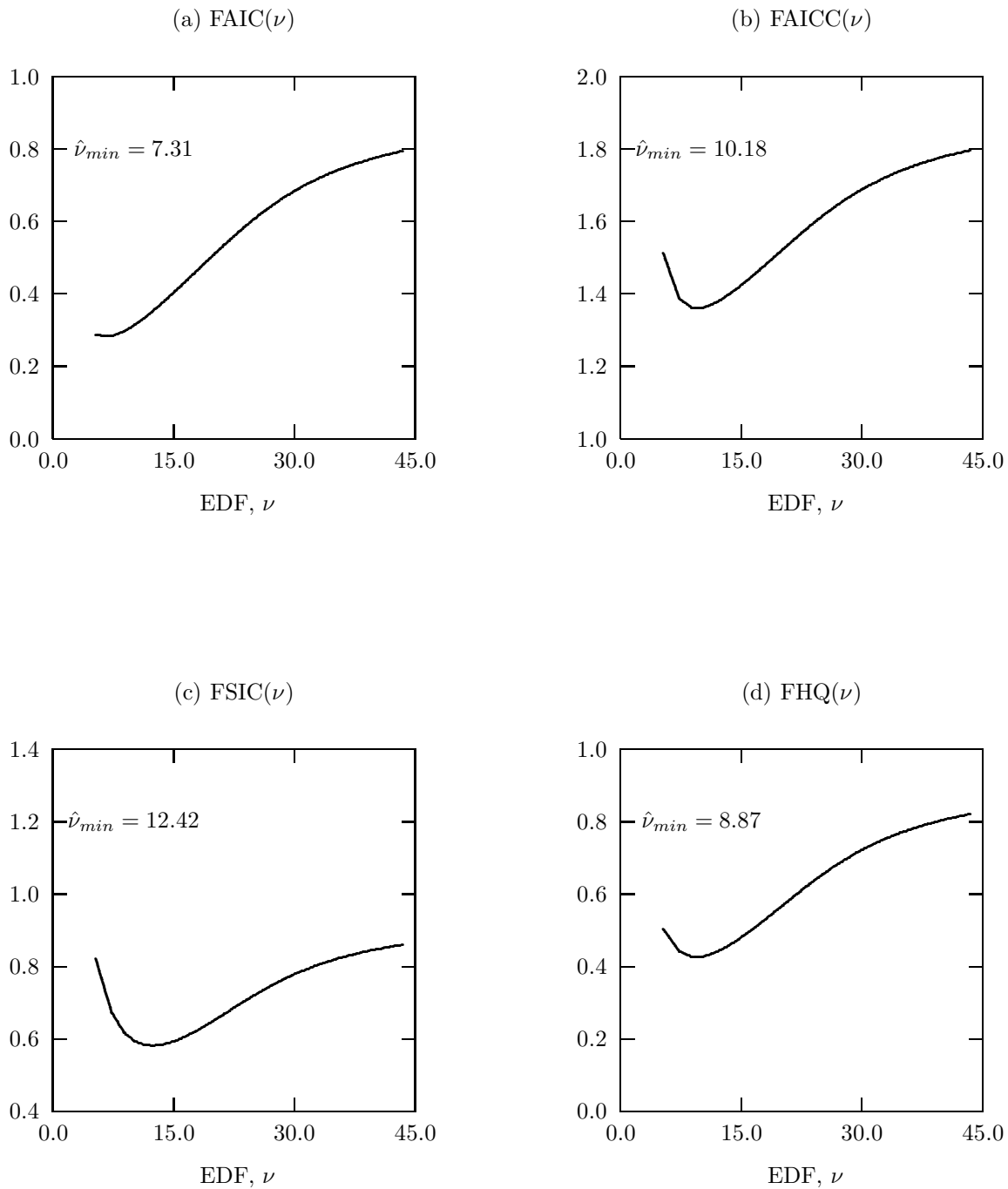


Abb. 6.9: Vergleich von Modellselektionskriterien, iteratives Hamming-Verfahren, Schätzungen für $E[\text{FAIC}(\nu)]$, $E[\text{FAICC}(\nu)]$, $E[\text{FSIC}(\nu)]$, $E[\text{FHQ}(\nu)]$ gemittelt über 100 Realisationen des AR(8)-(Four-Peak)-Prozesses (6.27) mit $N = 128$, $N_E = 128$, $N_\omega = 65$. Minimum der Referenzdiskrepanz: $\hat{\nu}_{min}^* = 8.87$.

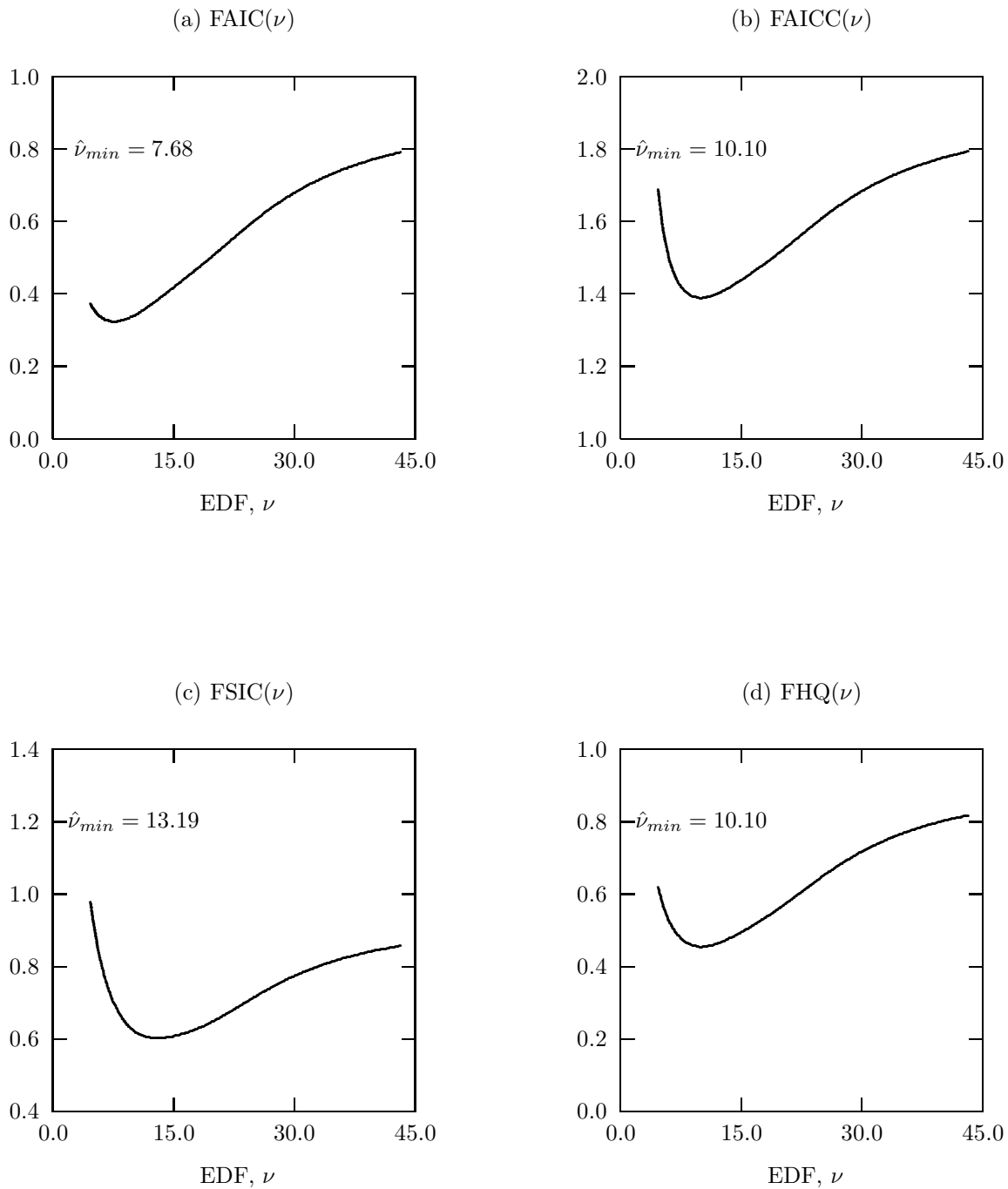


Abb. 6.10: Vergleich von Modellselektionskriterien, Parzen-Schätzer, Schätzungen für $E[\text{FAIC}(\nu)]$, $E[\text{FAICC}(\nu)]$, $E[\text{FSIC}(\nu)]$, $E[\text{FHQ}(\nu)]$ gemittelt über 100 Realisationen des AR(8)-(Four-Peak)-Prozesses (6.27) mit $N = 128$, $N_E = 128$, $N_\omega = 65$. Minimum der Referenzdiskrepanz: $\hat{\nu}_{min}^* = 8.05$.

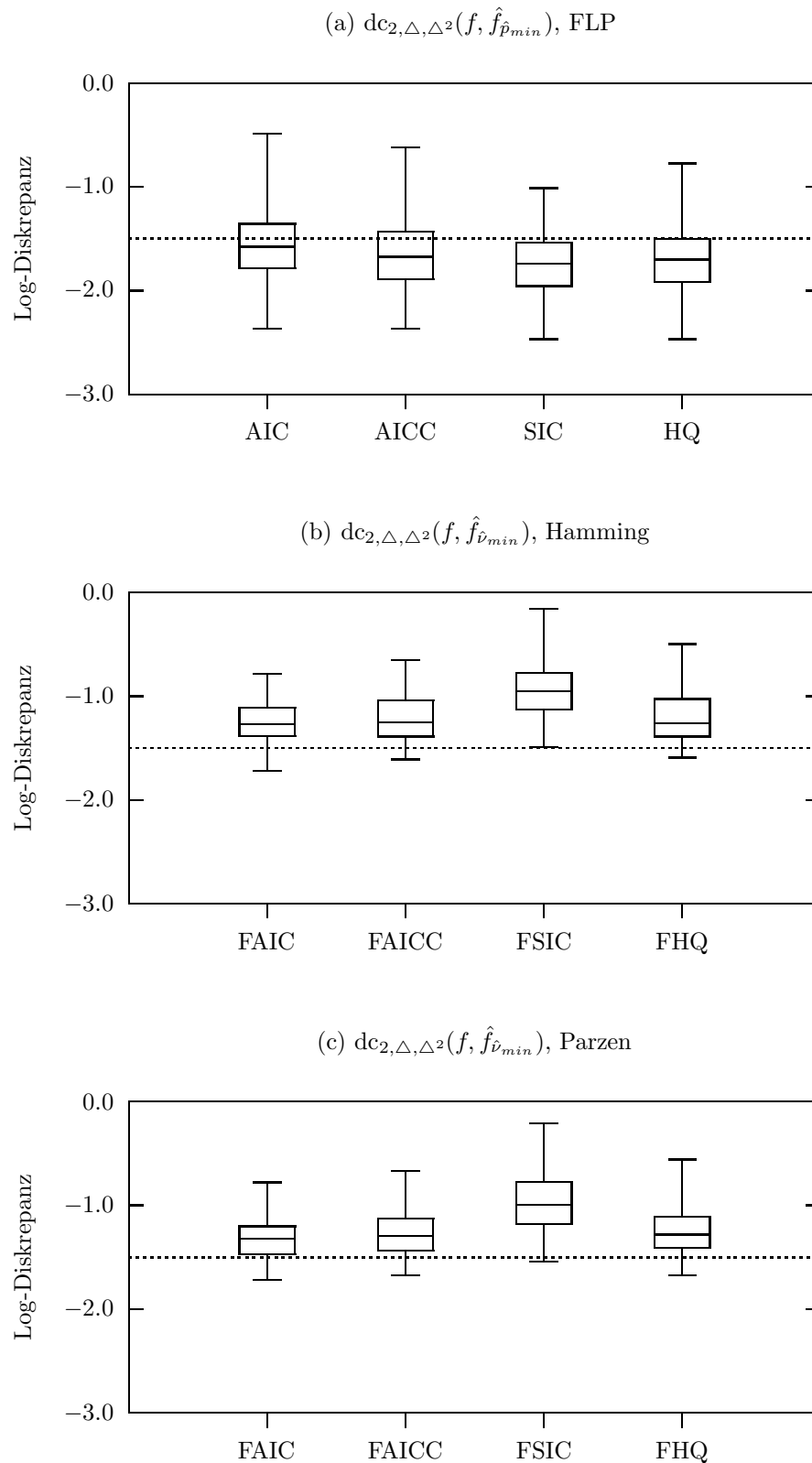


Abb. 6.11: Vergleich von parametrischen und nichtparametrischen Modellselektionskriterien, Boxplots (Minimum, 1. Quartil, Median, 3. Quartil, Maximum) basierend auf 100 Realisationen des AR(8)-(Four-Peak)-Prozesses (6.27) mit $N = 128$, $N_E = 128$, $N_\omega = 65$. Die Diskrepanz wurde für alle Kriterien jeweils mit der entsprechenden optimalen Spektralschätzung für jede Realisation berechnet.

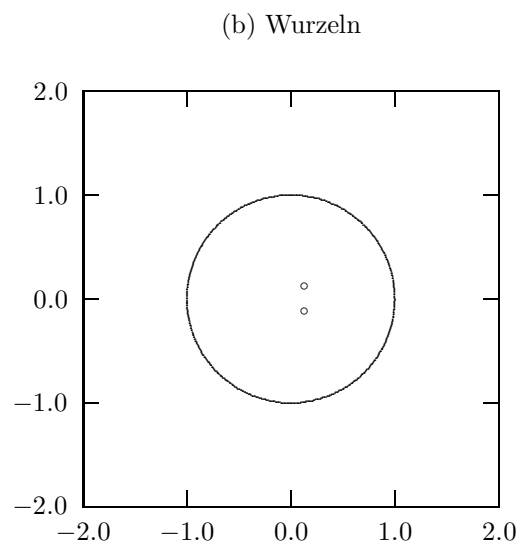
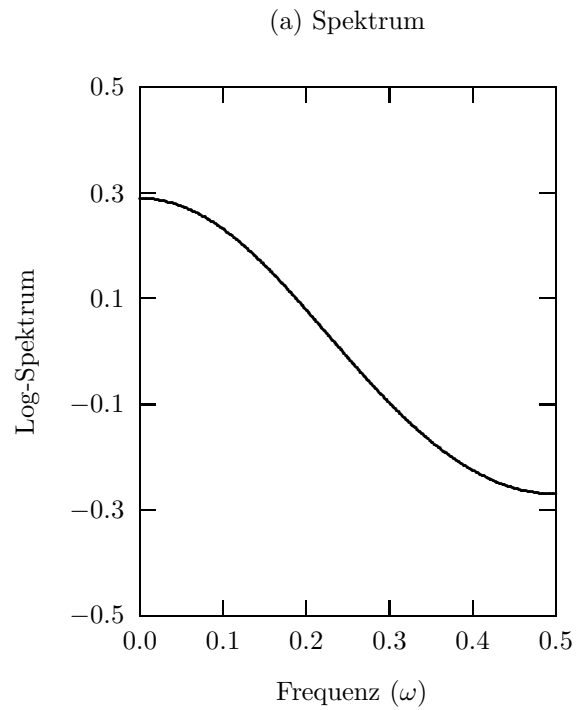


Abb. 6.12: (a) Theoretisches Spektrum des AR(2)-(Smooth)-Prozesses (6.28), (b) Lage der Wurzeln gegenüber dem Einheitskreis

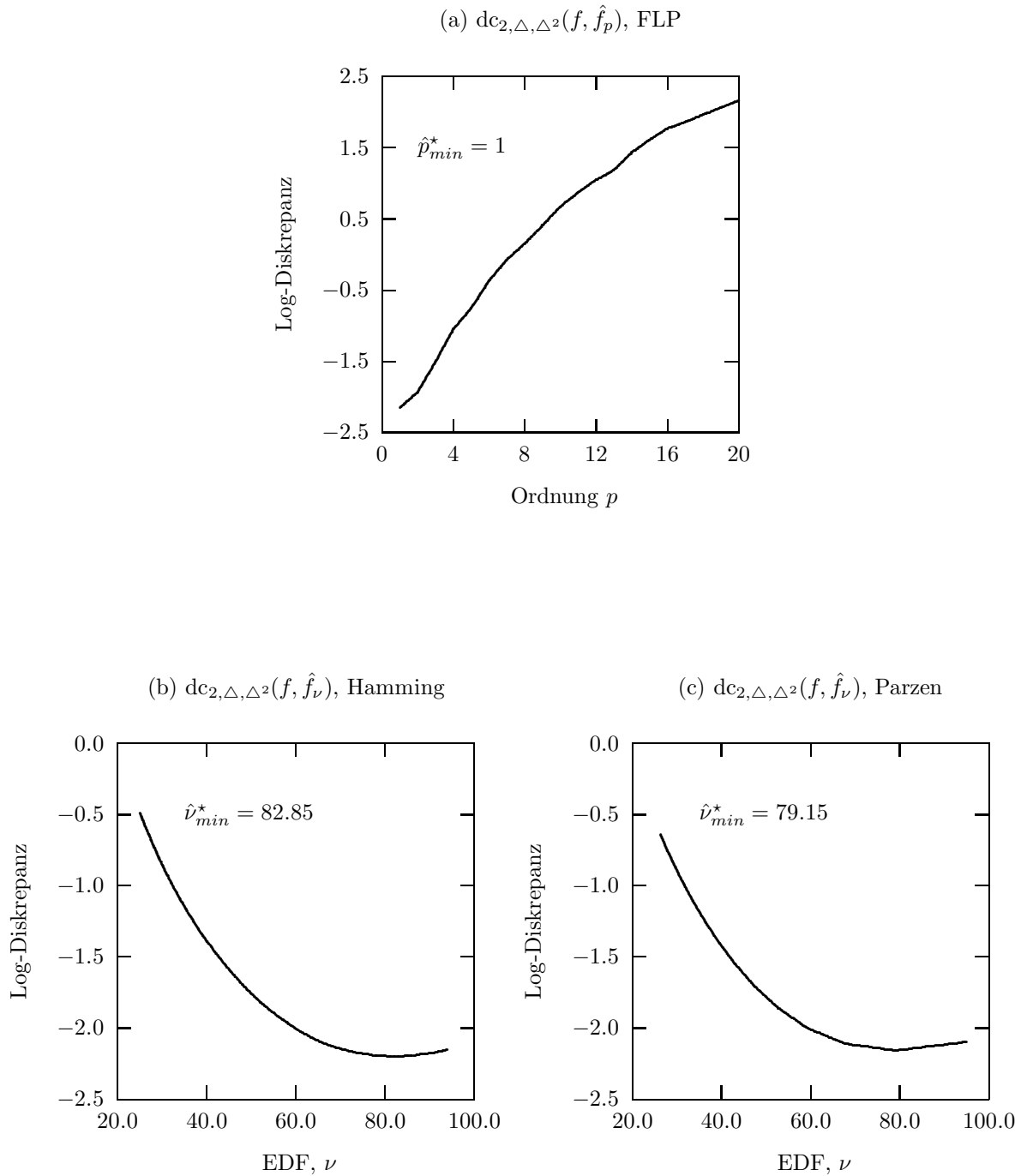


Abb. 6.13: Referenzdiskrepanzen, Schätzungen für $E[dc_{2,\Delta,\Delta^2}(f, \hat{f})]$, gemittelt über 100 Realisationen des AR(2)-(Smooth)-Prozesses (6.28) mit $N = 128$, $N_E = 128$, $N_\omega = 65$.

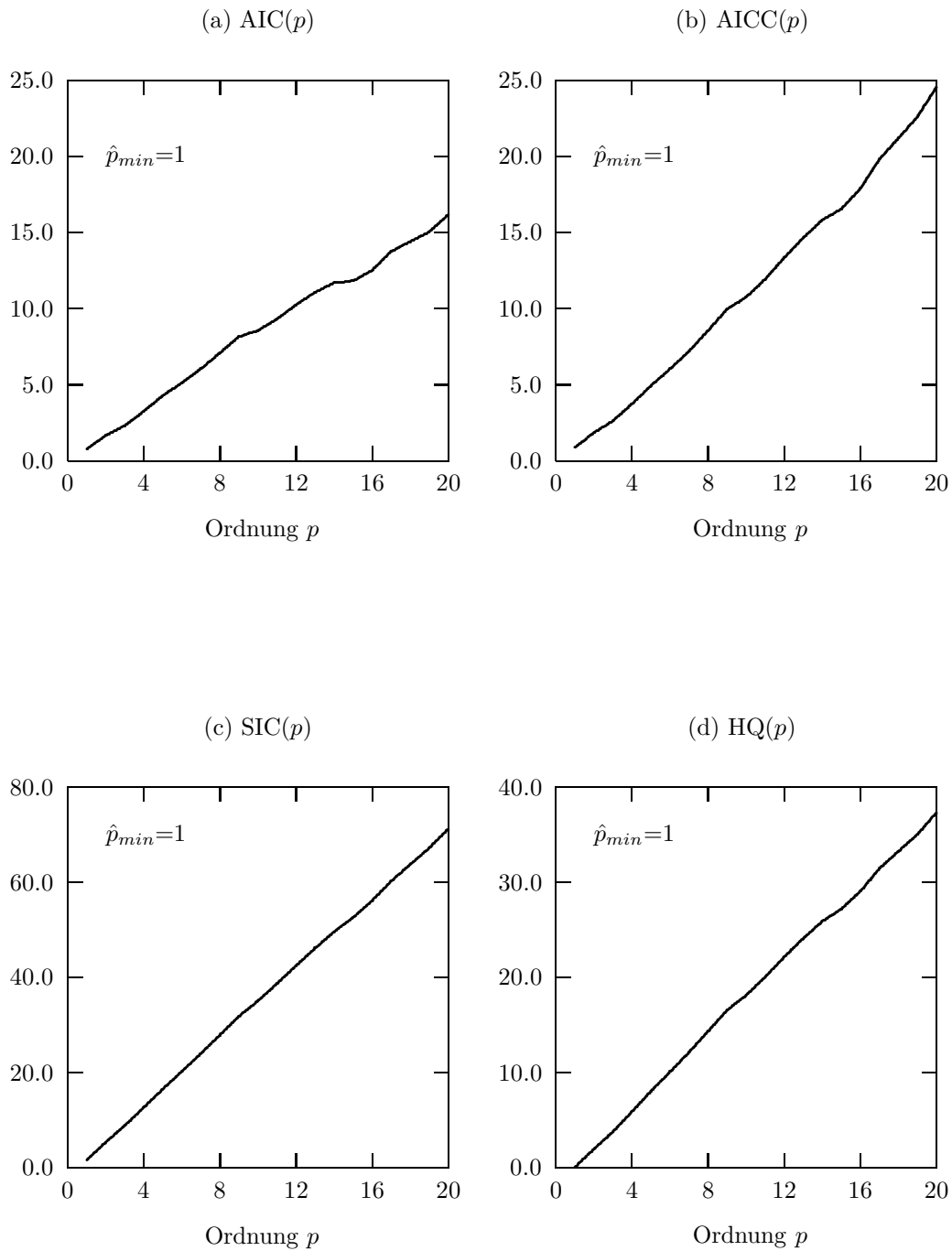


Abb. 6.14: Vergleich von Modellselektionskriterien, FLP-Methode, Schätzungen für $E[AIC(p)]$, $E[AICC(p)]$, $E[SIC(p)]$, $E[HQ(p)]$ gemittelt über 100 Realisationen des AR(2)-(Smooth)-Prozesses (6.28) mit $N = 128$, $N_E = 128$, $N_\omega = 65$. Minimum der Referenzdiskrepanz: $\hat{p}_{min}^* = 1$.

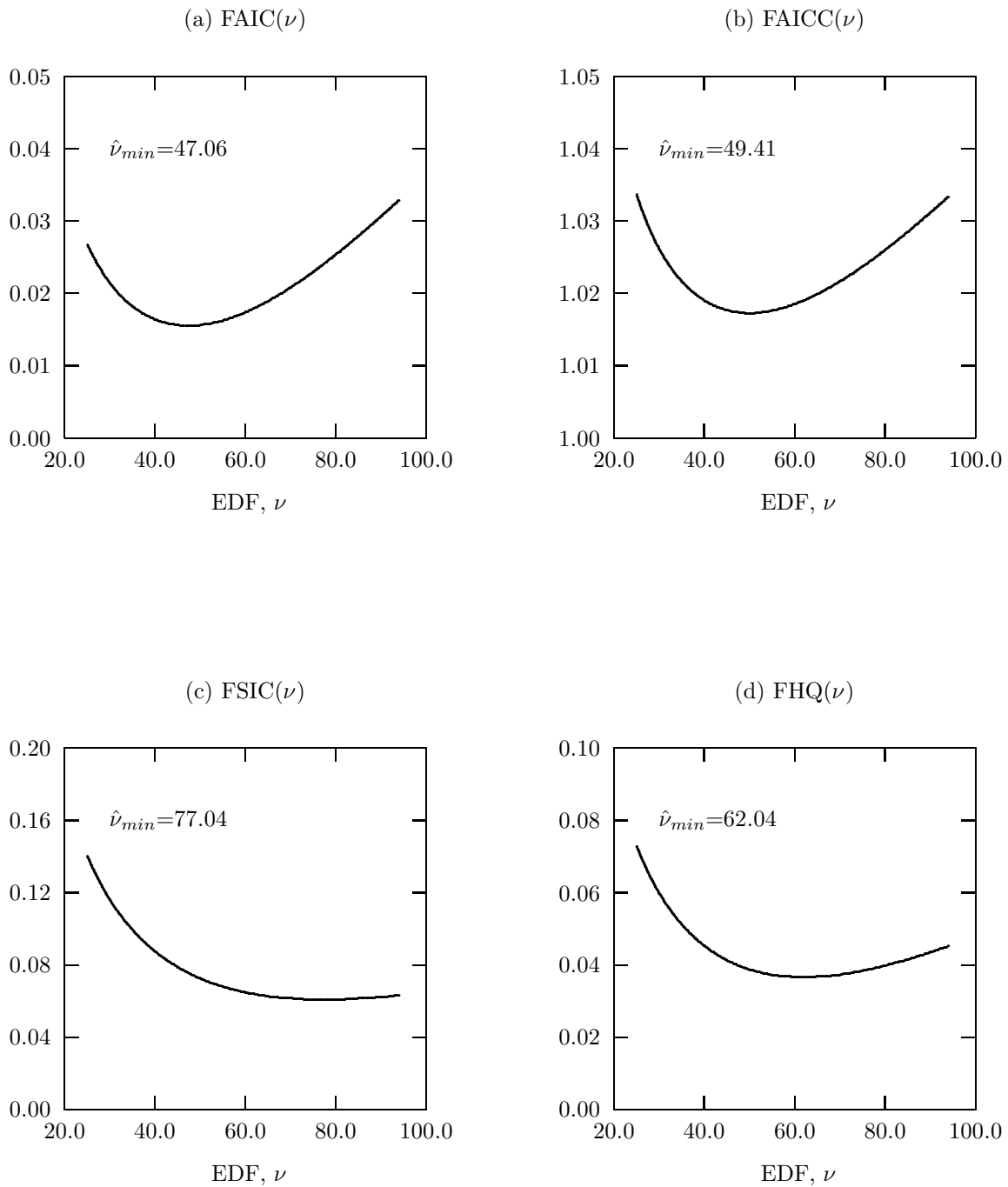


Abb. 6.15: Vergleich von Modellselektionskriterien, iteratives Hamming-Verfahren, Schätzungen für $E[\text{FAIC}(\nu)]$, $E[\text{FAICC}(\nu)]$, $E[\text{FSIC}(\nu)]$, $E[\text{FHQ}(\nu)]$ gemittelt über 100 Realisationen des AR(2)-(Smooth)-Prozesses (6.28) mit $N = 128$, $N_E = 128$, $N_\omega = 65$. Minimum der Referenzdiskrepanz: $\hat{\nu}_{min}^* = 82.85$.

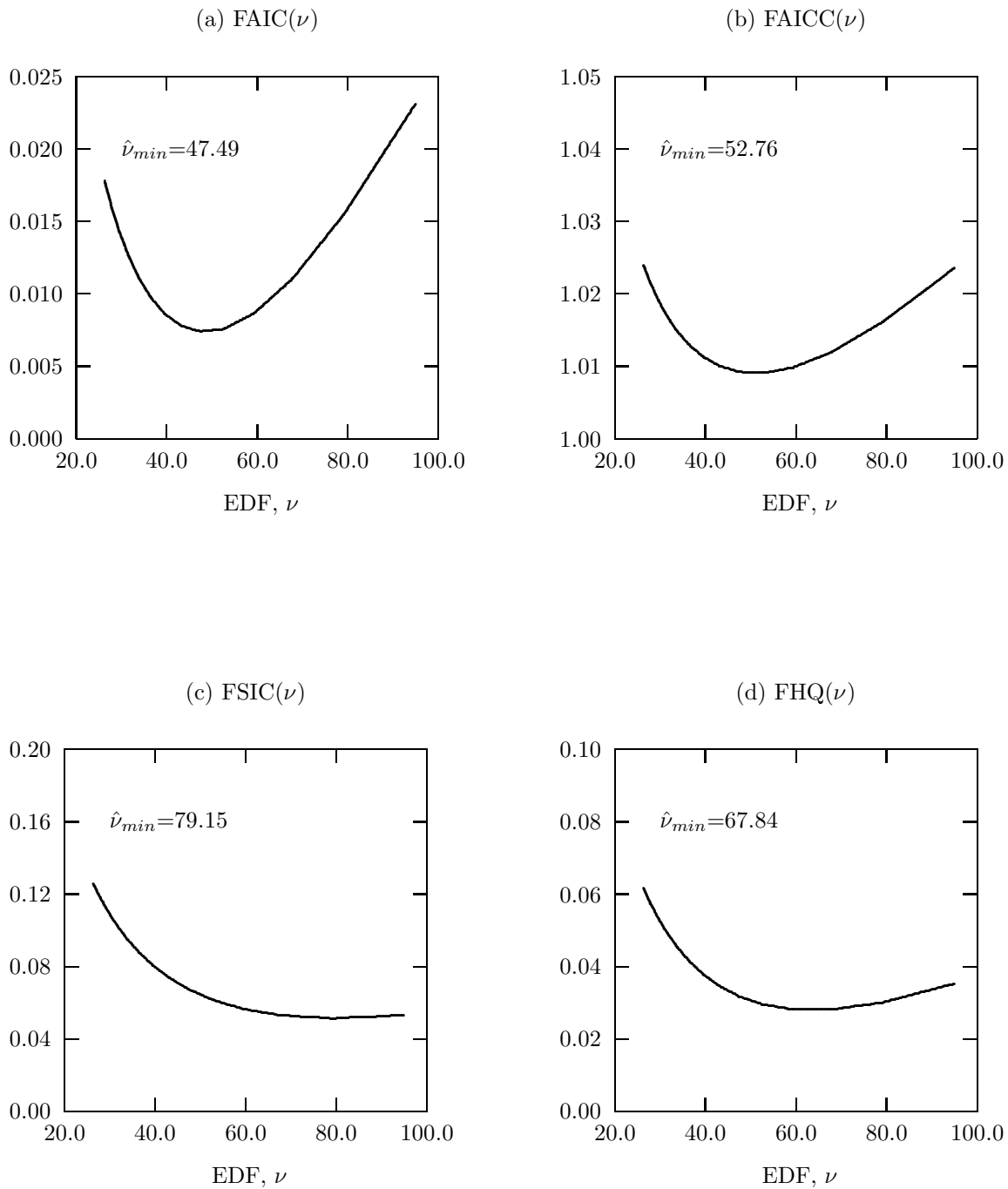


Abb. 6.16: Vergleich von Modellselektionskriterien, Parzen-Schätzer, Schätzungen für $E[\text{FAIC}(\nu)]$, $E[\text{FAICC}(\nu)]$, $E[\text{FSIC}(\nu)]$, $E[\text{FHQ}(\nu)]$ gemittelt über 100 Realisationen des AR(2)-(Smooth)-Prozesses (6.28) mit $N = 128$, $N_E = 128$, $N_\omega = 65$. Minimum der Referenzdiskrepanz: $\hat{\nu}_{min}^* = 79.15$.

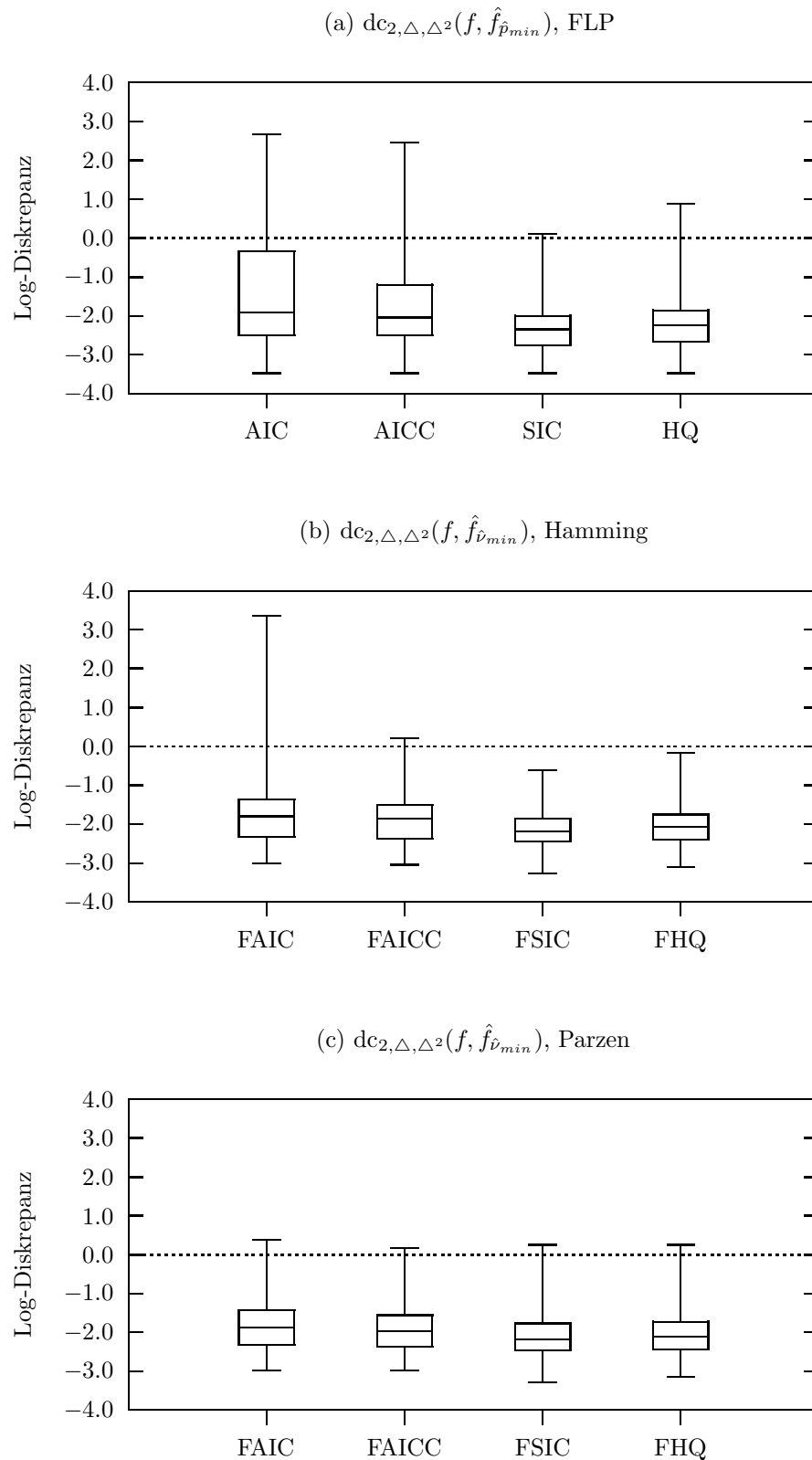


Abb. 6.17: Vergleich von parametrischen und nicht-parametrischen Modellselektionskriterien, Boxplots mit Minimum, 1. Quartil, Median, 3. Quartil und Maximum basierend auf 100 Realisationen des AR(2)-(Smooth)-Prozesses (6.28) mit $N = 128$, $N_E = 128$, $N_\omega = 65$. Die Diskrepanz wurde für alle Kriterien jeweils mit der entsprechenden optimalen Spektralschätzung für jede Realisation berechnet.

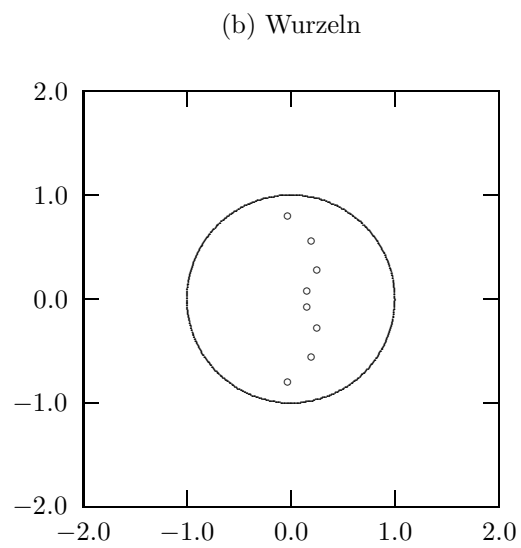
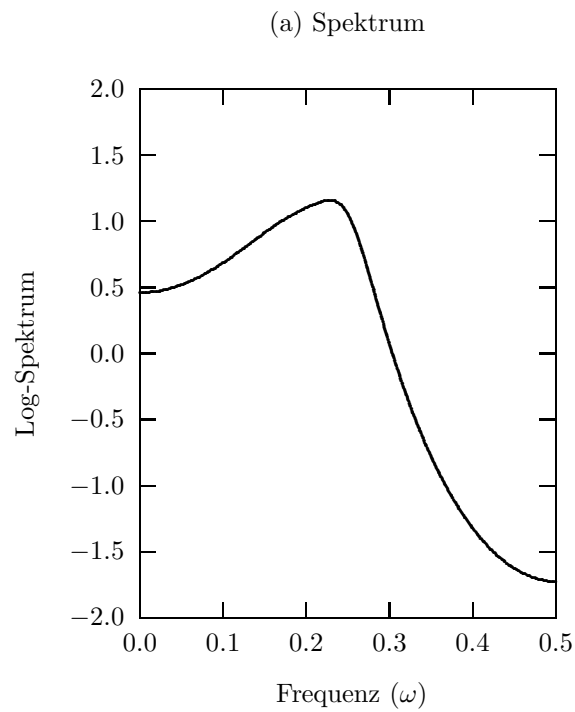


Abb. 6.18: (a) Theoretisches Spektrum des AR(8)-(One-Peak)-Prozesses (6.29), (b) Lage der Wurzeln gegenüber dem Einheitskreis

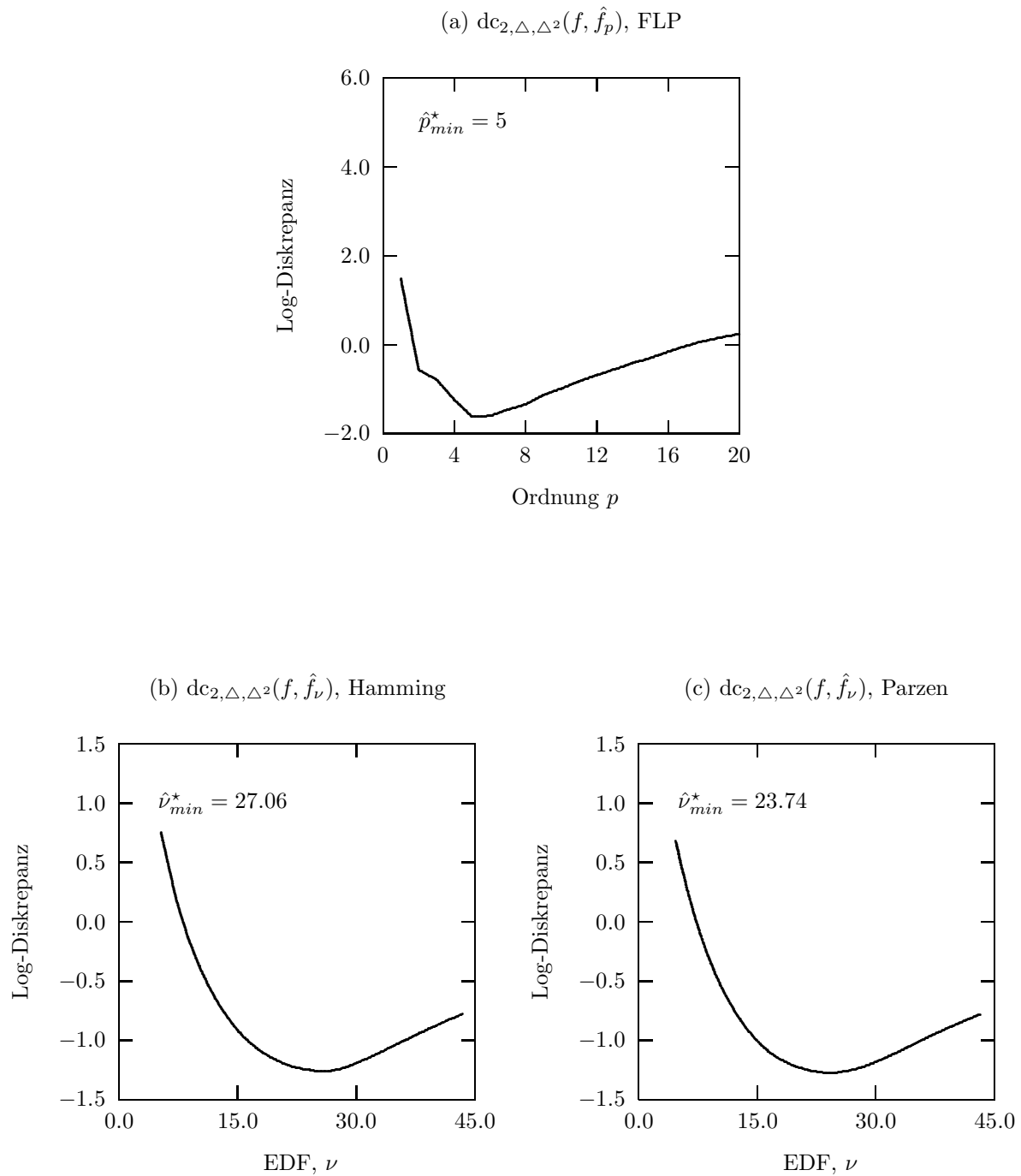


Abb. 6.19: Referenzdiskrepanzen, Schätzungen für $E[dc_{2,\Delta,\Delta^2}(f, \hat{f})]$, gemittelt über 100 Realisationen des AR(8)-(One-Peak)-Prozesses (6.29) mit $N = 128$, $N_E = 128$, $N_\omega = 65$.

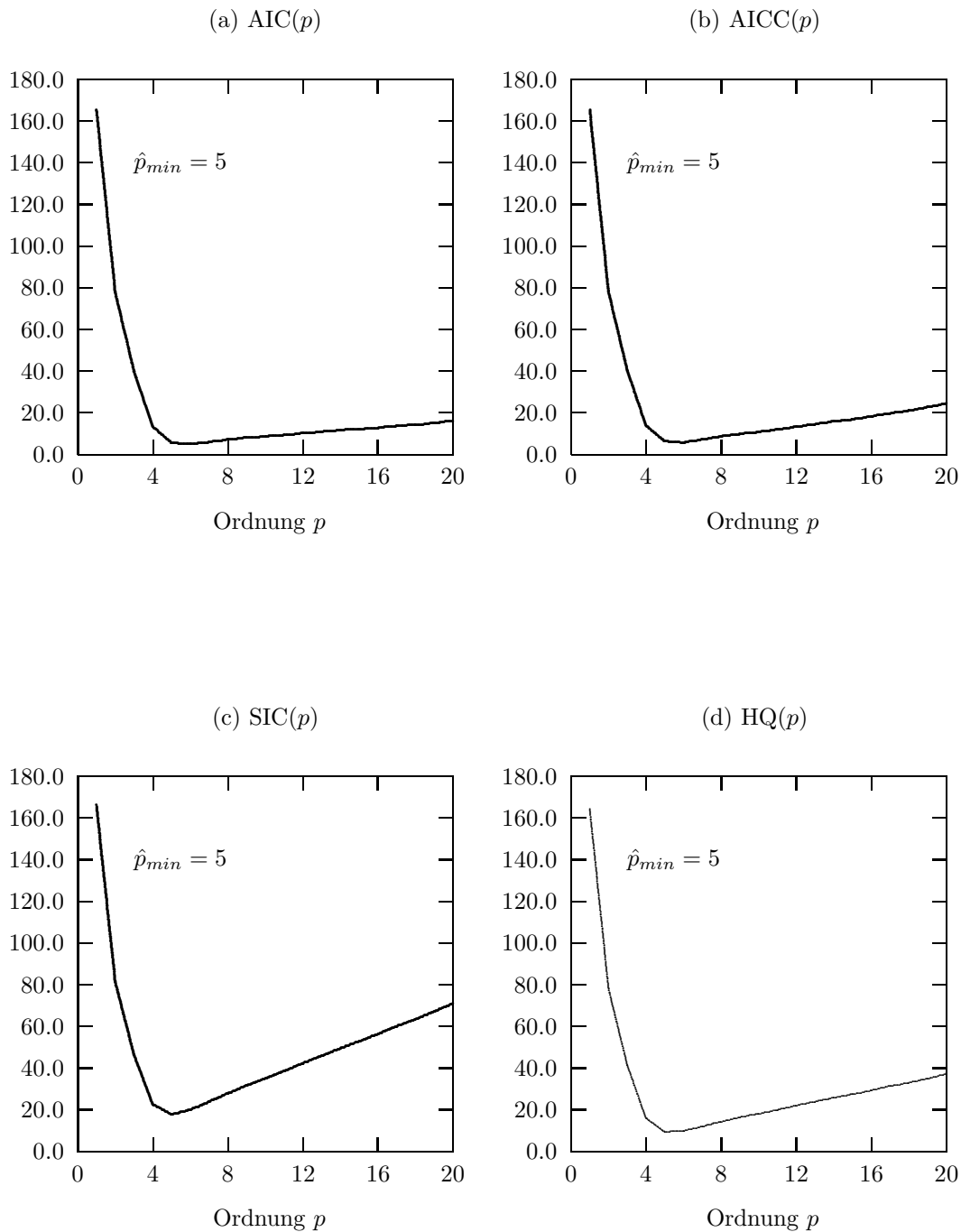


Abb. 6.20: Vergleich von Modellselektionskriterien, FLP-Methode, Schätzungen für $E[\text{AIC}(p)]$, $E[\text{AICC}(p)]$, $E[\text{SIC}(p)]$, $E[\text{HQ}(p)]$ gemittelt über 100 Realisationen des AR(8)-(One-Peak)-Prozesses (6.29) mit $N = 128$, $N_E = 128$, $N_\omega = 65$. Minimum der Referenzdiscrepanz: $\hat{p}_{min}^* = 5$.

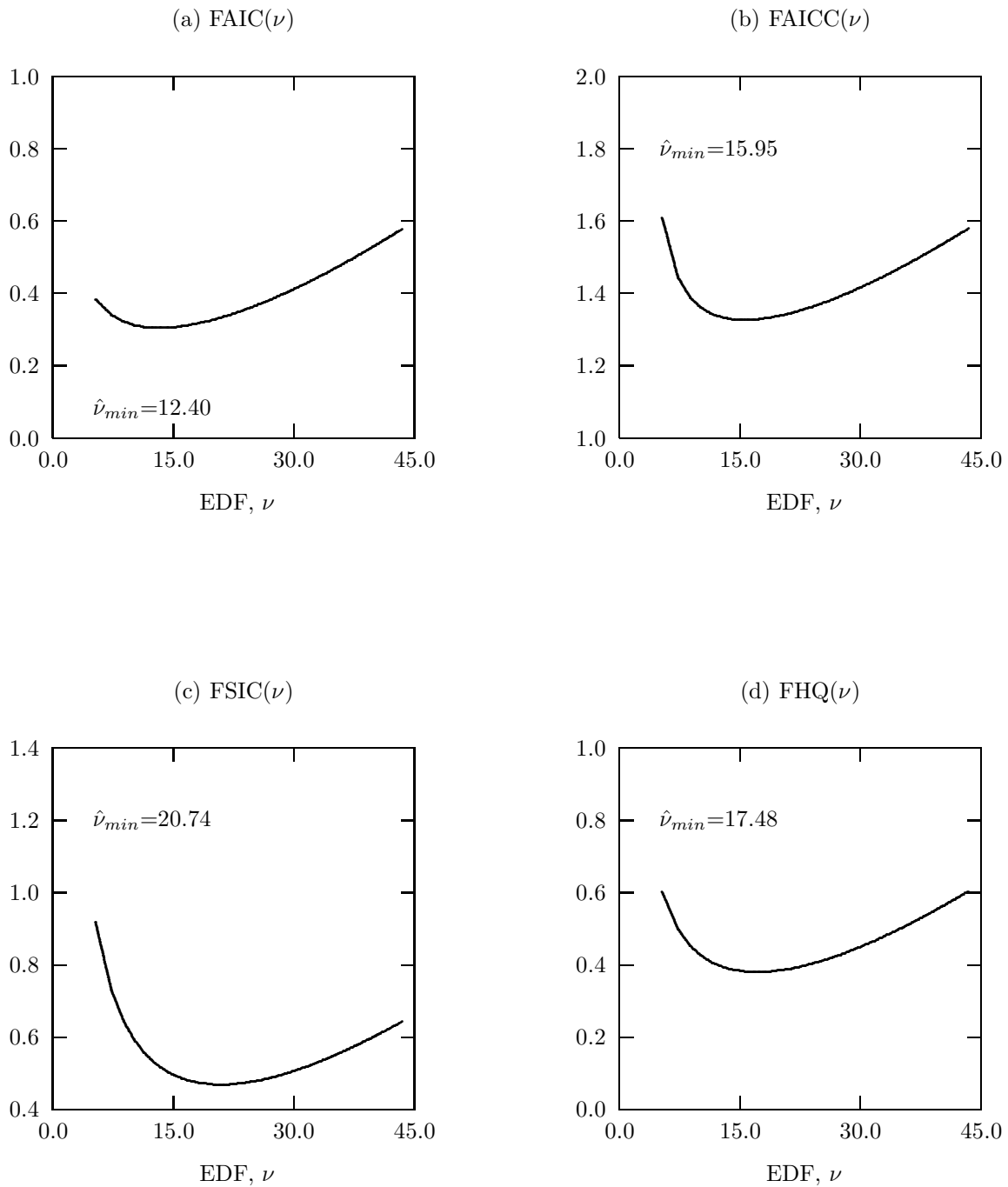


Abb. 6.21: Vergleich von Modellselektionskriterien, iteratives Hamming-Verfahren, Schätzungen für $E[\text{FAIC}(\nu)]$, $E[\text{FAICC}(\nu)]$, $E[\text{FSIC}(\nu)]$, $E[\text{FHQ}(\nu)]$ gemittelt über 100 Realisationen des AR(2)-(Smooth)-Prozesses (6.29) mit $N = 128$, $N_E = 128$, $N_\omega = 65$. Minimum der Referenzdiskrepanz: $\hat{\nu}_{min}^* = 27.06$.

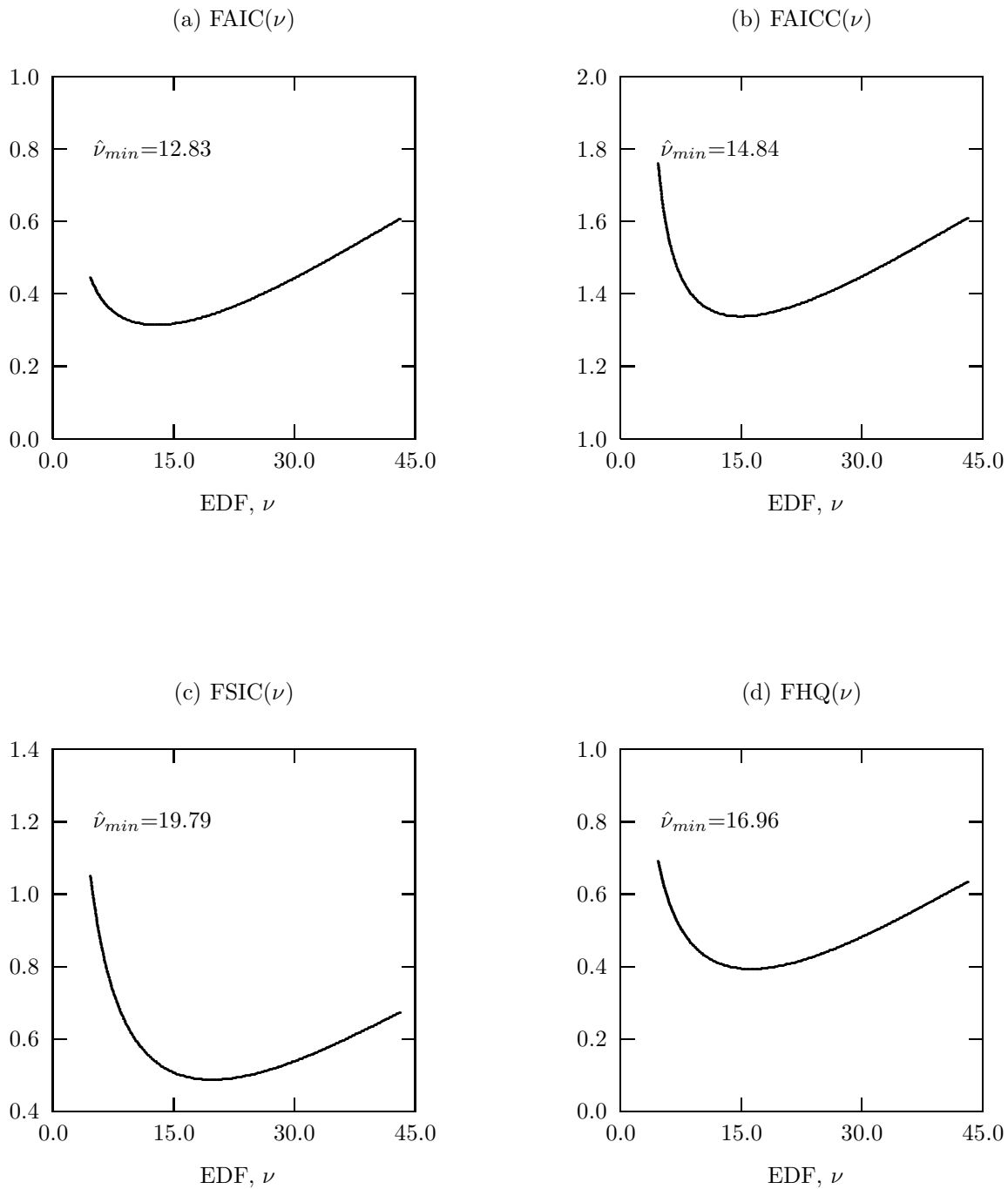


Abb. 6.22: Vergleich von Modellselektionskriterien, Parzen-Schätzer, Schätzungen für $E[\text{FAIC}(\nu)]$, $E[\text{FAICC}(\nu)]$, $E[\text{FSIC}(\nu)]$, $E[\text{FHQ}(\nu)]$ gemittelt über 100 Realisationen des AR(2)-(Smooth)-Prozesses (6.29) mit $N = 128$, $N_E = 128$, $N_\omega = 65$. Minimum der Referenzdiskrepanz: $\hat{\nu}_{min}^* = 23.74$.

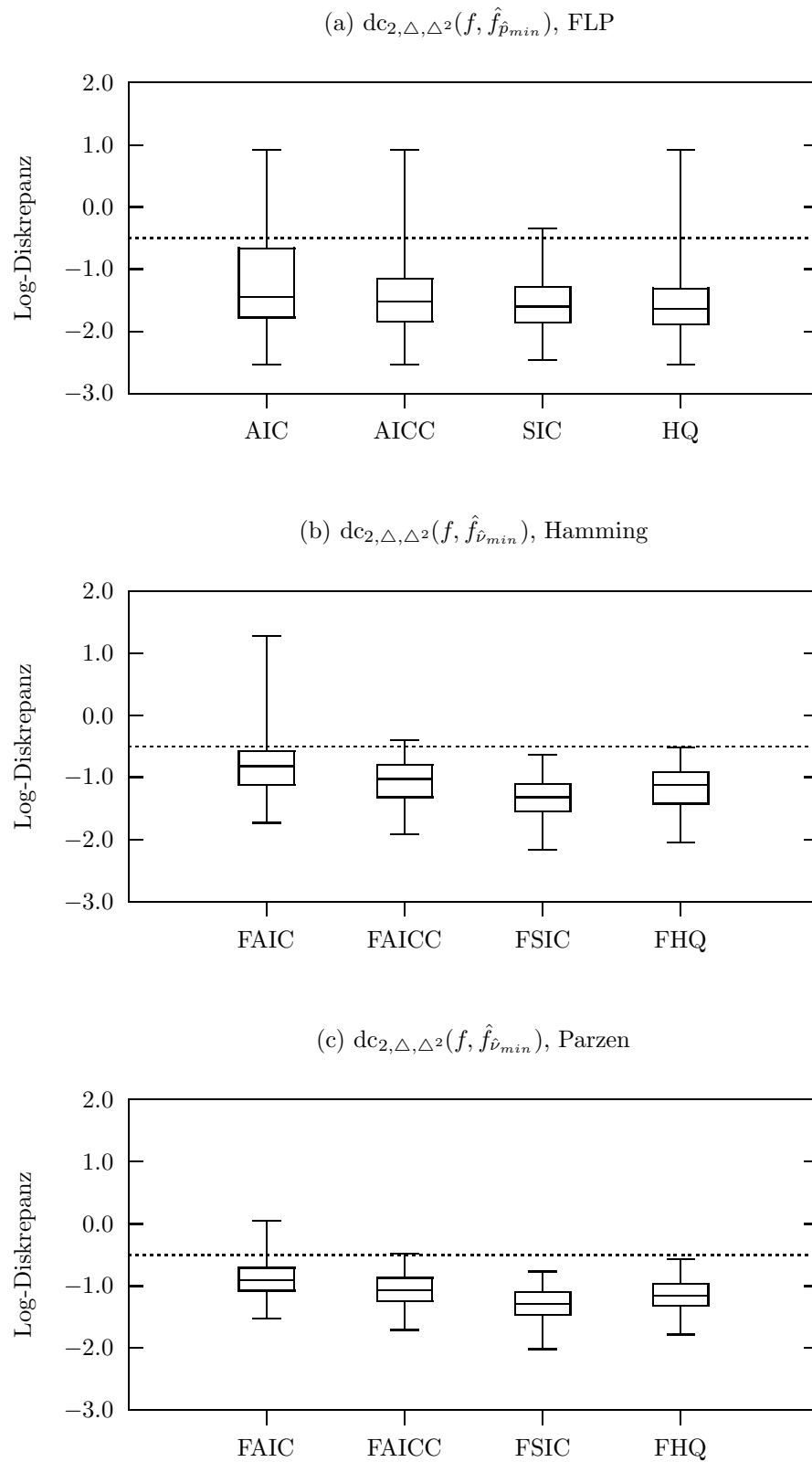


Abb. 6.23: Vergleich von parametrischen und nicht-parametrischen Modellselektionskriterien, Boxplots (Minimum, 1. Quartil, Median, 3. Quartil, Maximum) basierend auf 100 Realisationen des AR(8)-(One-Peak)-Prozesses (6.29) mit $N = 128$, $N_E = 128$, $N_\omega = 65$. Die Diskrepanz wurde für alle Kriterien jeweils mit der entsprechenden optimalen Spektralschätzung für jede Realisation berechnet.

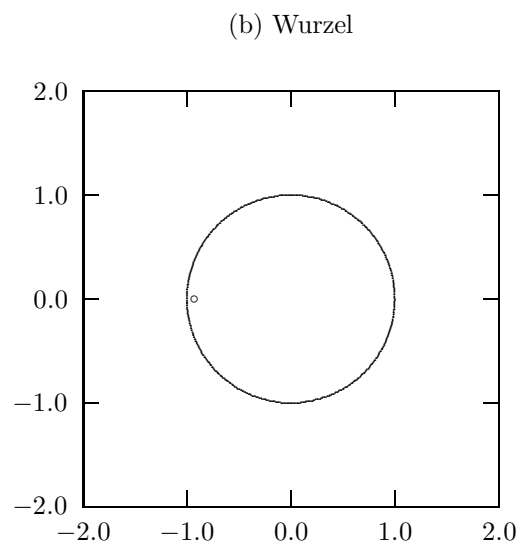
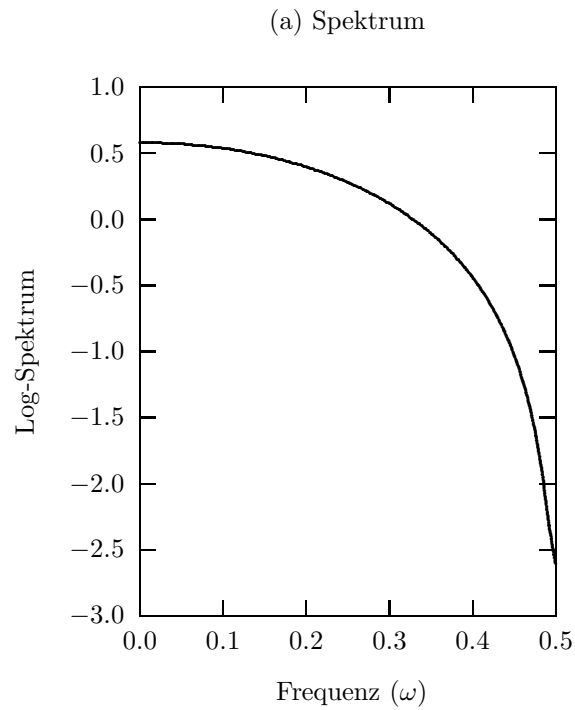


Abb. 6.24: (a) Theoretisches Spektrum des MA(1)-Prozesses $X_t = \epsilon_t + 0.95\epsilon_{t-1}$ (b) Lage der Wurzel gegenüber dem Einheitskreis

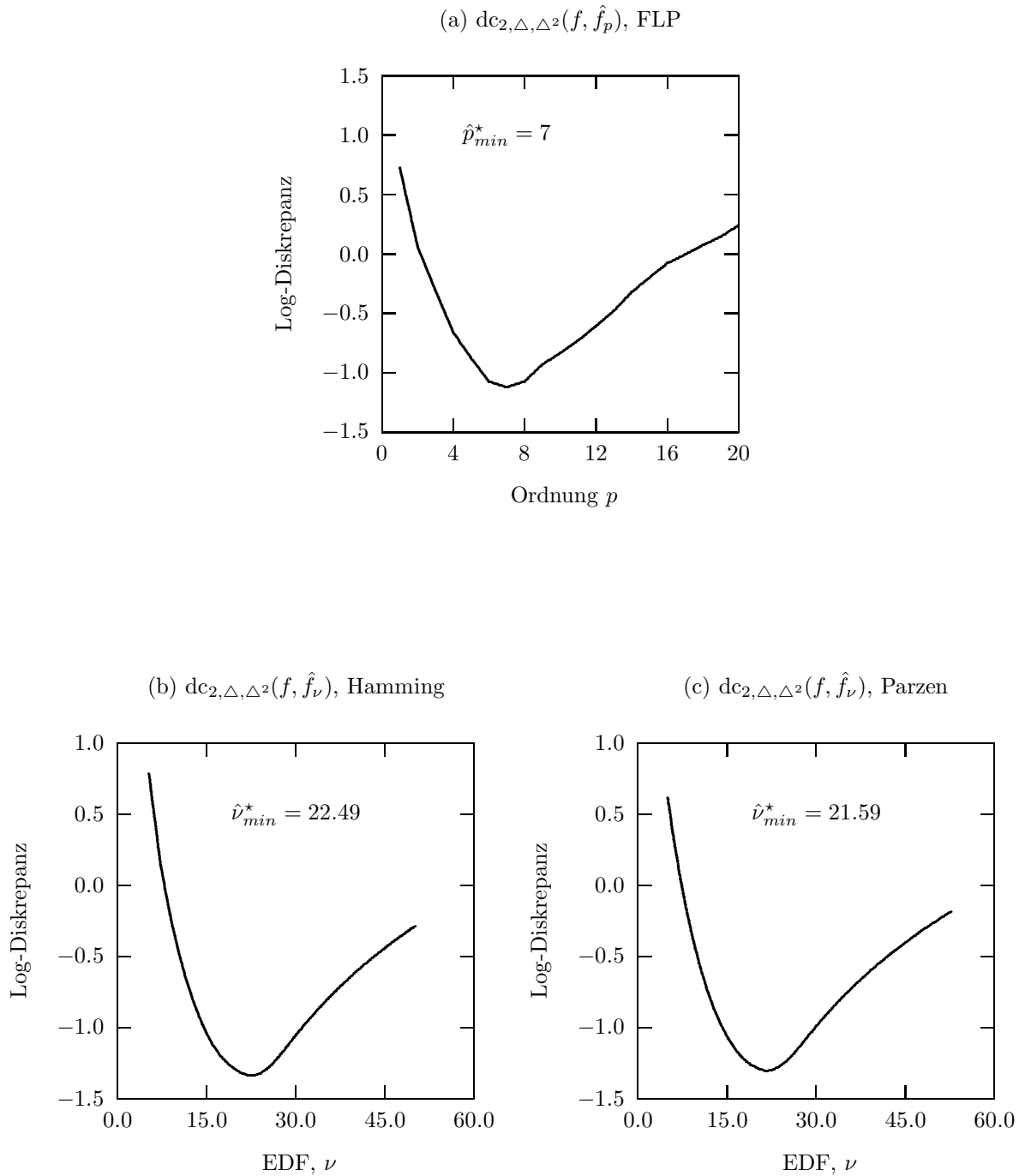


Abb. 6.25: Referenzdiskrepanzen, Schätzungen für $E[dc_{2,\Delta,\Delta^2}(f, \hat{f})]$, gemittelt über 100 Realisationen des MA(1)-(MA-Smooth)-Prozesses $X_t = \epsilon_t + 0.95\epsilon_{t-1}$ mit $N = 128$, $N_E = 128$, $N_\omega = 65$.

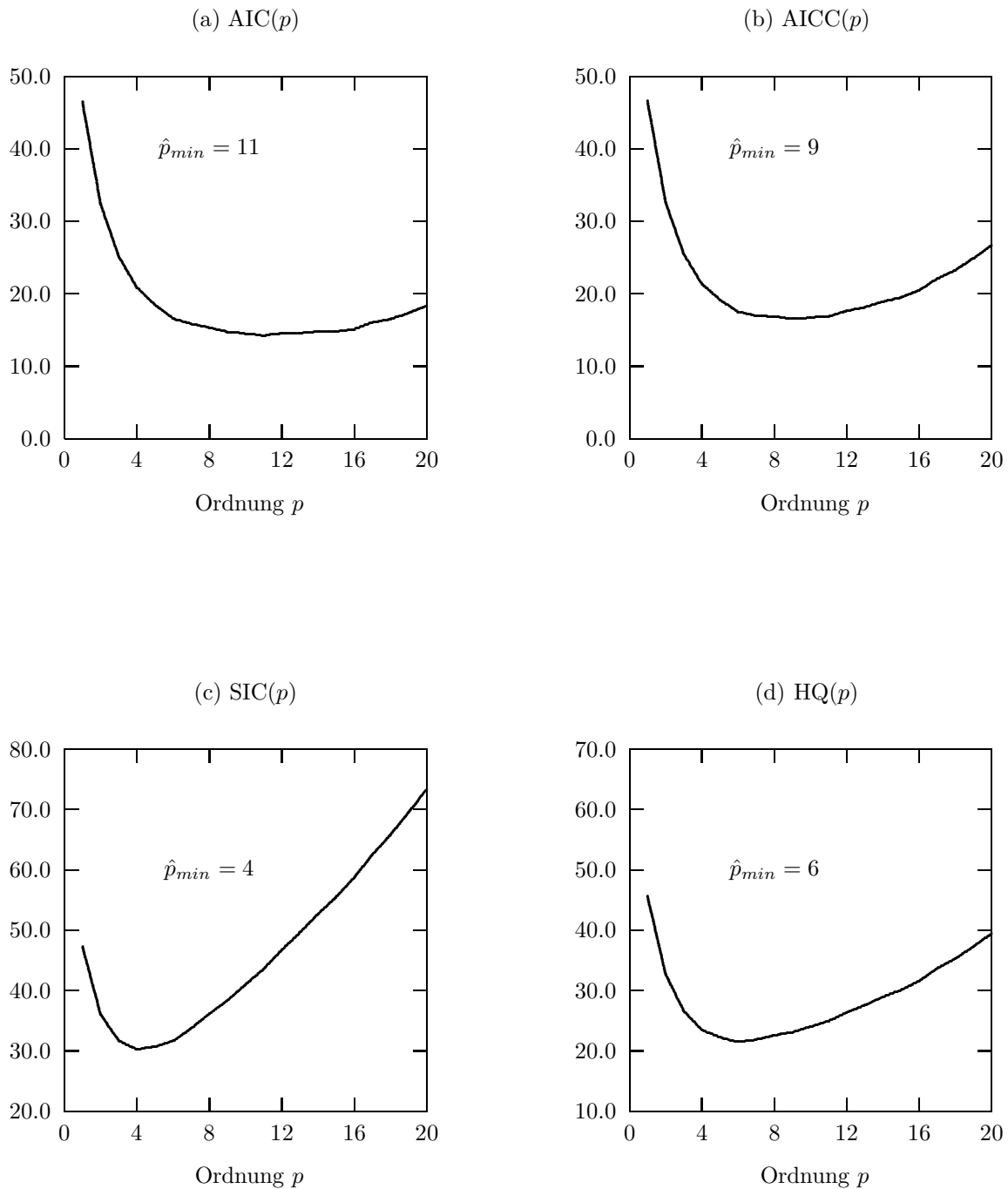


Abb. 6.26: Vergleich von Modellselektionskriterien, FLP-Methode, Schätzungen für $E[AIC(p)]$, $E[AICC(p)]$, $E[SIC(p)]$, $E[HQ(p)]$ gemittelt über 100 Realisationen des MA(1)-(MA-Smooth)-Prozesses $X_t = \epsilon_t + 0.95\epsilon_{t-1}$ mit $N = 128$, $N_E = 128$, $N_\omega = 65$. Minimum der Referenzdiskrepanz: $\hat{p}_{min}^* = 7$.

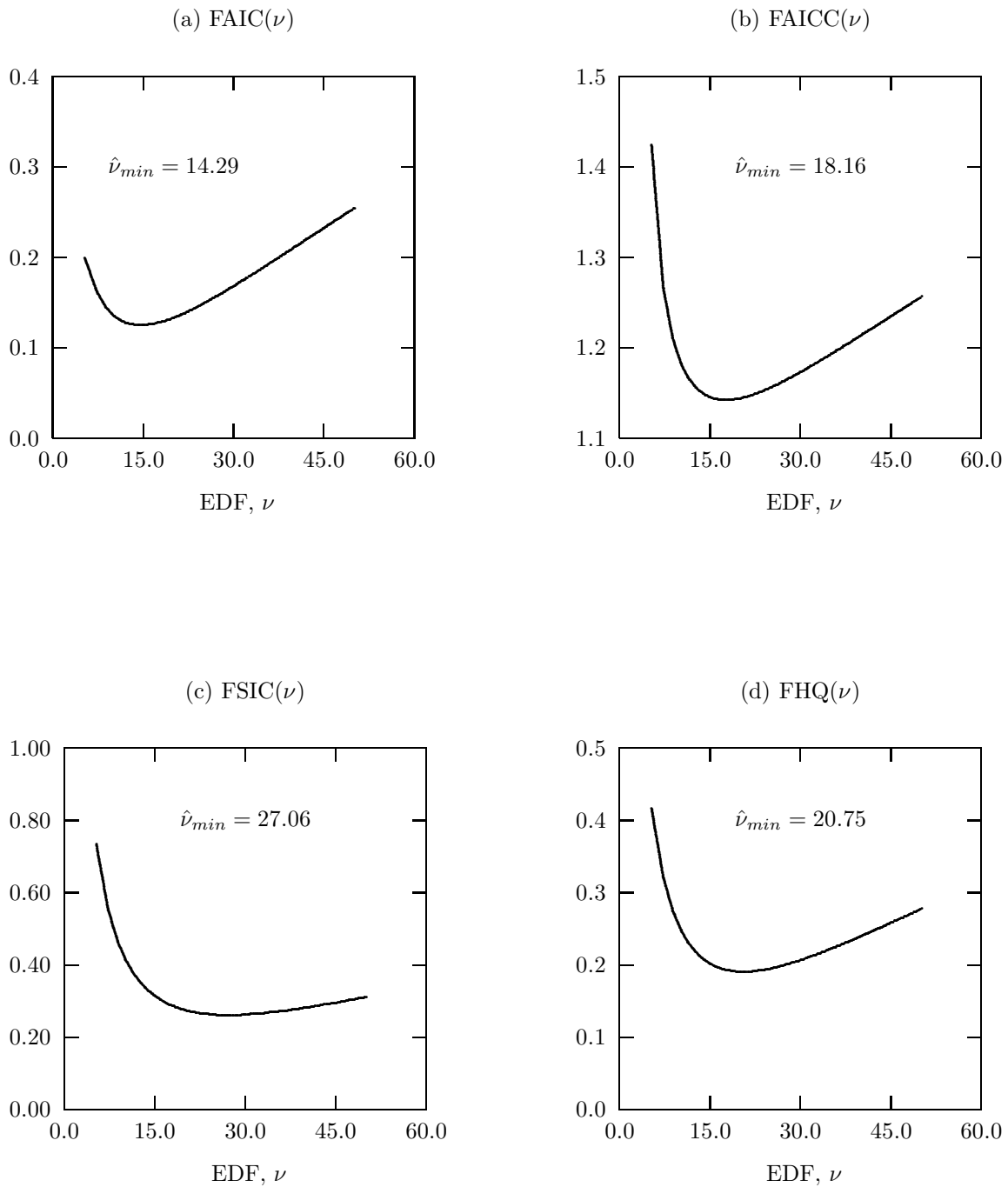


Abb. 6.27: Vergleich von Modellselektionskriterien, iteratives Hamming-Verfahren, Schätzungen für $E[\text{FAIC}(\nu)]$, $E[\text{FAICC}(\nu)]$, $E[\text{FSIC}(\nu)]$, $E[\text{FHQ}(\nu)]$ gemittelt über 100 Realisationen des MA(1)-Prozesses $X_t = \epsilon_t - 0.95\epsilon_{t-1}$ mit $N = 128$, $N_E = 128$, $N_\omega = 65$. Minimum der Referenzdiskrepanz: $\hat{\nu}_{min}^* = 22.49$.

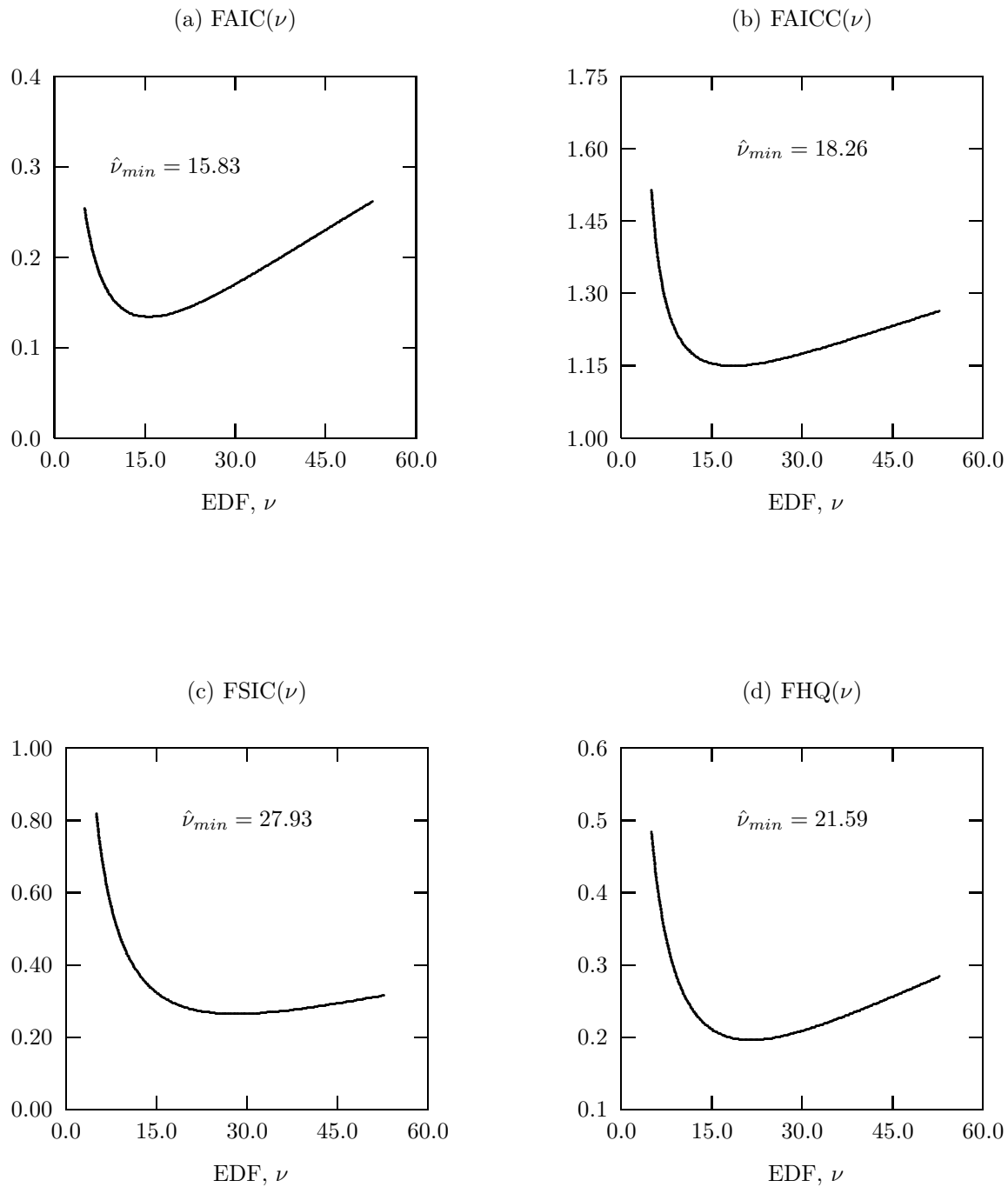


Abb. 6.28: Vergleich von Modellselektionskriterien, Parzen-Schätzung, Schätzungen für $E[\text{FAIC}(\nu)]$, $E[\text{FAICC}(\nu)]$, $E[\text{FSIC}(\nu)]$, $E[\text{FHQ}(\nu)]$ gemittelt über 100 Realisationen des MA(1)-Prozesses $X_t = \epsilon_t - 0.95\epsilon_{t-1}$ mit $N = 128$, $N_E = 128$, $N_\omega = 65$. Minimum der Referenzdiskrepanz: $\hat{\nu}_{min}^* = 21.59$.

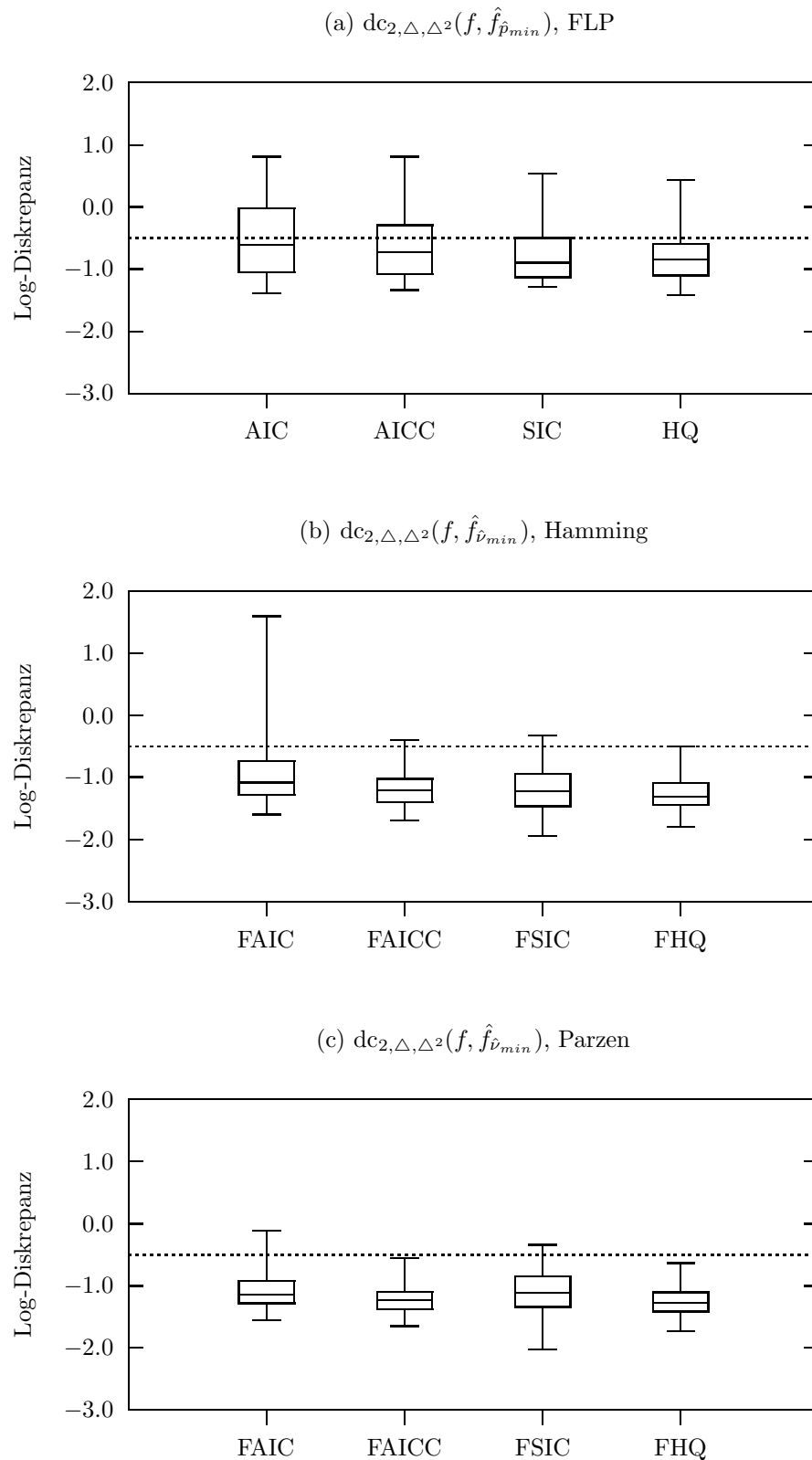


Abb. 6.29: Vergleich von parametrischen und nichtparametrischen Modellselektionskriterien, Boxplots (Minimum, 1. Quartil, Median, 3. Quartil, Maximum) basierend auf 100 Realisationen des MA(1)-Prozesses $X_t = \epsilon_t + 0.95\epsilon_{t-1}$ mit $N = 128$, $N_E = 128$, $N_\omega = 65$. Die Diskrepanz wurde für alle Kriterien jeweils mit der entsprechenden optimalen Spektralschätzung für jede Realisation berechnet.

6.4 Anwendungsbeispiele

In diesem Abschnitt erfolgt eine Evaluierung der parametrischen bzw. nichtparametrischen Modellselektionskriterien AIC, AICC, SIC, SICC bzw. FAIC, FAICC, FHQ, FSIC anhand jeweils einer Zeitreihe aus Ökonomie, Endokrinologie und Astronomie. Eine Beurteilung der automatischen Spektralschätzungen erfolgt weitgehend nach pragmatisch-methodischen Gesichtspunkten, eine substanzwissenschaftliche Interpretation der Ergebnisse wird nicht angestrebt.

6.4.1 Ökonomie: Festgeldzinsen

Die Reihe für den dreimonatigen nominalen Festgeldzinsatz der BRD umfaßt $N = 88$ Quartalsdaten von 1970:I bis 1991:IV. Die Originalreihe ist in Abb. 6.30 (a) wiedergegeben. Die Reihe enthält im vierten Quartal 1973 und im dritten Quartal 1981 zwei ausgeprägte Peaks. Die empirische Autokorrelationsfunktion (vgl. Abb. 6.30 (b)) weist eine lange gedämpfte Schwingung auf.

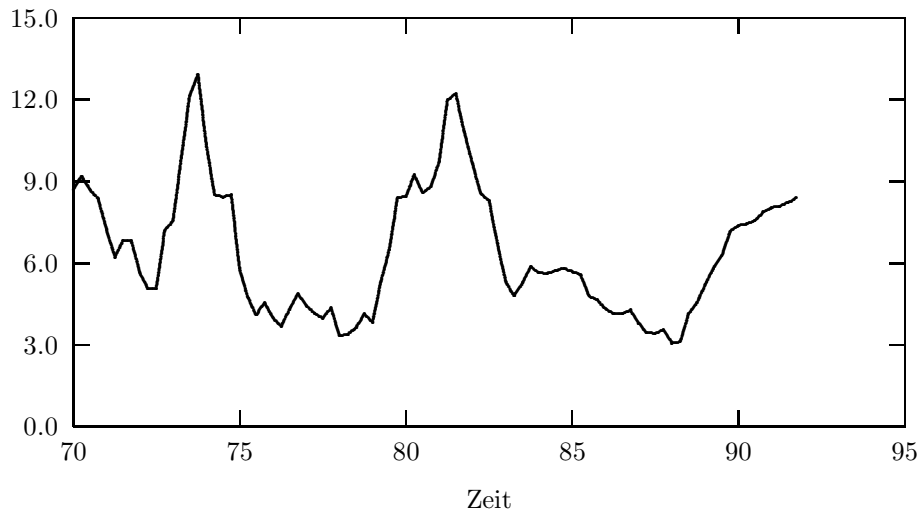
Die parametrischen Modellselektionskriterien AIC, AICC, SIC ermitteln bei Vorgabe von $p_{max} = 30$ übereinstimmend eine optimale Modellordnung von $\hat{p}_{min} = 24$, das ad hoc Kriterium SICC ermittelt eine erheblich niedrigere optimale Modellordnung von $\hat{p}_{min} = 2$ (vgl. Abb. 6.31 (a) - (d)). Entsprechend unterschiedlich sehen die zugehörigen autoregressiven Spektralschätzungen aus (vgl. Abb. 6.32 (a) und (b)): Die autoregressive Spektralschätzung mit $p = 24$ enthält zahlreiche Peaks, die Schätzung mit $p = 2$ verläuft glatt und monoton fallend. Beide Schätzungen erscheinen im Vergleich zur Autokorrelationsfunktion unplausibel: Mit der Modellordnung $p = 24$ ergibt sich ein sehr rauhes Spektrum mit mehreren Peaks im hochfrequenten Bereich von $[0.3 \leq \omega \leq 0.5]$. Bei der Modellordnung $p = 2$ wird ein glattes Spektrum ohne lokales Maximum ausgewiesen, obwohl anhand der Autokorrelationsfunktion ein lokales Maximum im Spektrum zu erwarten ist. Ein Blick auf die Kurvenverläufe der parametrischen Modellselektionskriterien zeigt, daß bei allen Kriterien zahlreiche lokale Minima auftreten: Damit hängt die ermittelte Modellordnung im wesentlichen von der Vorgabe der maximalen Modellordnung p_{max} ab. Wird die maximale Modellordnung weiter heraufgesetzt, z.B. auf $p_{max} = 50$ dann tritt das bereits in Abschnitt 6.1 besprochene Phänomen auf und das globale Minimum stimmt immer mit der maximalen Modellordnung überein.

Für die nichtparametrischen Modellselektionskriterien FAIC, FAICC, FSIC und FHQ

wurden indirekte Spektralschätzungen mit dem Parzenfenster für die Stützpunkte $T = 1, \dots, 75$ durchgeführt, dies entspricht für die Zahl äquivalenter Freiheitsgrade einem Bereich von $[326.48 \geq \nu \geq 4.35]$. In den Abbildungen 6.33 (a) - (d) ist nur der Bereich $[65.30 \geq \nu \geq 4.35]$ wiedergegeben, dies entspricht Stützpunkten von $T = 5, \dots, 75$. Das FAIC ermittelt den größten optimalen Stützpunkt bzw. die kleinste Zahl von äquivalenten Freiheitsgraden ($\hat{T}_{min} = 54$ bzw. $\hat{\nu}_{min} = 6.05$), das FSIC ermittelt den kleinsten Stützpunkt bzw. die größte Zahl von äquivalenten Freiheitsgraden ($\hat{T}_{min} = 29$ bzw. $\hat{\nu}_{min} = 11.26$). Der gravierendste Unterschied im Vergleich zu den parametrischen Modellselektionskriterien ergibt sich bezüglich der lokalen Minima: Lokale Minima treten bei den nichtparametrischen Modellselektionskriterien nicht auf, die jeweiligen optimalen Minima sind ausschließlich eindeutige globale Minima. Die nichtparametrischen Modellselektionskriterien können für die maximal zulässigen Bereiche der Stützpunkte T bzw. der äquivalenten Zahl von Freiheitsgraden ν berechnet werden, ohne daß störende Anomalien auftreten. Bei den gemäß FAICC, FSIC und FHQ als optimal ermittelten Spektralschätzungen ergibt sich trotz unterschiedlicher optimaler Stützpunkte im großen und ganzen das gleiche Bild (vgl. Abb. 6.39 (b) - (d)). Das optimale Spektrum des FAIC weist im Vergleich zu den anderen optimalen nichtparametrischen Spektralschätzungen mehrere zusätzliche Peaks auf (vgl. Abb. 6.39 (a)). Alle optimalen nichtparametrischen Spektren weisen an der gleichen Frequenz $\omega = 0.027$ den dominierenden Hauptpeak aus, was einer Periode von etwas mehr als 9 Jahren entspricht. Untergeordnete lokale Nebenmaxima sind weitgehend übereinstimmend an den Frequenzen $\omega \approx 0.155$ und $\omega \approx 0.250$ wahrzunehmen. Ein lokales Maximum an der letzten Frequenz kann bei Quartalsdaten als saisonaler Effekt interpretiert werden.

Für die Reihe Festgeldzinsen läßt sich feststellen, daß die nichtparametrischen Modellselektionskriterien, insbesondere FAICC, FSIC und FHQ, plausiblere Ergebnisse als die autoregressiven Spektralschätzungen nach AIC, AICC, SIC bzw. SICC liefern. Für die Kriterien AIC und AICC war bereits bei den Simulationen aufgefallen, daß die Resultate erhebliche Schwankungen aufweisen können. Überraschend sind bei der Reihe Festgeldzinsen insbesondere die vielen lokalen Minima bei dem Kriterium SIC, das bei den Simulationen am besten abgeschnitten hatte.

(a) Nominaler Festgeldzinssatz, 3 Monate



(b) Autokorrelationsfunktion

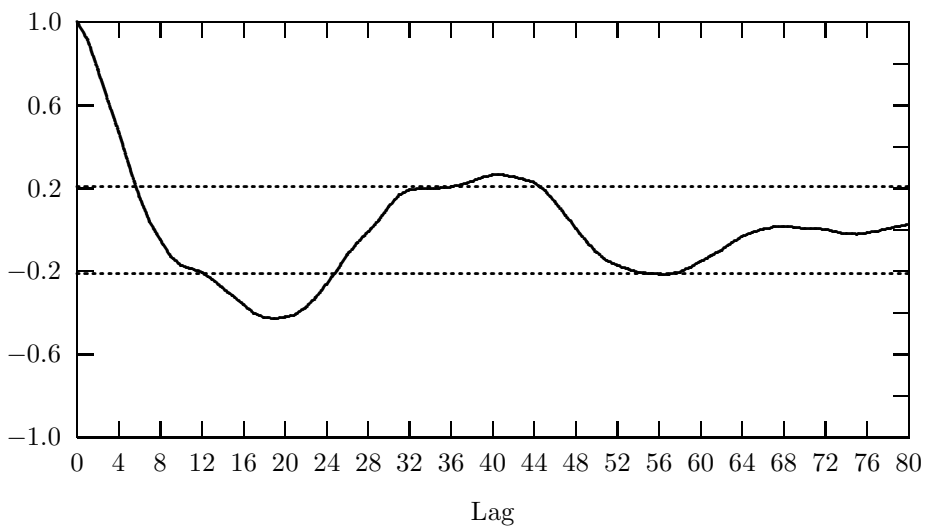


Abb. 6.30: (a) Nominaler Zinssatz für Festgelder, 3 Monate unter 1 Mio., Quartalsdaten, 1970:I bis 1991:IV, $N = 88$. (b) Empirische Autokorrelationsfunktion des Festgeldzinssatzes mit approximativen 95%-Konfidenzintervallen $\pm 1.96/\sqrt{N}$.

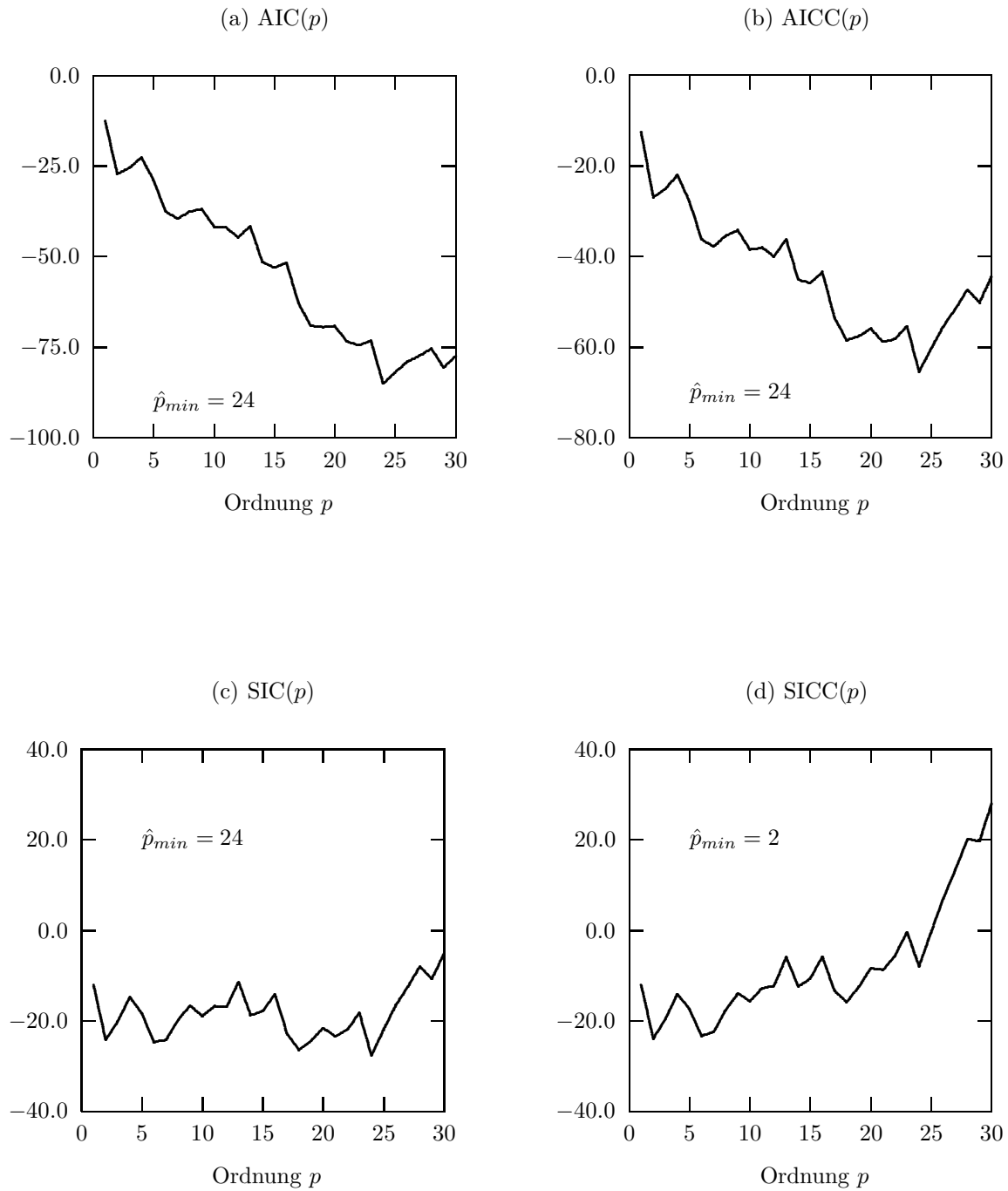


Abb. 6.31: Vergleich von Modellselektionskriterien anhand von FLP-Schätzungen für die Reihe Festgeldzinssatz.

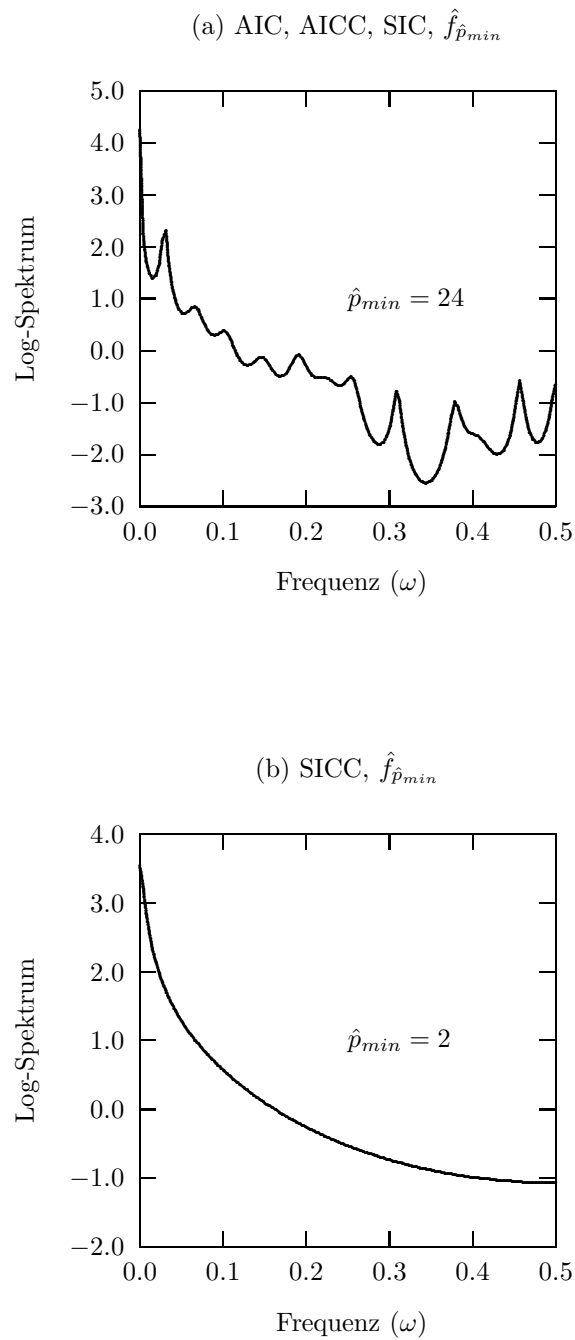


Abb. 6.32: Vergleich von Modellselektionskriterien, autoregressive Spektralschätzungen für die Reihe Festgeldzinssatz.

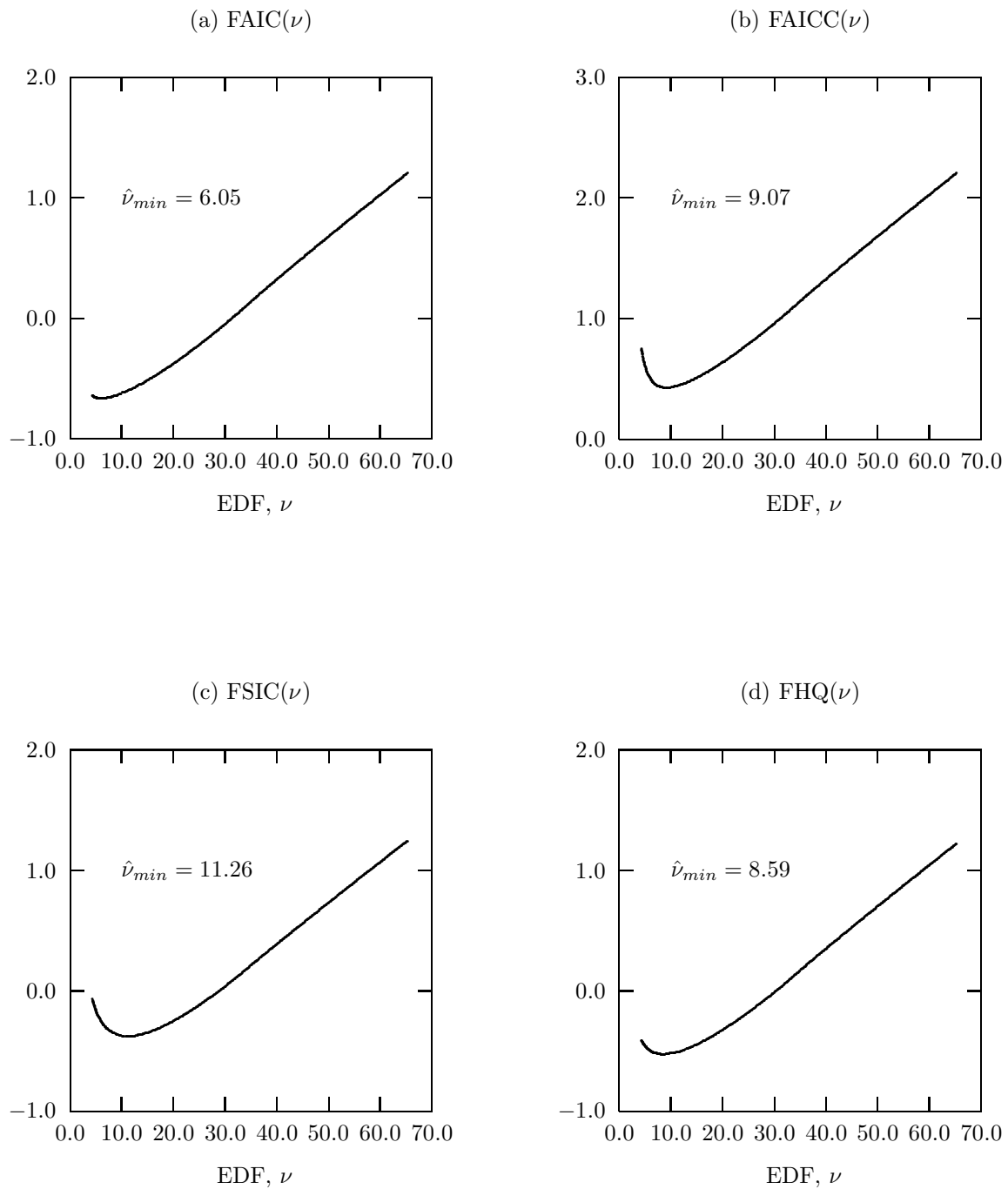


Abb. 6.33: Vergleich von Modellselektionskriterien anhand von indirekten Spektralschätzungen mit dem Parzen-Fenster.

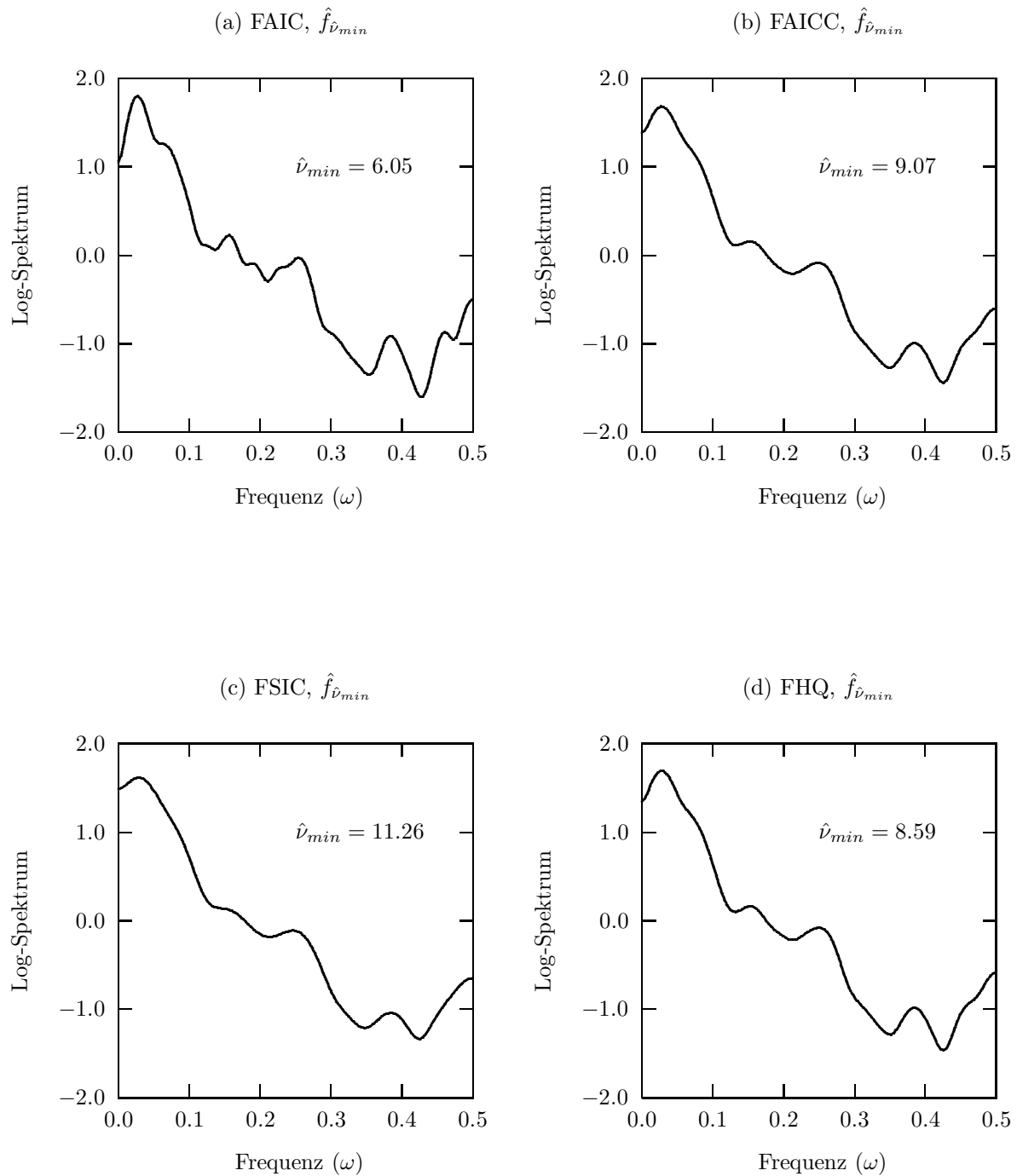


Abb. 6.34: Vergleich von Modellselektionskriterien, indirekte Spektralschätzungen mit dem Parzen-Fenster für die Reihe Festgeldzinssatz.

6.4.2 Endokrinologie: Luteinisierendes Hormon (LH)

Die Reihe luteinisierendes Hormon (LH) besteht aus 144 Messungen die in 10 minütigen Abständen über 24 Stunden hinweg an einer Kuh vorgenommen wurden (vgl. NEWTON[1988, S. 591ff, Kuh 3]).

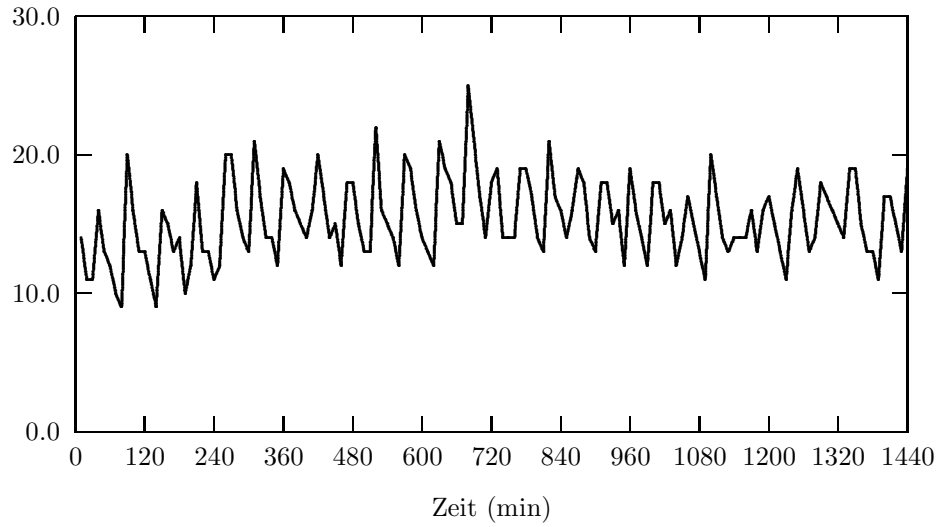
Die Originalreihe ist in Abb. 6.35 (a) wiedergegeben. Die empirische Autokorrelationsfunktion besitzt ein nicht einfach interpretierbares gedämpftes Schwingungsmuster. Für die parametrischen Modellselektionskriterien wurde als maximale Modellordnung $p_{max} = 50$ vorgegeben. Das AIC ermittelt eine optimale Modellordnung von $\hat{p}_{min} = 48$, das AICC liegt mit $\hat{p}_{min} = 19$ bereits erheblich darunter, SIC als auch SICC ermitteln als optimale Modellordnung $\hat{p}_{min} = 5$. Bei den Kriterien AIC und AICC treten die bereits bei der Reihe Festgeldzins beschriebenen Probleme sogar in verstärkter Form auf (vgl. Abb. 6.36 (a) und (b)): Es existieren für weite Bereiche der Modellordnung p zahlreiche lokale Minima, die sich numerisch kaum unterscheiden. Beim AIC kann die Wahl der Modellordnung nur als Anomalie aufgrund einer zu großen maximalen Modellordnung p_{max} interpretiert werden, beim AICC entspricht die Wahl der Modellordnung eher einer Zufallsauswahl aus einem Bereich nahezu äquivalenter lokaler Minima. Dagegen entspricht der Kurvenverlauf von SIC und SICC sehr viel besser der zu erwartenden idealtypischen Form, das globale Minimum jeweils bei $\hat{p}_{min} = 5$ ist deutlich ausgeprägt (vgl. Abb. 6.36 (c) und (d)).

Das Spektrum, das mit der optimalen Modellordnung des AIC von $\hat{p}_{min} = 48$ berechnet wurde, ist aufgrund der zahlreichen unplausiblen Peaks für eine Interpretation ungeeignet (vgl. Abb. 6.37 (a)). Das optimale Spektrum des AICC besitzt ein breites lokales Maximum im Bereich der Frequenz $\omega = 0.2$, dies entspricht einer Periode von ca. 50 Minuten. Im Bereich der Frequenz $\omega = 0.4$ treten zwei weitere lokale Nebenmaxima auf (vgl. Abb. 6.37 (b)). Das optimale Spektrum von SIC bzw. SICC verläuft sehr glatt und besitzt zwei klare lokale Maxima bei $\omega = 0.2$ und $\omega = 0.4$ (vgl. Abb. 6.37 (c)). Alle optimalen parametrischen Spektren weisen zudem ein globales Randmaximum bei der Frequenz $\omega = 0.0$ auf. Dies kann ein Hinweis darauf sein, daß bei der Originalreihe eine Trendbereinigung oder Differenzenbildung erforderlich sein kann.

Die nichtparametrischen Modellselektionskriterien wurden für die Stützpunkte $T = 1, \dots, 120$ bzw. einer Zahl äquivalenter Freiheitsgrade im Bereich $[534.24 \geq \nu \geq 4.45]$ berechnet. In den Abbildungen ist der Bereich $T = 6, \dots, 120$ bzw. $[89.04 \geq \nu \geq 4.45]$ wiedergegeben. Bei den Kurvenverläufen der nichtparametrischen

Modellselektionskriterien fällt wiederum auf, daß bei allen ein eindeutiges globales Minimum ohne weitere lokale Minima vorliegt. Das FAICC und das FHQ ermitteln das gleiche optimale $\hat{\nu}_{min} = 9.89$, was einem Stützpunkt von $T = 54$ entspricht. Das FAIC ermittelt die kleinste optimale Zahl von äquivalenten Freiheitsgraden $\hat{\nu}_{min} = 7.03$ bzw. den größten Stützpunkt $T = 76$. Das FSIC ermittelt den kleinsten Stützpunkt, $T = 38$ und die größte Zahl von äquivalenten Freiheitsgraden, $\hat{\nu}_{min} = 14.06$. Die identischen optimalen Spektralschätzungen von FAICC und FHQ stimmen weitgehend mit der optimalen Spektralschätzung des FSIC überein, neben dem Hauptpeak bei $\omega = 0.2$ weisen alle drei Spektren im Bereich der Frequenz $\omega = 0.4$ zwei lokale Nebenmaxima auf (vgl. 6.39 (b) - (c)). Bei der optimalen Spektralschätzung des FAIC wird das Hauptmaximum bei $\omega = 0.2$ bereits in zwei Hauptmaxima aufgespalten (vgl. 6.39 (a)). Das Beispiel der Reihe luteinisierendes Hormon zeigt, daß die Kriterien AIC und AICC aufgrund ihres problematischen, erheblich von der idealtypischen Form abweichenden Kurvenverlaufes als Modellselektionskriterien ungeeignet sind. Von den nichtparametrischen Kriterien erscheint das FAIC am wenigsten geeignet. Alle anderen Kriterien können grundsätzlich als Modellselektionskriterien in Betracht gezogen werden, eine genauere Auswahl kann nur im Rahmen einer substanzwissenschaftlichen Fragestellung erfolgen.

(a) Luteinisierendes Hormon



(b) Autokorrelationsfunktion

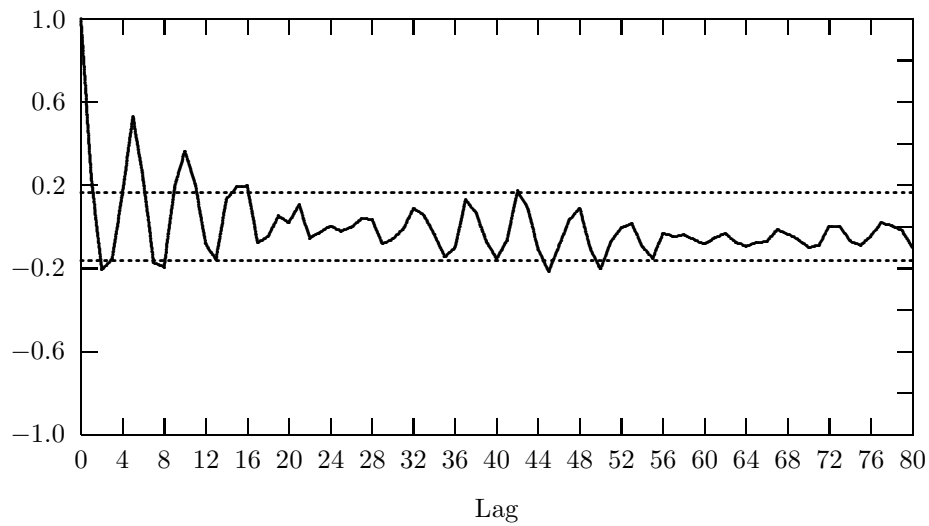


Abb. 6.35: (a) Luteinisierendes Hormon, im 10 Minuten-Abstand an einer Kuh gemessene Werte, $N = 144$, Daten aus NEWTON[1988, S. 592, Kuh 3]. (b) Empirische Autokorrelationsfunktion des luteinisierenden Hormons mit approximativen 95%-Konfidenzintervallen $\pm 1.96/\sqrt{N}$.

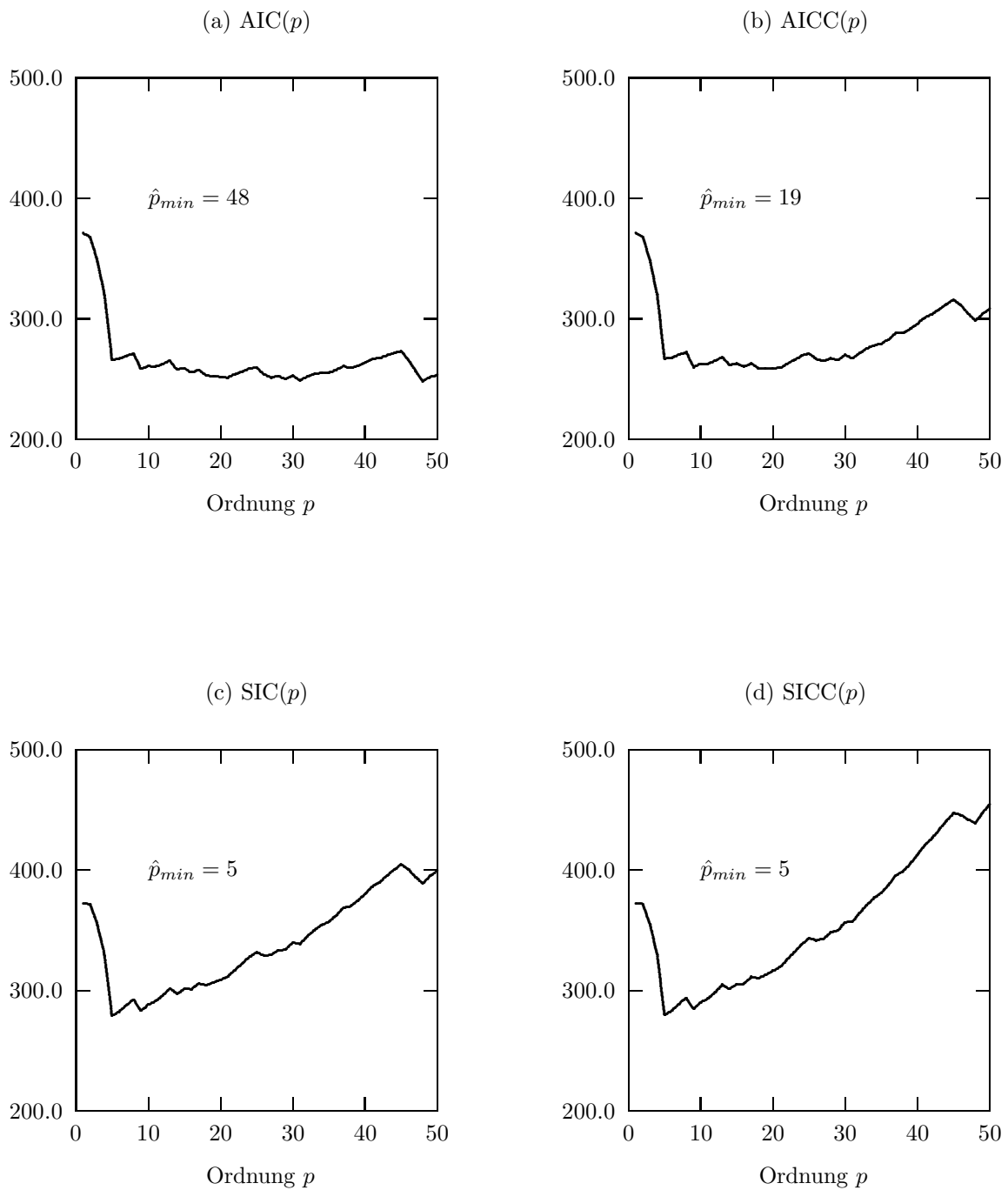


Abb. 6.36: Vergleich von Modellselektionskriterien anhand von FLP-Schätzungen für die Reihe luteinisierendes Hormon.

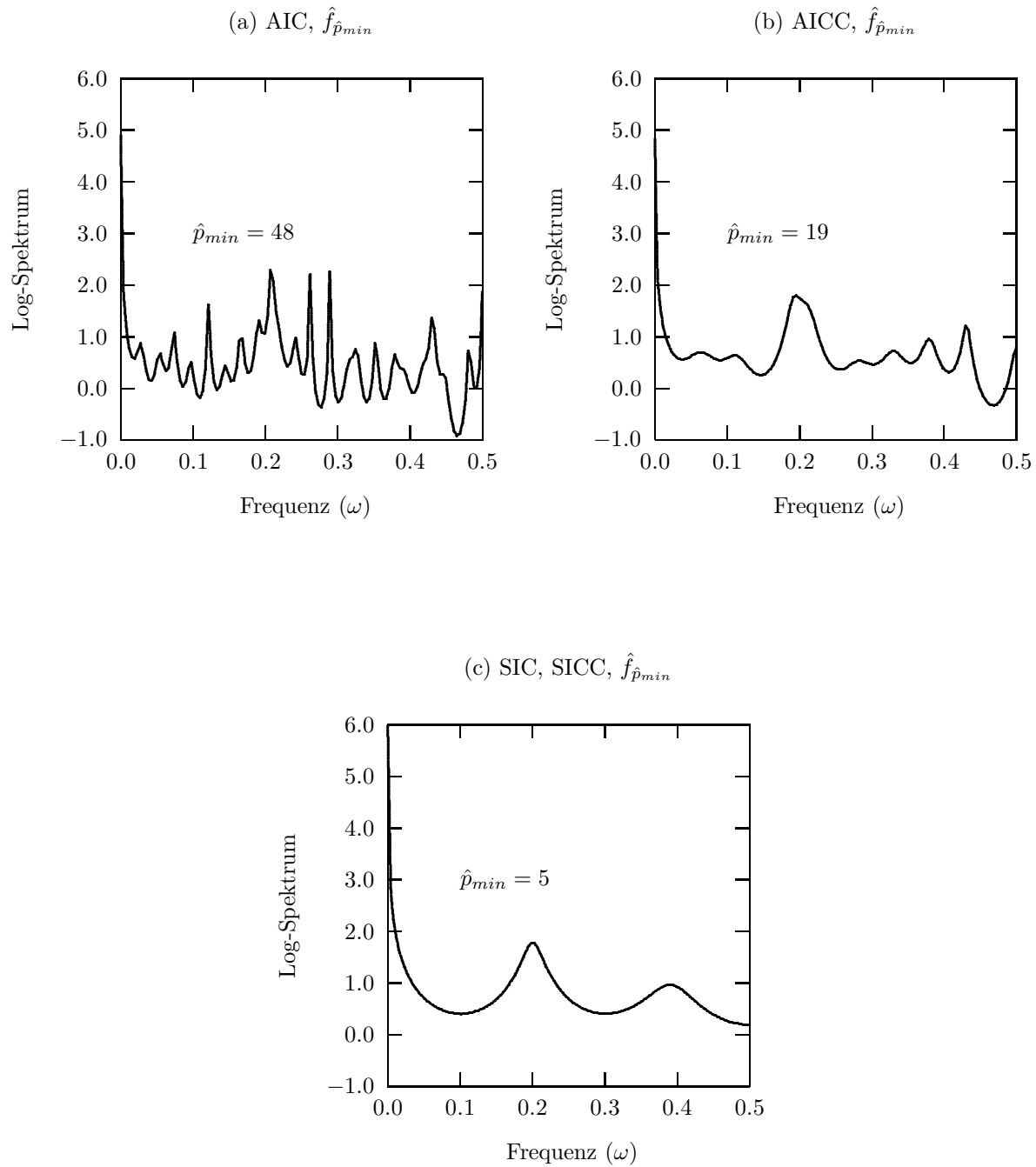


Abb. 6.37: Vergleich von Modellselektionskriterien, autoregressive Spektralschätzungen nach der FLP-Methode für die Reihe luteinisierendes Hormon.

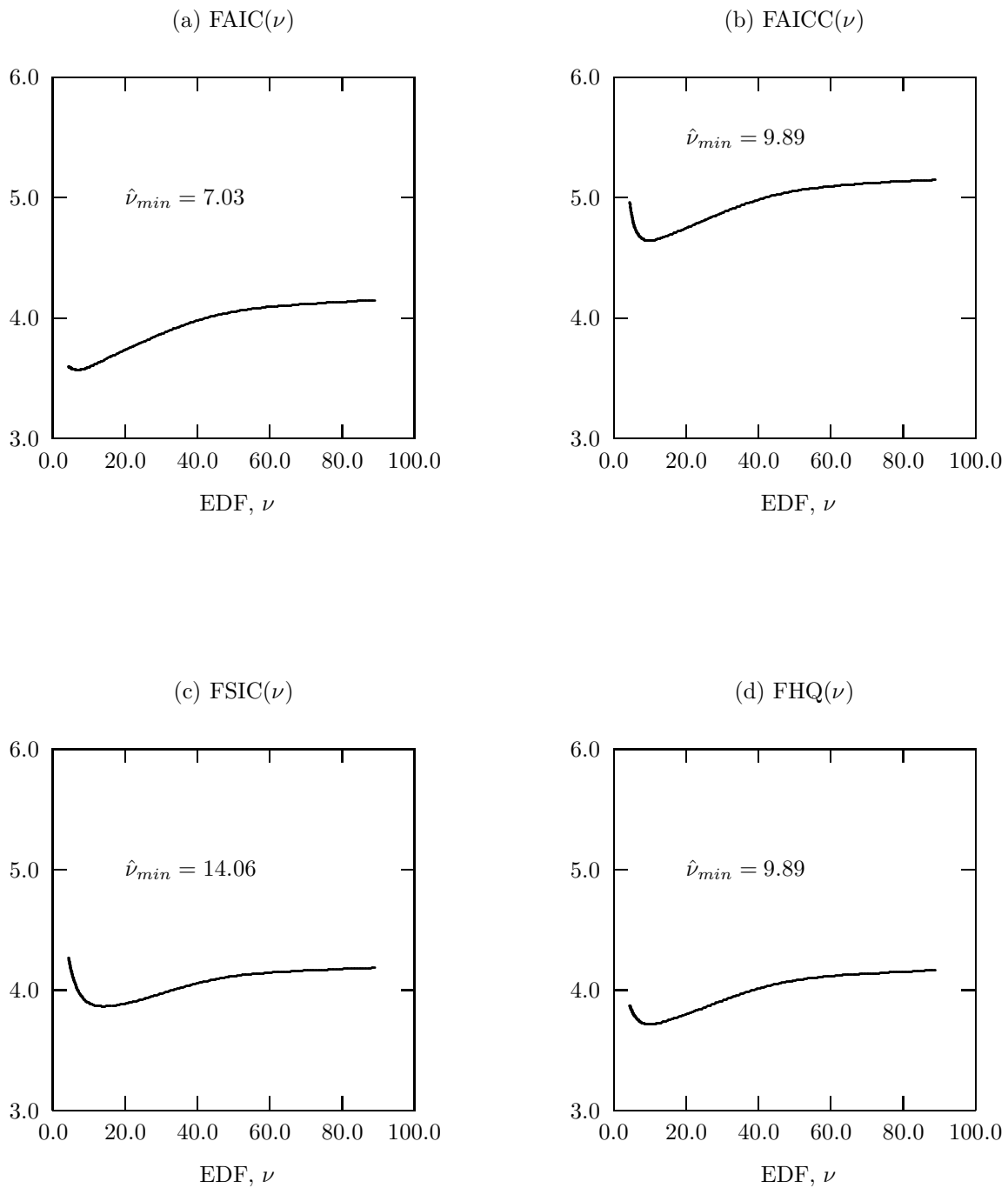


Abb. 6.38: Vergleich von Modellselektionskriterien anhand von indirekten Spektralschätzungen mit dem Parzen-Fenster für die Reihe luteinisierendes Hormon.

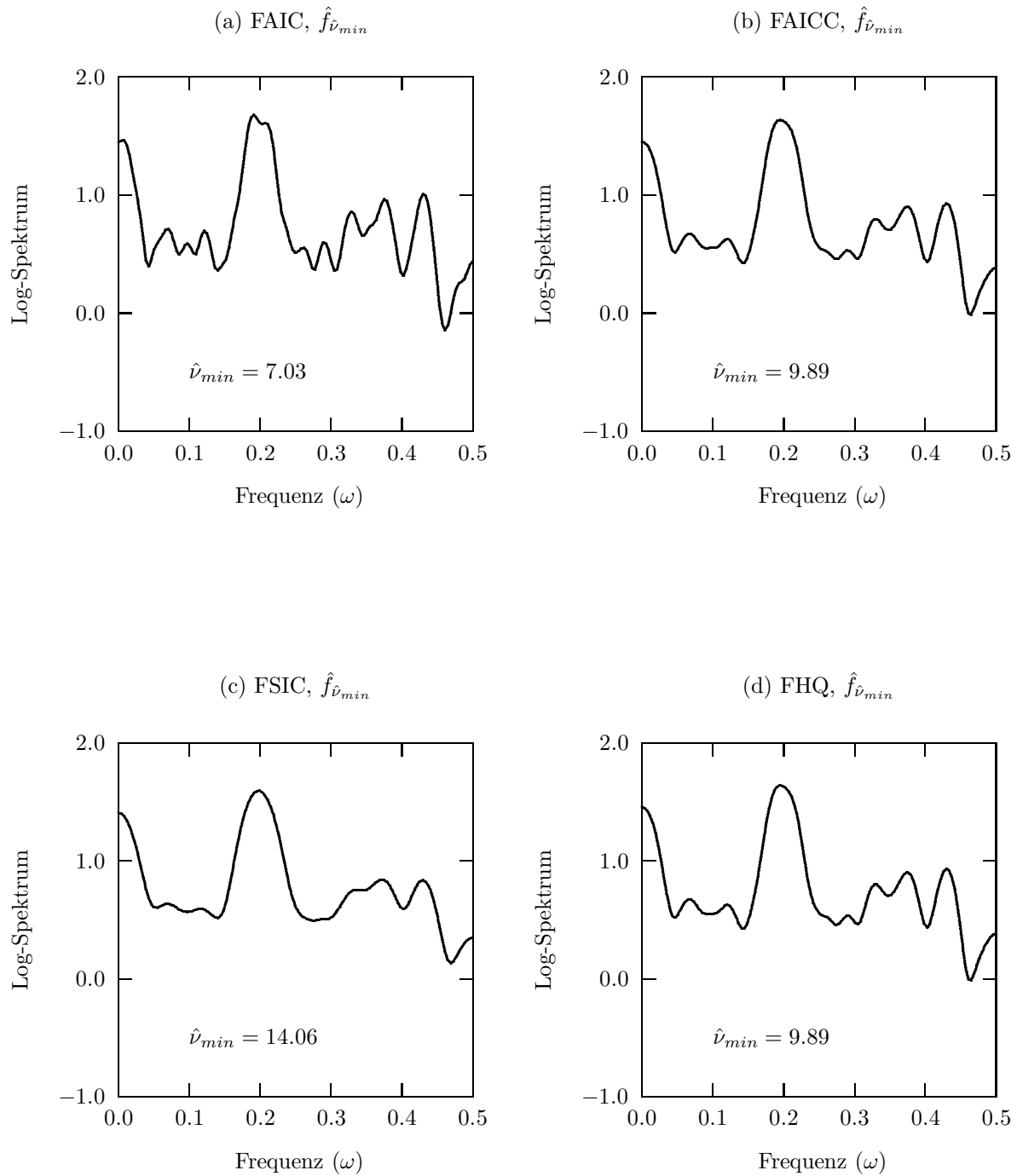


Abb. 6.39: Vergleich von Modellselektionskriterien, indirekte Spektralschätzungen mit dem Parzen-Fenster für die Reihe luteinisierendes Hormon.

6.4.3 Astronomie: Sonnenfleckindex

Seit YULE[1927] ist die Reihe des Sonnenfleckindex ein klassisches Anwendungsbeispiel für Spektralanalyse. Der hier verwendeten Reihe liegen Monatsdaten aus MARPLE[1987, S. 470ff] für den Zeitraum von 1700 - 1984 zugrunde. Eine genauere Beschreibung über Berechnung des Sonnenfleckindex ist ebenfalls bei MARPLE[1987, S. 13ff] zu finden. Da für die ersten 50 Jahre nur grobe Schätzungen der Monatswerte vorliegen, wurde die Reihe auf $N = 285$ Jahreswerte aggregiert, indem jeweils der Wert des Monats Juni ausgewählt wurde.

Die auf Jahresdaten aggregierte Reihe ist in Abb. 6.40 (a) dargestellt. Die empirische Autokorrelationsfunktion zeigt dasjenige typische Schwingungsverhalten, welches auftritt, wenn der datenerzeugende Prozeß z.B. ein AR-Prozeß mit mindestens einem Paar konjugiert-komplexer Wurzeln am Einheitskreis ist (vgl. Abb. 6.40 (b)).

Für die parametrischen Modellselektionskriterien wurde die maximale Modellordnung mit $p_{max} = 50$ vorgegeben. Im Gegensatz zur Reihe Festgeldzins und luteinisierendes Hormon weisen bei der Reihe Sonnenfleckindex alle Verlaufskurven der parametrischen Modellselektionskriterien eine äquivalente idealtypische Form auf, als optimale Modellordnung wird von allen Kriterien der Wert $\hat{p}_{min} = 10$ ermittelt (vgl. Abb. 6.41 (a) - (d)). Entsprechend sind alle optimalen Spektralschätzungen für AIC, AICC, SIC und SICC identisch (vgl. Abb. 6.42): Das Spektrum besitzt ein lokales Hauptmaximum bei $\omega \approx 0.1$ was einer Periodenlänge von ca. 10 Jahren entspricht, weitere Nebenmaxima sind bei $\omega \approx 0.2$ und $\omega \approx 0.3$. Die Nebenmaxima sind wahrscheinlich als Oberschwingungen der Grundfrequenz des Hauptmaximums von $\omega = 0.1$ zu interpretieren. Zusätzlich existiert noch ein globales Randmaximum bei der Frequenz $\omega = 0.0$.

Für die nichtparametrischen Modellselektionskriterien FAIC, FAICC, FSIC und FHQ wurden indirekte Spektralschätzungen mit dem Parzen-Fenster für die Stützpunkte $T = 1, \dots, 120$ durchgeführt, dies entspricht für die Zahl äquivalenter Freiheitsgrade einem Bereich von $[1057.35 \geq \nu \geq 8.81]$. In den Abbildungen 6.43 (a) - (d) ist nur der Bereich $[132.17 \geq \nu \geq 8.81]$ wiedergegeben, dies entspricht Stützpunkten von $T = 8, \dots, 120$. Die Kurvenverläufe weisen wiederum idealtypische Formen auf, allerdings besitzt das Kriterium FAIC ein globales Randminimum bei $\hat{\nu}_{min} = 8.81$ bzw. $\hat{T}_{min} = 120$ und erzielt damit, wie auch bei den anderen beiden Zeitreihen, die kleinste optimale äquivalente Zahl von Freiheitsgraden. Da das FAIC ohnehin eine Tendenz zur Überschätzung des Stützpunktes bzw. zur Unterschätzung der äquivalenten Zahl

von Freiheitsgraden besitzt, wurde darauf verzichtet den maximalen Stützpunkt noch weiter zu erhöhen. Den kleinsten optimalen Stützpunkt bzw. die größte optimale Zahl von äquivalenten Freiheitsgraden ermittelt das Kriterium FSIC ($\hat{T}_{min} = 59$ bzw. $\hat{\nu}_{min} = 17.92$). Die entsprechenden optimalen Spektralschätzungen (vgl. Abb. 6.44 (a) - (d)) weisen erheblich mehr lokale Maxima auf als die parametrisch ermittelte Spektralschätzung. Neben dem Hauptmaximum bei $\omega \approx 0.1$ existiert z.B. bei den optimalen Spektren nach FAIC, FAICC und FHQ ein Nebenmaximum bei der Frequenz $\omega \approx 0.008$, was ca. einer Periode von 125 Jahren entspräche. Bei den optimalen Spektren nach FAIC und FAICC erscheint weiterhin ein lokales Maximum an der Frequenz $\omega \approx 0.12$, was ca. einer Periode von etwas mehr als 8 Jahren entspräche.

Für die Reihe Sonnenfleckensindex kann festgestellt werden, daß die parametrischen Modellselektionskriterien eine identische und plausiblere Spektralschätzung ermitteln, als die nichtparametrischen Kriterien. Von den nichtparametrischen Modellselektionskriterien erscheint lediglich das Kriterium FSIC akzeptabel.

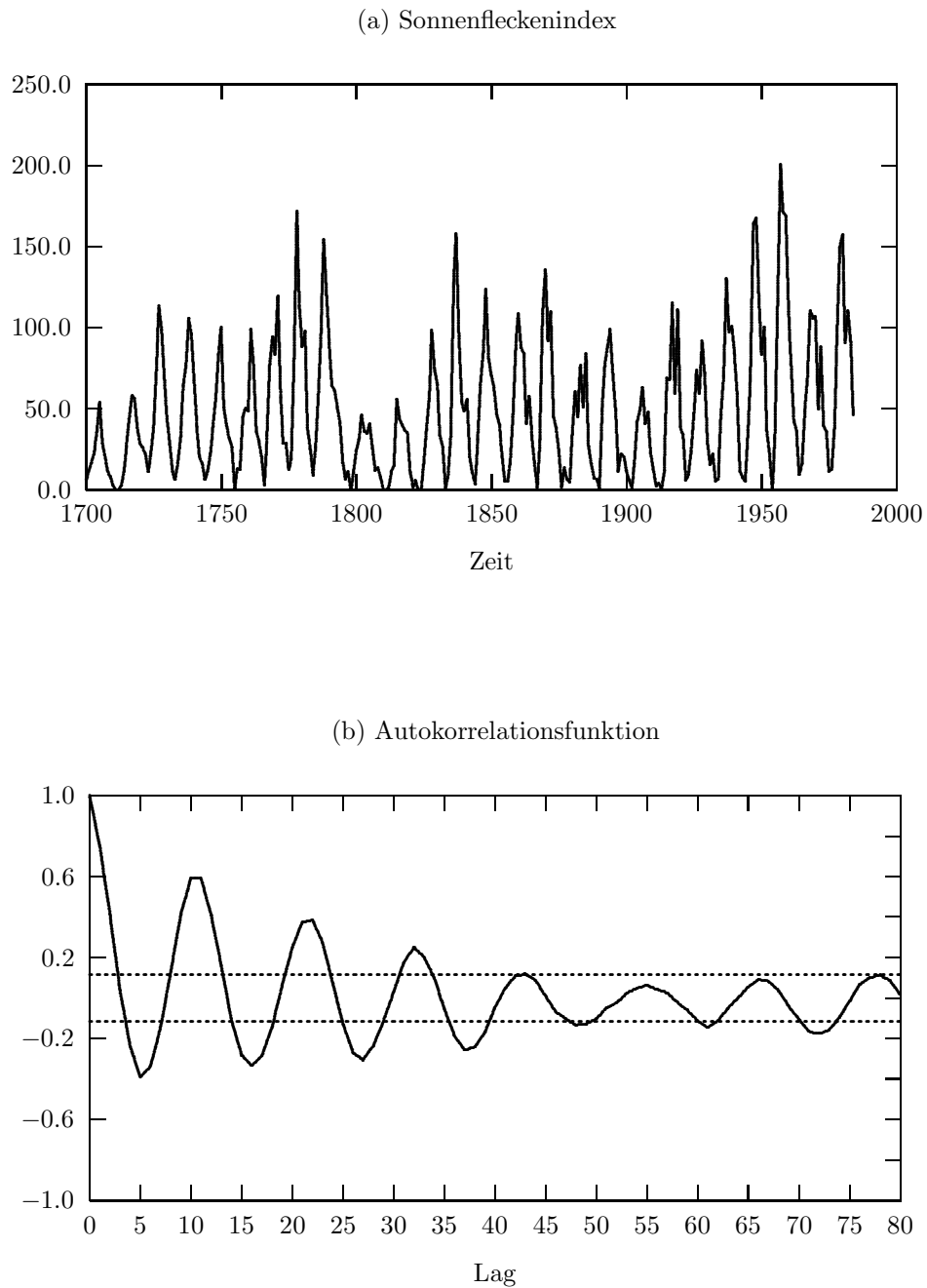


Abb. 6.40: (a) Sonnenfleckenindex, jeweils die Werte des Monats Juni von 1700 - 1984, $N = 285$, Daten aus MARPLE[1987, S. 470ff]. (b) Empirische Autokorrelationsfunktion des Sonnenfleckenindex mit approximativen 95%-Konfidenzintervallen $\pm 1.96/\sqrt{N}$.

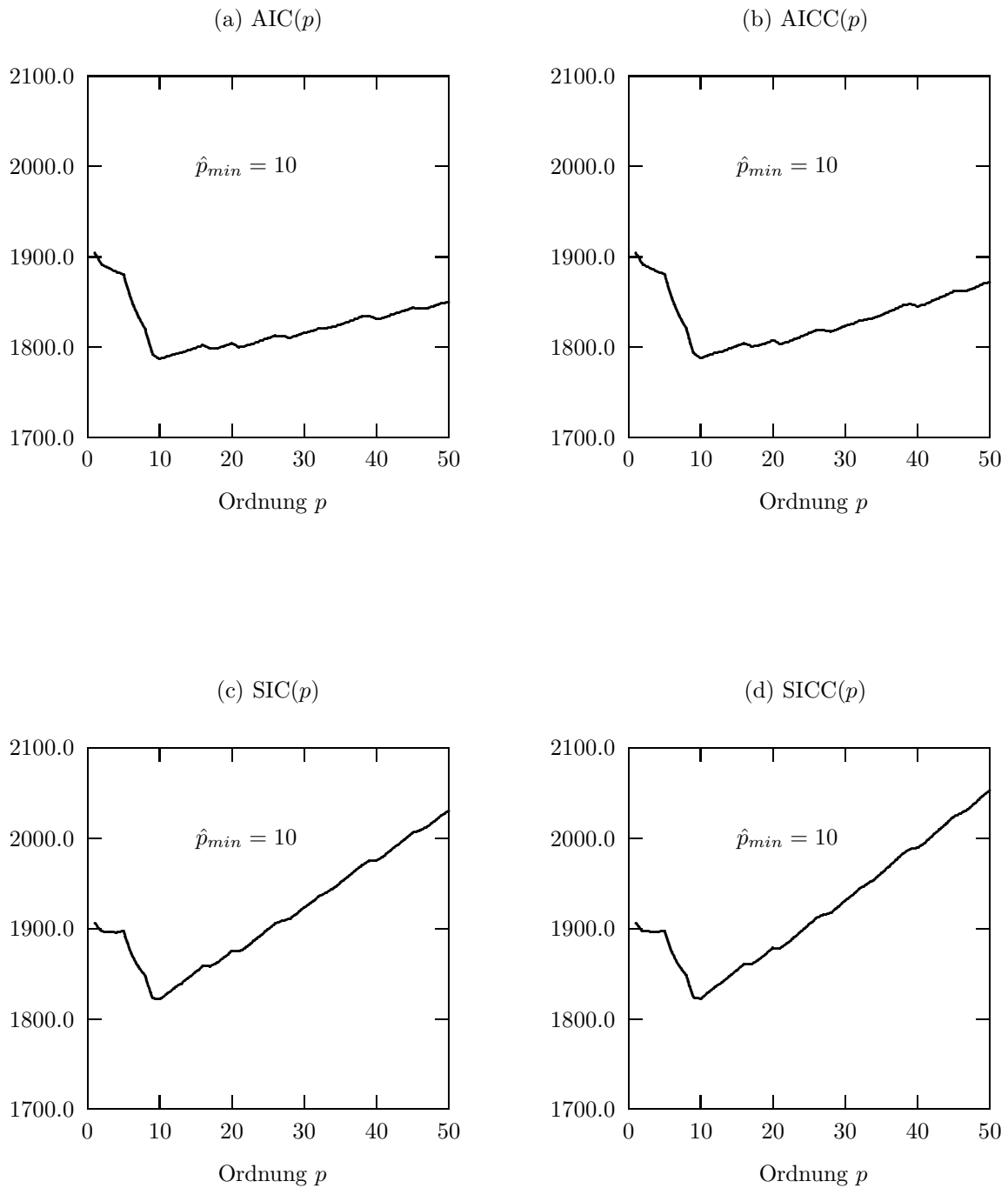


Abb. 6.41: Vergleich von Modellselektionskriterien, FLP-Schätzungen für Sonnenfleckenin-
dex.

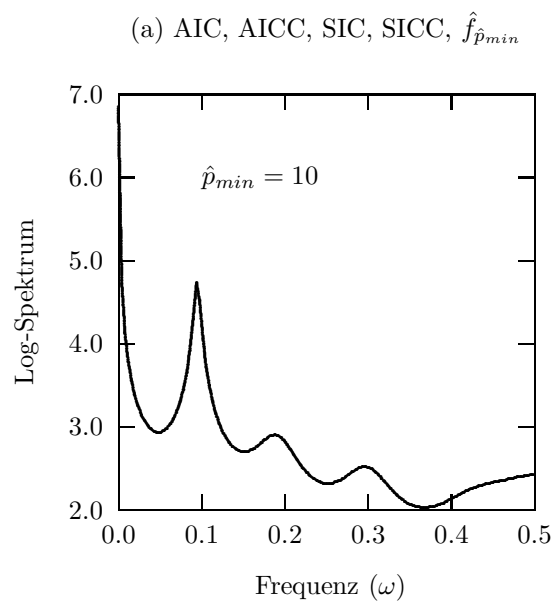


Abb. 6.42: Vergleich von Modellselektionskriterien, autoregressive Spektralschätzungen nach der FLP-Methode für die Reihe Sonnenfleckenindex.

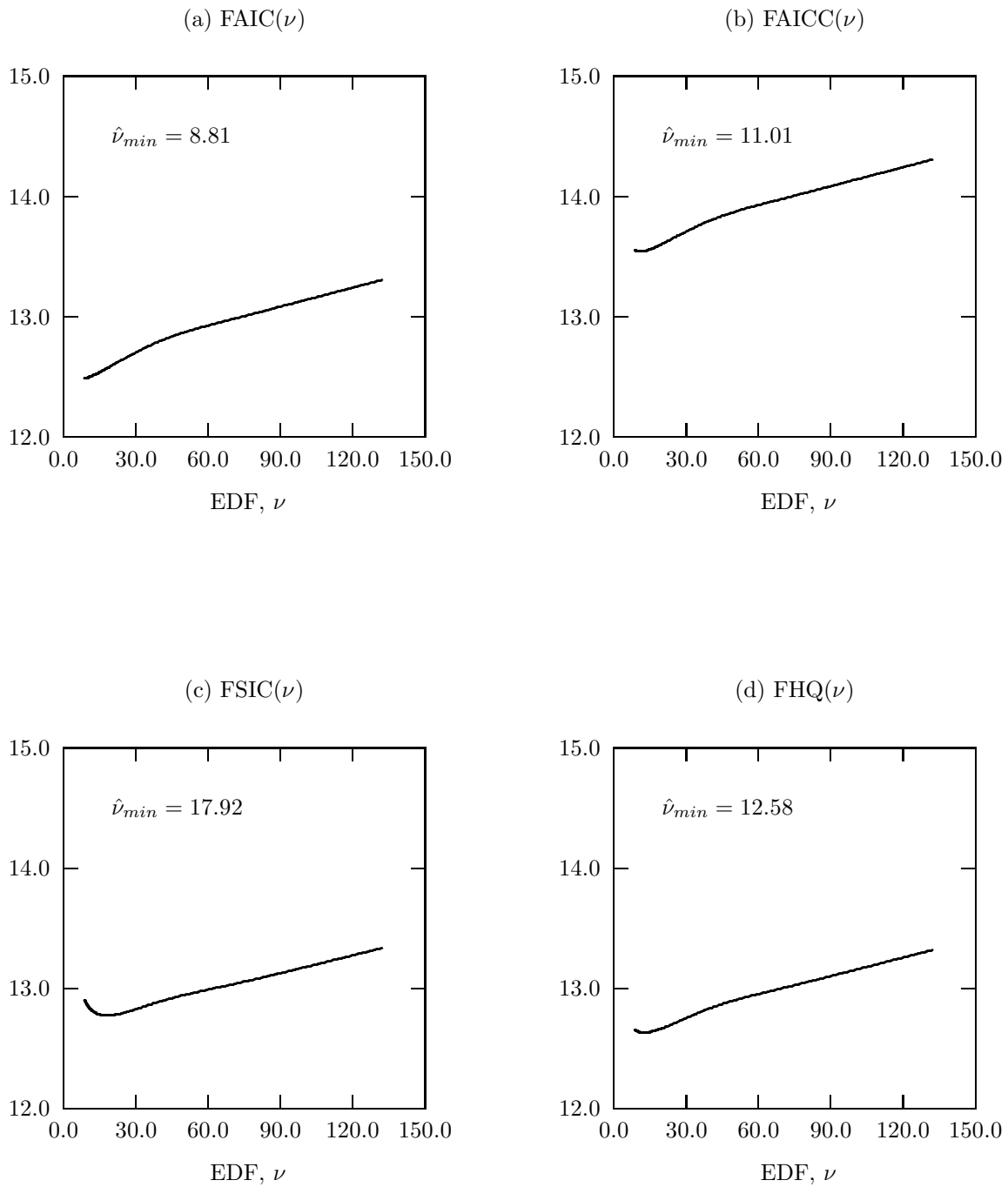


Abb. 6.43: Vergleich von Modellselektionskriterien an Hand von indirekten Spektralschätzungen mit dem Parzen-Fenster für die Reihe Sonnenfleckindex.

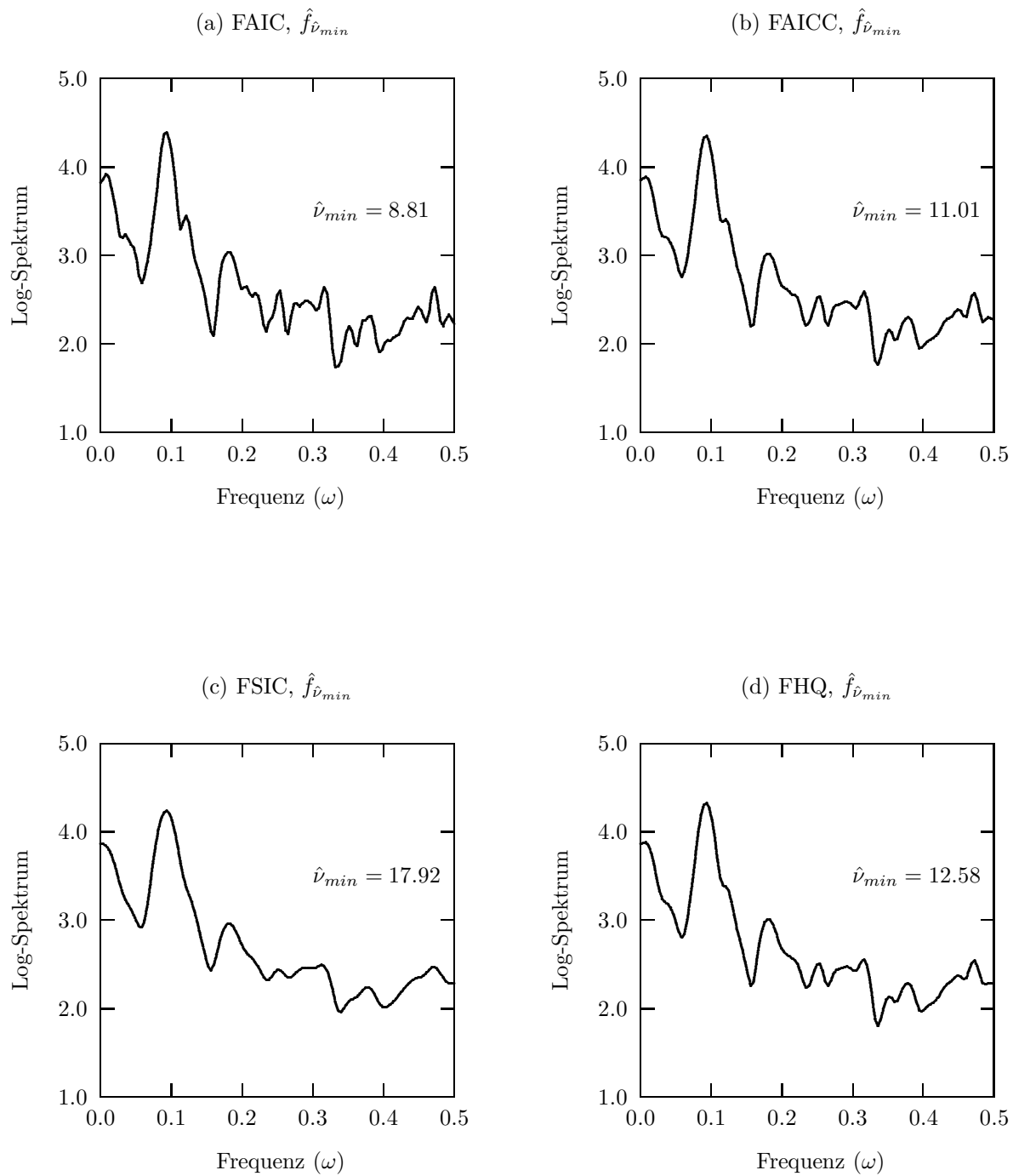


Abb. 6.44: Vergleich von Modellselektionskriterien, indirekte Spektralschätzungen mit dem Parzen-Fenster für die Reihe Sonnenfleckensindex.

Kapitel 7

Zusammenfassung

Das Problem der automatischen parametrischen und nichtparametrischen Spektralschätzung wurde in einem geschlossenen und einheitlichen Rahmen diskutiert und verschiedene methodisch-statistische Erweiterungen entwickelt.

Für autoregressive Spektralschätzungen wurden umfassend alternative Schätzverfahren sowie deren Bedeutung bei speziellen Problemstellungen diskutiert. Insbesondere bei hochauflösenden Spektralschätzungen ist der Wahl eines geeigneten autoregressiven Schätzverfahrens besondere Bedeutung zuzumessen.

Anhand von Beispielen war gezeigt worden, daß klassische Diskrepanzmaße, wie z.B. der integrierte quadratische Fehler oder das Kullback-Leibler-Kriterium im Frequenzbereich, im allgemeinen keine sinnvollen Diskrepanzmaße für den Vergleich von Spektraldichtefunktionen sind. Da sinnvolle Diskrepanzmaße aber unabdingbare Voraussetzung zur Evaluierung von automatischen Verfahren der Spektralschätzung sind, wurde eine neue Klasse von global korrigierten Diskrepanzmaßen entwickelt, mit denen sich die Mängel der klassischen Diskrepanzmaße korrigieren lassen. Zur Evaluierung von automatischen Verfahren der Spektralschätzung wurde ein spezielles global korrigiertes Diskrepanzmaß vorgeschlagen, das besonders geeignet scheint, auch die visuell wahrnehmbaren Unterschiede zwischen zwei Spektren zu beschreiben.

Geläufige parametrische Modellselektionsverfahren wurden vorgestellt und hinsichtlich verschiedener Optimalitätskonzepte kritisch diskutiert. Es mußte festgestellt werden, daß keines der Optimalitätskonzepte klare theoretische Anhaltspunkte dafür liefert, daß allgemein ein bestimmtes parametrisches Modellselektionskriterium den anderen vorgezogen werden kann. Insbesondere das Konzept der Ordnungskonsistenz bezüglich eines „wahren“ datenerzeugenden Prozesses muß als Beurteilungsmaßstab für para-

metrische Modellselektionskriterien zurückgewiesen werden. Auf spezielle Probleme von parametrischen Modellselektionskriterien bei kleinen Stichprobenumfängen wurde hingewiesen, in Anlehnung an HURVICH/TSAI[1989] wurden weitere korrigierte parametrische Modellselektionskriterien für kleine Stichprobenumfänge vorgeschlagen.

Für nichtparametrische Spektralschätzungen wurden zunächst zwei Kreuzvalidierungsverfahren im Frequenzbereich vorgestellt. Als Verallgemeinerung der Ansätze von HANNAN/RISSANEN[1988], HURVICH/BELTRÃO[1990] und HURVICH[1992] wurde eine neue Klasse von strafenden nichtparametrischen Modellselektionskriterien vorgeschlagen. Ein Vorteil dieser Klasse von Modellselektionsverfahren im Vergleich zu Kreuzvalidierungsverfahren im Frequenzbereich ist insbesondere die einfache und schnelle Berechenbarkeit. Darüberhinaus können diese nichtparametrischen Kriterien in unmittelbarer Analogie zu den üblichen parametrischen Modellselektionskriterien interpretiert werden, der Strafterm kann entsprechend leicht modifiziert und an spezielle Problemstellungen angepaßt werden.

Die Simulationsergebnisse, basierend auf vier verschiedenen datenerzeugenden Prozessen (drei AR-Prozesse, ein MA-Prozeß) und jeweils vier parametrischen und vier nichtparametrischen Modellselektionskriterien können wie folgt zusammengefaßt werden.

Bei den parametrischen Modellselektionskriterien und Anwendung der FLP-Schätzung erzielte das SIC, unabhängig von dem datenerzeugenden Prozeß, die besten Ergebnisse. Handelte es sich bei dem datenerzeugenden Prozeß nicht um einen autoregressiven Prozeß, so konnten mit den nichtparametrischen Schätzverfahren bessere Ergebnisse erzielt werden.

Beschränkt man sich auf einen Vergleich der nichtparametrischen Modellselektionskriterien, so kann kein gleichmäßig bestes Kriterium ermittelt werden. In Abhängigkeit von dem datenerzeugenden Prozeß können entweder mit dem ad hoc Kriterium FAIC, dem ad hoc Kriterium FSIC oder dem ad hoc Kriterium FHQ die besten Ergebnisse erzielt werden.

Daß die Aussagen der Simulationsstudie, insbesondere was die Überlegenheit des SIC betrifft, nicht ohne weiteres verallgemeinert werden dürfen, zeigen die Anwendungsbeispiele mit echten Zeitreihen. Bei der Reihe Festgeldzins treten z.B. auch mit dem Kriterium SIC Probleme aufgrund eines extrem atypischen Kurvenverlaufs auf, statt eines deutlich ausgeprägten globalen Minimums waren zahlreiche lokale Minima zu beobachten.

Sowohl die Simulationsergebnisse als auch die Anwendungsbeispiele belegen, daß keine allgemeine Empfehlung für ein bestimmtes automatisches Spektralschätzverfahren gegeben werden kann. Ein allgemeiner methodischer Rahmen zur Entwicklung und Evaluierung von verschiedenen automatischen Schätzverfahren wurde jedoch in dieser Arbeit aufgezeigt. Eine sinnvolle Auswahl von Modellselektionskriterien zur automatischen Spektralschätzung ist nur im Rahmen einer konkreten substanzwissenschaftlichen Fragestellung bei Verfügbarkeit von a priori Informationen möglich. Erfolgreiche praktische Anwendungen sind der letztendlich gültige Beurteilungsmaßstab.

Anhang A

Toeplitzmatrizen

Die Eigenschaften von Toeplitzmatrizen und insbesondere die rekursive Berechnung von inversen Toeplitzmatrizen mit Hilfe der Levinson-Durbin-Rekursion sind grundlegend für die numerische Implementierung von Schätzern für AR-Prozesse. Die Darstellung hier folgt MARPLE[1987, S. 80ff].

Def. A.0.1 (*Toeplitz-Matrizen*)

Sei $\{t_l, l \in \mathcal{Z}\}$ eine komplexwertige Folge. Dann heißt eine Matrix der Form

$$(A.1) \quad \mathbf{T}_N = \begin{pmatrix} t_0 & t_{-1} & \dots & t_{-N} \\ t_1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & t_{-1} \\ t_N & \dots & t_1 & t_0 \end{pmatrix}$$

mit $t_{lm} = t_{l-m}$, $l, m = 1, \dots, (N + 1)$ Toeplitzmatrix der Ordnung $(N + 1)$. Für hermitesche Toeplitzmatrizen mit $\mathbf{T}_N^H = \mathbf{T}_N$ gilt zusätzlich: $t_l = t_{-l}^*$. ■

Toeplitzmatrizen sind symmetrisch bezüglich der Gegendiagonalen (persymmetrisch), d.h. es gilt

$$\mathbf{T} = \mathbf{J}\mathbf{T}'\mathbf{J},$$

wobei \mathbf{J} eine ordnungskonforme Reflektionsmatrix ist, die nur Einsen auf der Gegendiagonalen besitzt:

$$\mathbf{J} = \begin{pmatrix} 0 & \dots & 0 & 1 \\ 0 & \dots & 1 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 1 & \dots & 0 & 0 \end{pmatrix}.$$

Die Inverse einer Toeplitzmatrix ist ebenfalls persymmetrisch und soll sich darstellen lassen als:

$$\mathbf{U}_N = \mathbf{T}_N^{-1} = \begin{pmatrix} u_{00} & \dots & u_{0N} \\ \vdots & \ddots & \vdots \\ u_{N0} & \dots & u_{NN} \end{pmatrix}.$$

Die Inverse Matrix \mathbf{U}_N kann vollständig bestimmt werden, wenn die erste Spalte von \mathbf{U}_N

$$(A.2) \quad \begin{pmatrix} t_0 & t_{-1} & \dots & t_{-N} \\ t_1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & t_{-1} \\ t_N & \dots & t_1 & t_0 \end{pmatrix} \begin{pmatrix} u_{00} \\ \vdots \\ u_{(N-1)0} \\ u_{N0} \end{pmatrix} = \begin{pmatrix} 1 \\ \vdots \\ 0 \\ 0 \end{pmatrix}$$

sowie die letzte Spalte von von \mathbf{U}_N

$$(A.3) \quad \begin{pmatrix} t_0 & t_{-1} & \dots & t_{-N} \\ t_1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & t_{-1} \\ t_N & \dots & t_1 & t_0 \end{pmatrix} \begin{pmatrix} u_{0N} \\ \vdots \\ u_{(N-1)N} \\ u_{NN} \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}$$

berechnet worden ist (vgl. z.B. MARPLE[1987, S. 81ff]).

Für die weitere Darstellung ist es einfacher, Gleichung (A.2) und Gleichung (A.3) umzuformen: Gleichung (A.2) wird auf der linken und rechten Seite durch u_{00} dividiert und Gleichung (A.3) wird auf der linken und rechten Seite durch u_{NN} dividiert.

Mit $a_{lN} = u_{l0}/u_{00}$ für $l = 1, \dots, N$ und $c_N^a = 1/u_{00}$ ergibt sich für Gleichung (A.2):

$$(A.4) \quad \begin{pmatrix} t_0 & t_{-1} & \dots & t_{-N} \\ t_1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & t_{-1} \\ t_N & \dots & t_1 & t_0 \end{pmatrix} \begin{pmatrix} 1 \\ \vdots \\ a_{(N-1)N} \\ a_{NN} \end{pmatrix} = \begin{pmatrix} c_N^a \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

und mit $b_{lN} = u_{(N-l)N}/u_{NN}$ für $l = 1, \dots, N$ und $c_N^b = 1/u_{NN}$ ergibt sich für Gleichung (A.3):

$$(A.5) \quad \begin{pmatrix} t_0 & t_{-1} & \dots & t_{-N} \\ t_1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & t_{-1} \\ t_N & \dots & t_1 & t_0 \end{pmatrix} \begin{pmatrix} b_{NN} \\ \vdots \\ b_{1N} \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ c_N^b \end{pmatrix}.$$

Aus der Persymmetrie von \mathbf{U}_N folgt $u_{00} = u_{NN}$ so daß gilt: $c_N^a = c_N^b = c_N$.

Grundlegend für die rekursive Berechnung der Koeffizienten $\mathbf{a}_n = (a_{1n}, \dots, a_{nn})'$ und $\mathbf{b}_n = (b_{1n}, \dots, b_{nn})'$ für $n = 1, \dots, N$ ist die Darstellung einer Toeplitzmatrix als geränderte Matrix:

$$(A.6) \quad \mathbf{T}_n = \begin{pmatrix} \mathbf{T}_{n-1} & \mathbf{J}\mathbf{s}_n \\ \mathbf{r}'_n\mathbf{J} & t_0 \end{pmatrix} = \begin{pmatrix} t_0 & \mathbf{s}'_n \\ \mathbf{r}_n\mathbf{J} & \mathbf{T}_{n-1} \end{pmatrix},$$

mit den Vektoren $\mathbf{r}_n = (t_1, \dots, t_n)'$ und $\mathbf{s}_n = (t_{-1}, \dots, t_{-n})'$ sowie der ordnungskonformen Reflektionsmatrix \mathbf{J} .

Durch Induktion läßt sich ableiten, daß sich die Koeffizienten des Vektors \mathbf{a}_n mit Hilfe der Vektoren \mathbf{a}_{n-1} und \mathbf{b}_{n-1} darstellen lassen als:

$$(A.7) \quad \begin{pmatrix} 1 \\ \mathbf{a}_n \end{pmatrix} = \begin{pmatrix} 1 \\ \mathbf{a}_{n-1} \\ 0 \end{pmatrix} + a_{nn} \begin{pmatrix} 0 \\ \mathbf{J}\mathbf{b}_{n-1} \\ 1 \end{pmatrix},$$

bzw. bei elementweiser Betrachtung für $l = 1, \dots, (n-1)$:

$$(A.8) \quad a_{ln} = a_{l(n-1)} + a_{nn}b_{(n-l)(n-1)}.$$

Zur Bestimmung des Skalars a_{nn} wird die Gleichung (A.7) auf beiden Seiten von links mit \mathbf{T}_n multipliziert, wobei für \mathbf{T}_n unter Verwendung von (A.6) eine geeignete Partitionierung gewählt wird. Damit ergibt sich:

$$(A.9) \quad \mathbf{T}_n \begin{pmatrix} 1 \\ \mathbf{a}_n \end{pmatrix} = \begin{pmatrix} \mathbf{T}_{n-1} & \mathbf{J}\mathbf{s}_n \\ \mathbf{r}'_n\mathbf{J} & t_0 \end{pmatrix} \begin{pmatrix} 1 \\ \mathbf{a}_{n-1} \\ 0 \end{pmatrix} + a_{nn} \begin{pmatrix} t_0 & \mathbf{s}'_n \\ \mathbf{r}_n\mathbf{J} & \mathbf{T}_{n-1} \end{pmatrix} \begin{pmatrix} 0 \\ \mathbf{J}\mathbf{b}_{n-1} \\ 1 \end{pmatrix}.$$

Unter Verwendung von Gleichung (A.4) und (A.5) ergibt sich aus (A.9):

$$(A.10) \quad \begin{pmatrix} c_n \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} c_{n-1} \\ 0 \\ \vdots \\ 0 \\ \Delta_n \end{pmatrix} + a_{nn} \begin{pmatrix} \nabla_n \\ 0 \\ \vdots \\ 0 \\ c_{n-1} \end{pmatrix}$$

mit

$$(A.11) \quad \Delta_n = \mathbf{r}'_n\mathbf{J} \begin{pmatrix} 1 \\ \mathbf{a}_{n-1} \end{pmatrix} = \sum_{l=1}^n t_l a_{(n-l)(n-1)}, \quad a_{0(n-1)} = 1$$

$$(A.12) \quad \nabla_n = \mathbf{s}'_n \begin{pmatrix} \mathbf{J}\mathbf{b}_{n-1} \\ 1 \end{pmatrix} = \sum_{l=1}^n t_{-l} b_{(n-l)(n-1)}, \quad b_{0(n-1)} = 1.$$

Aus der letzten Zeile von (A.10) ergibt sich:

$$(A.13) \quad a_{nn} = -\Delta_n/c_{n-1},$$

so daß sich die erste Zeile von (A.10) darstellen läßt als:

$$(A.14) \quad c_n = c_{n-1} + a_{nn}\nabla_n = -\Delta_n\nabla_n/c_{n-1}.$$

Analog ergibt sich als Update-Gleichung für den Vektor \mathbf{b}_n :

$$(A.15) \quad \begin{pmatrix} \mathbf{J}\mathbf{b}_n \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ \mathbf{J}\mathbf{b}_{n-1} \\ 1 \end{pmatrix} + b_{nn} \begin{pmatrix} 1 \\ \mathbf{a}_{n-1} \\ 0 \end{pmatrix},$$

bzw. bei elementeiseer Betrachtung für $l = 1, \dots, (n-1)$:

$$(A.16) \quad b_{ln} = b_{l(n-1)} + b_{nn}a_{(n-l)(n-1)}.$$

Prämultiplikation von (A.15) mit \mathbf{T}_n ergibt:

$$(A.17) \quad \mathbf{T}_n \begin{pmatrix} \mathbf{J}\mathbf{b}_n \\ 1 \end{pmatrix} = \begin{pmatrix} t_0 & \mathbf{s}'_n \\ \mathbf{r}_n\mathbf{J} & \mathbf{T}_{n-1} \end{pmatrix} \begin{pmatrix} 0 \\ \mathbf{J}\mathbf{b}_{n-1} \\ 1 \end{pmatrix} + b_{nn} \begin{pmatrix} \mathbf{T}_{n-1} & \mathbf{J}\mathbf{s}_n \\ \mathbf{r}'_n\mathbf{J} & t_0 \end{pmatrix} \begin{pmatrix} 1 \\ \mathbf{a}_{n-1} \\ 0 \end{pmatrix},$$

was sich zusammenfassen läßt zu:

$$(A.18) \quad \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ c_n \end{pmatrix} = \begin{pmatrix} \nabla_n \\ 0 \\ \vdots \\ 0 \\ c_{n-1} \end{pmatrix} + b_{nn} \begin{pmatrix} c_{n-1} \\ 0 \\ \vdots \\ 0 \\ \Delta_n \end{pmatrix}.$$

Δ_n und ∇_n in (A.18) sind wie oben in (A.11) und (A.12) definiert.
Aus der ersten Zeile von (A.18) folgt:

$$(A.19) \quad b_{nn} = -\nabla_n/c_{n-1}.$$

Mit (A.19) läßt sich (A.14) darstellen als:

$$(A.20) \quad c_n = c_{n-1} + a_{nn}(-b_{nn}c_{n-1}) = c_{n-1}(1 - a_{nn}b_{nn}).$$

Mit den Gleichungen (A.11), (A.12), (A.13), (A.19), (A.8), (A.16) und (A.14) können die Vektoren \mathbf{a}_n und \mathbf{b}_n rekursiv für $n = 0, \dots, N$ berechnet werden. Die vollständige Darstellung der Inversen \mathbf{U}_N unter Verwendung von \mathbf{a}_n und \mathbf{b}_n läßt sich aus der folgenden Darstellung der partitionierten Matrix \mathbf{U}_N ableiten (vgl. MARPLE[1987, S. 85ff]):

$$(A.21) \quad \mathbf{U}_N = \frac{1}{c_N} \begin{pmatrix} 1 & \mathbf{b}'_N \\ \mathbf{a}_N \mathbf{J} & c_N \mathbf{U}_{N-1} + \mathbf{a}_N \mathbf{b}'_N \end{pmatrix} = \frac{1}{c_N} \begin{pmatrix} c_N \mathbf{U}_{N-1} + \mathbf{J} \mathbf{b}_N \mathbf{a}'_N \mathbf{J} & \mathbf{J} \mathbf{b}_N \\ \mathbf{a}'_N \mathbf{J} & 1 \end{pmatrix}.$$

Mit Hilfe von (A.21) ergeben sich die einzelnen Elemente $u_N[l, k]$ als:

$$(A.22) \quad u_N[0, 0] = u_N[N, N] = 1/c_N$$

$$(A.23) \quad u_N[0, k] = u_N[N - k, N] = b_N[k]/c_N \quad \text{für } 1 \leq k \leq N$$

$$(A.24) \quad u_N[l, 0] = u_N[N, N - l] = a_N[l]/c_N \quad \text{für } 1 \leq l \leq N$$

$$(A.25) \quad u_N[l + 1, k + 1] = u_N[N, N - l] = a_N[l, N]/c_N \quad \text{für } 1 \leq l \leq N$$

(A.26)

$$u_N[l+1, k+1] = u_{N-1}[l, k] + a_N[l+1]b_N[k+1]/c_N \quad \text{für } 0 \leq l, k \leq N-1$$

$$(A.27) \quad u_N[l, k] = u_{N-1}[l, k] + b_N[N-l]a_N[N-k]/c_N \quad \text{für } 0 \leq l, k \leq N-1$$

(A.26) und (A.27) lassen sich so zusammenfassen, daß $u_{N-1}[l, k]$ eliminiert wird:

$$(A.28) \quad u_N[l+1, k+1] = u_N[l, k] + \frac{1}{c_N}(a_N[l+1]b_N[k+1] - b_N[N-l]a_N[N-k])$$

$$\text{für } 0 \leq l, k \leq N-1.$$

Für den Spezialfall einer hermiteschen Toeplitzmatrix mit $\mathbf{T}^H = \mathbf{T}$ gilt $\mathbf{r}_n = \mathbf{s}_n^*$, so daß folgt:

$$(A.29) \quad \mathbf{a}_n = \mathbf{b}_n^*$$

und

$$(A.30) \quad \Delta_n = \nabla_n^*.$$

c_n ergibt sich in diesem Fall als:

$$(A.31) \quad c_n = c_{n-1}(1 - |a_{nn}|^2),$$

so daß c_n reell und positiv ist.

Zur Lösung des linearen Gleichungssystems

$$(A.32) \quad \mathbf{T}_N \mathbf{x}_N = \mathbf{y}_N,$$

mit dem unbekanntem Lösungsvektor $\mathbf{x}_N = (x_0, \dots, x_N)$ muß die inverse Matrix \mathbf{T}^{-1}_N nicht explizit berechnet werden. Nach Ermittlung von x_0 über den Ansatz

$$(A.33) \quad t_0 x_0 = y_0,$$

kann der Lösungsvektor $\mathbf{x}_n = (x_0, \dots, x_n)'$ rekursiv für $n = 1, \dots, N$ über die Update-Gleichung

$$(A.34) \quad \mathbf{x}_n = \begin{pmatrix} \mathbf{x}_{n-1} \\ 0 \end{pmatrix} + \alpha_n \begin{pmatrix} \mathbf{J}\mathbf{b}_{n-1} \\ 1 \end{pmatrix}$$

berechnet werden. Der Skalar α_n kann bestimmt werden, indem Gleichung (A.34) mit \mathbf{T}_N prämultipliziert wird, so daß sich \mathbf{y}_n unter Verwendung einer geeigneten Partionierung gemäß (A.6) darstellen läßt als:

$$(A.35) \quad \mathbf{y}_n = \begin{pmatrix} \mathbf{y}_{n-1} \\ \beta_n \end{pmatrix} + \alpha_n \begin{pmatrix} 0 \\ \vdots \\ 0 \\ c_n \end{pmatrix}$$

mit

$$(A.36) \quad \beta_n = \mathbf{r}'_n \mathbf{J} \mathbf{x}_{n-1} = \sum_{l=0}^{n-1} x_l t_{n-l}.$$

Aus der letzten Zeile von (A.35) α_n folgt:

$$(A.37) \quad \alpha_n = \frac{y_n - \beta_n}{c_n}.$$

Die vorstehenden Ausführungen werden im folgenden Satz zusammengefaßt:

Satz A.0.1 (*Toeplitz-Gleichungssystem*)

Sei \mathbf{T}_N eine Toeplitzmatrix der Ordnung $(N + 1)$. Dann kann die Lösung \mathbf{x}_N des Gleichungssystems

$$\mathbf{T}_N \mathbf{x}_N = \mathbf{y}_N$$

rekursiv mit den Gleichungen (A.11), (A.12), (A.13), (A.19), (A.8), (A.16), (A.14), (A.37), (A.36) und (A.34) berechnet werden. ■

Die Eigenvektoren von hermiteschen Toeplitzmatrizen besitzen verschiedene spezielle Eigenschaften (vgl. z.B. CANTONI/BUTLER[1976]). Von Interesse ist hier noch der folgende Satz (vgl. z.B. MAKHOUL[1981, 1985]):

Satz A.0.2 (*Eigenvektoren hermitescher Toeplitzmatrizen*)

Sei \mathbf{T}_N eine Toeplitzmatrix der Ordnung $(N + 1)$. \mathbf{v}_{min} und \mathbf{v}_{max} seien die Eigenvektoren, die zu dem kleinsten und dem größten Eigenwert λ_{min} und λ_{max} gehören, die algebraische Vielfachheit von λ_{min} und λ_{max} sei gleich Eins.

Dann besitzen die Polynome

$$\sum_{k=0}^N \mathbf{v}_{min}[k + 1]z^{-k}$$

und

$$\sum_{k=0}^N \mathbf{v}_{max}[k + 1]z^{-k}$$

N Wurzeln auf dem Einheitskreis. ■

Literatur

- Ables, J. G. [1972]: *Maximum Entropy Spectral Analysis*. Proc. Symp. on the Collection and Analysis of Astrophysical Data, Reprint in Childers (Ed.) [1978, S. 23-31]
- Ahlberg, D. A. [1992]: *Error measures and the choice of a forecast method*. International Journal of Forecasting, Vol. 8, S. 99-100
- Akaike, H. [1969a]: *Fitting autoregressive models for prediction*. Ann. Inst. Statist. Math., Vol. 21, S. 243-247
- Akaike, H. [1969b]: *Power Spectrum Estimation through autoregressive model fitting*. Ann. Inst. Statist. Math., Vol. 21, S. 407-419
- Akaike, H. [1974]: *A new look at statistical model identification*. IEEE Trans. Autom. Control, Vol. AC-19, S. 716-723
- Akaike, H. [1977]: *On Entropy Maximization Principle*. in: Krishnaiah, P. R., (Ed.), Applications of Statistics, S. 27-41
- Akaike, H. [1977]: *On the Likelihood of a Time Series Model*. The Statistician, Vol. 27, S. 217-235
- Akaike, H. [1978a]: *A Bayesian analysis of the minimum AIC procedure*. Ann. Inst. Statist. Math., Vol. 30A, S. 9-14
- Akaike, H. [1979]: *A Bayesian extension of the minimum AIC procedure of autoregressive modell fitting*. Biometrika, Vol. 66, S. 237-242
- Alagón, J. [1989]: *Spectral Discrimination for Two Groups of Time Series*. Journal of Time Series Analysis, Vol. 10, S. 203-214
- Alonso, M./Finn, E. J. [1992]: *Physics*. Addison Wesley, New York u.a.

- Anděl, J. [1982]: *Fitting models to time series*. Math. Operationsforsch. Statist., Ser. Statistics, Vol. 13, S. 121-143
- Anderson, T. W. [1971]: *The Statistical Analysis of Time Series*. John Wiley & Sons, New York u.a.
- Andrews, D. W. K. [1991]: *Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation*. Econometrica, Vol. 59, S. 817-858
- Aoki, M. [1987]: *State Space Modelling of Time Series*. Springer Verlag, Berlin Heidelberg New York
- Åström, K. J./Hågander, P. I./Sternby, J. [1984]: *Zeros of sampled systems*. Automatica, Vol. 20, S. 31-38
- Bandemer, H./Gottwald, S. [1993]: *Einführung in Fuzzy-Methoden*. 4. erweiterte Aufl. Akademie Verlag, Berlin
- Basseville, M. [1989]: *Distance Measures for Signal Processing and Pattern Recognition*. Signal Processing, Vol. 18, S. 349-369
- Båth, C. [1974]: *Spectral Analysis in Geophysics*. Elsevier, Amsterdam
- Beltrão, K. I./Bloomfield, P. [1987]: *Determining the Bandwidth of a Kernel Spectrum Estimate*. Journal of Time Series Analysis, Vol. 8, S. 21-38
- Blaffert, T. [1984]: *Computer-assisted multicomponent spectral analysis with fuzzy data sets*. Analytica Chimica Acta, Vol. 161, S. 135-148
- Bloomfield, P. [1976]: *Fourier Analysis of Time Series: An Introduction*. John Wiley & Sons, New York u.a.
- van den Bos, A. [1971]: *Alternative interpretation of Maximum Entropy Spectral Analysis*. IEEE Trans. Inform. Theory, IT-17, S. 493-494
- Box, G. E. P./Jenkins, G. M. [1976]: *Time Series Analysis: Forecasting and Control, 2nd ed.* Holden-Day, San Francisco u.a.
- Bretthorst, G. L. [1988]: *Bayesian Spectrum Analysis and Parameter Estimation*. Springer Verlag, New York u.a.
- Brillinger, D. R. [1975]: *Time Series. Data Analysis and Theory*. Holt, Rinehart and Winston, New York u. a.
- Brillinger, D. R., Krishnaiah, P. R. (eds.) [1983]: *Handbook of Statistics, Vol. 3*. Elsevier Science Publisher B. V.

- Brillinger, D. R./Caines, P./Geweke, J./Parzen, E./ Rosenblatt, M./Taqqu, M. S. (eds.) [1992]: *New Directions in Time Series Analysis. Part I* Springer Verlag, New York Berlin Heidelberg
- Brillinger, D. R./Kaiser, R. [1992]: *Fourier and Likelihood Analysis in NMR Spectroscopy*. in: Brillinger et al. [S. 41-64]
- Brockwell, P. J./Davis, R. A. [1991]: *Time Series: Theory and Methods*. 2. erw. Aufl., Springer Verlag, New York Berlin Heidelberg
- Burg, J. P. [1967]: *Maximum Entropy Spectral Analysis*. 37th Annual International S.E.G. Meeting, Oklahoma City, Reprint in Childers (Ed.) [1978, S. 34-41]
- Burg, J. P. [1968]: *A new analysis technique for time series data*. Paper presented at NATO Advanced Study Institute on Signal Processing with Emphasis on Underwater Acoustics, 12.-23. Aug. 1968 Reprint in: Childers (Ed.) [1978, S. 42-48]
- Byrd, R. H./Tapia, R. A./Thompson, J. R.[1978]: *Optimal Smoothing of Direct Estimates of the Power Spectrum*. Commun. Statist. - Simul. Comput., Vol. B7(4), S. 335-344
- Cameron, M. A. [1987]: *An Automatic Non-Parametric Spectrum Estimator*. Journal of Time Series Analysis, Vol. 8, S. 379-387
- Cantoni, A./Butler, P. [1976]: *Eigenvalues and Eigenvectors of Symmetric Centrosymmetric Matrices*. Linear Algebra and its Applications, Vol. 13, S. 275-288
- Chatfield, C. [1992]: *A commentary on error measures*. International Journal of Forecasting, Vol. 8, S. 100-102
- Chen, W. Y./Stegen, G. R. [1974]: *Experiments with maximum entropy spectral analysis*. J. Geophys. Res., Vol. 79, S. 3019-3022
- Childers, D. G. (Ed.) [1978]: *Modern Spectrum Analysis*. IEEE Press, New York
- Coates, D. S./Diggle, P. J. [1986]: *Tests for Comparing Two Estimated Spectral Densities*. Journal of Time Series Analysis, Vol. 7, S. 7-20
- Cogburn, I./Davis, H. T. [1974]: *Periodic Splines and Spectral Estimation*. Annals of Statistics, Vol. 2, S. 1108-1126
- Dahlhaus, R. [1983]: *Spectral Analysis with Tapered Data*. Journal of Time Series Analysis, Vol. 4, S. 163-175
- Dahlhaus, R. [1984]: *Parameter Estimation of Stationary Processes with Spectra Containing Strong Peak*. in: Franke/Härdle/Martin (Eds.) [1984, S. 50-67]

- Dahlhaus, R. [1987]: Kommentar zu Hannan[1987]
- Dahlhaus, R. [1993]: *Statistical Methods in Spectral Estimation*. Beiträge zur Statistik Nr. 2, Universität Heidelberg
- Dargahi-Noubary, G. R./Laycock, P. J. [1981]: *Spectral ratio discriminants and information theory*. Journal of Time Series Analysis, Vol. 2, S. 71-86
- Diggle, P. J. [1990]: *Time Series. A Biostatistical Introduction*. Oxford University Press, Oxford
- Dreher, W. [1990]: *Beiträge zur Anwendung der Methode der maximalen Entropie in der 1-dimensionalen NMR-Spektroskopie*. Diss., AdW Berlin
- Ephraim, Y./Lev-Ari, H./Gray, R. M. [1988]: *Asymptotic Minimum Discrimination Information Measure for Asymptotically Weakly Stationary Processes*. IEEE Trans. on Information Theory, Vol. 34, S. 1033-1040
- Fougere, P. F./Zawalick, E. J./Radoski, H. R. [1976]: *Spontaneous Line Splitting in Maximum Entropy Power Spectral Analysis*. Physics of the Earth and Planetary Interiors, Vol. 12, S. 201-207
- Fougere, P. [1985]: *A Review of the Problem of Spontaneous Line Splitting in Maximum Entropy Power Spectral Analysis*. in: Smith/Grandy (Eds.) [1985, S. 303-315]
- Funke, C. [1991]: *Kubische Glättungssplines und Verallgemeinerte Kreuzvalidierung in Rahmen der nichtparametrischen Regressionsanalyse*. Diplomarbeit, TU Berlin, unveröffentlicht
- Franke, J./Härdle, W./Martin, D. (Eds.) [1984]: *Robust and Non-Linear Time-Series-Analysis*. Springer Verlag, New York Berlin Heidelberg
- Friedrich, D. [1974]: *Schätzprobleme sowie Verfahren und Möglichkeiten der Analyse ökonomischer Modelle*. J.C.B. Mohr (Paul Siebeck), Tübingen
- Fuller, W. A. [1976]: *Introduction to Statistical Time Series*. John Wiley & Sons, New York u.a.
- Gani, J. (Ed.) [1988]: *A celebration of Applied Probability*. Applied Probability Trust, Sheffield
- Gersch, W./Martinelli, F./Yonemoto, J./Low, M. D./Mc Ewan, J. A. [1979]: *Automatic Classifications of Electroencephalograms: Kullback-Leibler Nearest Neighbor Rules*. Science, Vol. 205, S. 193-195

- Geyer, A. [1985]: *Maximum Entropie Spektralanalyse ökonomischer Zeitreihen*. Diss., Wirtschaftsuniversität Wien
- Giaşu, S. [1977]: *Information Theory with Applications*. Mc Graw Hill, New York
- Granger, C. W. J./Hatanaka, M. [1964]: *Spectral Analysis of Economic Time Series*. Princeton University Press, Princeton, NJ
- Gray, R. M./Markel J. D. [1976]: *Distance Measures for Speech Processing*. IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. ASSP-24, S. 380-391
- Gray, R. M./Buzo, A./Gray, A. H./Matsuyama, Y. [1980]: *Distortion Measures for Speech Processing*. IEEE ASSP, Vol. ASSP-28, S. 367-76
- den Haan, W. J./Levin, A. [1994]: *Inferences from Parametric and Non-Parametric Covariance Matrix Estimation Procedures*. Fifth (EC)², 15.-17. Dez. 1994 in Berlin
- Hainz, G. [1992]: *The Asymptotic Properties of Burg Estimators*. Beiträge zur Statistik Nr. 18, Universität Heidelberg
- Hall, P./Marron, J. S. [1991]: *Local Minima in Cross-Validation Functions*. J. Royal Stat. Soc. B, Vol. 53, S. 245-252
- Hannan, E. J. [1987]: *Rational Transfer Function Approach*. Statistical Science, Vol. 2, S. 135-161
- Hannan, E. J./Krishnaiah, P.R./Rao, M. M. (Eds.) [1985]: *Handbook of Statistics, Vol. 5*. Elsevier Science Publishers B.V.
- Hannan, E. J./Quinn, B. G. [1979]: *The determination of the order of an autoregression*. J. R. Statist. Soc. B, Vol. 41, S. 190-195
- Hannan, E. J./Quinn, B. G. [1989]: *The Resolution of Closely Adjacent Spectral Lines*. Journal of Time Series Analysis, Vol. 10, S. 13-31
- Hannan, E. J./Rissanen, J. [1988]: *The Width of a Spectral Window*. in: Gani[1988, 301-307]
- Harris, B. (Ed.) [1967]: *Spectral Analysis of Time Series*. John Wiley & Sons, New York u.a.
- Haykin, S. (Ed.) [1983]: *Nonlinear Methods of Spectral Analysis*. Springer Verlag, Berlin New York Heidelberg, 2. Aufl.
- Haykin, S. [1989]: *Modern filters*. Macmillan Publ. Comp., New York
- Haykin, S. Kesler, S. [1983]: *Prediction-Error-Filtering and Maximum-Entropie Spectral Estimation*. in: Haykin (Ed.) [1983, S. 9-72]

- Helme, B./Nikias, C. L. [1985]: *Improved Spectrum Performance via Data-Adaptive Weighted Burg Technique*. IEEE Trans. Acoust. Speech Signal Process., Vol. ASSP-33, S. 903-910
- Herring, R. W. [1980]: *The Cause of Line Splitting in Burg Maximum-Entropy Spectral Analysis*. IEEE Transactions on Acoustics, Speech and Signal Processing, Vol. 28, S. 692-701
- Huang, D. [1990]: *Levinson-Type Recursive Algorithms for Least-Squares Autoregression*. Journal of Time Series Analysis, Vol. 11, S. 295-315
- Hurvich, C. M. [1985]: *Data-Driven Choice of a Spectrum Estimate: Extending the Applicability of Cross-Validation Methods*. JASA, Vol. 80, S. 933-940
- Hurvich, C. M. [1986]: *Data-Dependent Spectral Windows: Generalizing the Classical Framework to Include Maximum Entropy Estimates*. Technometrics, Vol. 28, S. 259-268
- Hurvich, C. M. [1987]: *Automatic Selection of a Linear Predictor Through Frequency Domain Cross-Validation*. Commun. Statist. – Theory Meth., Vol. 16(11), S. 3199-3234
- Hurvich, C. M. [1988]: *A mean squared error criterion for time series data windows*. Biometrika, Vol. 75, S. 485-490
- Hurvich, C. M. [1992]: *Selection of Time Series Models and Spectrum Estimates Using a Bias-Corrected Generalization of AIC*. in: Brillinger et. al [1992, 155-168]
- Hurvich, C. M., Beltrão, K. I. [1990]: *Cross-Validatory Choice of a Spectrum Estimate and its Connection with AIC*. Journal of Time Series Analysis, Vol. 11, S. 121-137
- Hurvich, C. M., Tsai, C.-L. [1989]: *Regression and Time Series Model Selection in Small Samples*. Biometrika, Vol. 76, S. 297-307
- Hurvich, C. M., Tsai, C.-L. [1993]: *A Corrected Akaike Information Criterion for Vector Autoregressive Model Selection*. Journal of Time Series Analysis, Vol. 14, S. 271-279
- Hurvich, C. M., Zeger, S. L. [1990]: *A Frequency Domain Selection Criterion for Regression With Autocorrelated Errors*. JASA, Vol. 85, S. 705-714
- Hurvich, C. M., Shumway, R. Tsai, C.-L. [1990]: *Improved estimators of Kullback-Leibler information for autoregressive model selection in small samples*. Biometrika, Vol. 77, S. 709-719
- Ingmann, D./Merlis, Y. [1992]: *Maximum Entropy Signal Reconstruction with Neural Networks*. IEEE Transactions on Neural Networks, Vol. 3, S. 195-201

- Itakura, F./Saito, S. [1970]: *A Statistical Method for Estimation of Speech Spectral Density and Formant Frequencies*. Electr. and Comm. in Japan, Vol. 53-A, S. 36-43
- Jaynes, E. T. [1957a]: *Information Theory and Statistical Mechanics I*. Phys. Rev. Vol. 106, S. 620-630, Reprint in: Jaynes[1983]
- Jaynes, E. T. [1957b]: *Information Theory and Statistical Mechanics II*. Phys. Rev. Vol. 108, S. 171-190, Reprint in: Jaynes[1983]
- Jaynes, E. T. [1983]: *Papers on Probability, Statistics and Statistical Physics, a reprint collection*. D. Reidel, Dordrecht-Holland
- Jenkins, G. M./Watts, D. G. [1968]: *Spectral Analysis and its Applications*. Holden-Day, San Francisco u.a.
- Jones, L. K./Byrne, Ch. L. [1990]: General Entropy Criteria for Inverse Problems, with Applications to Data Compression, Pattern Classification, and Cluster Analysis. IEEE Transactions on Information Theory, Vol. 36, S. 23-30
- Johnson, D. H./Dudgeon, D. E. [1993]: *Array Signal Processing*. Prentice Hall, Englewood Cliffs, NJ 07632
- Kailath, Th. [1967]: *The Divergence and Battacharryya Distance Measures in Signal Selection*. IEEE Transactions on Communication Technology, Vol. COM-15, S. 52-60
- Kashyap, K. L. [1980]: *Inconsistency of the AIC Rule for Estimating the Order of Autoregressive Models*. IEEE Transactions on Automatic Control, Vol. AC-25, S. 996-998
- Kay, S. M. [1988]: *Modern Spectral Estimation*. Prentice Hall, Englewood Cliffs, NJ 07632
- Kay, S. M./Marple, S. L., Jr. [1979]: *Sources of and Remedies for Spectral Line Splitting in Autoregressive Spectrum Analysis*. Proceedings of the 1979 IEEE International Conference on Acoustics, Speech Signal Processing, S. 151-154
- Kazakos, D./Papantoni-Kazakos, P. [1980]: *Spectral Distance Measures between Gaussian Processes*. IEEE Trans. on Automatic Control, Vol. AC-25, S. 950-958
- Kesler, S. B. (Ed.) [1986]: *Modern Spectrum Analysis II*. IEEE Press, New York
- Klir, G. J./Folger, T. A. [1988]: *Fuzzy Sets, Uncertainty, and Information*. Prentice Hall International, London
- König, H./Wolters, J. [1972]: *Einführung in die Spektralanalyse ökonomischer Zeitreihen*. Anton Hain Verlag, Meisenheim am Glan
- Kooperberg, C./Stone, C. J. [1991]: *A study of logspline density estimation*. Computational Statistics and Data Analysis, Vol.12, S. 327-347

- Kullback, S. [1959]: *Information, Theory and Statistics*. John Wiley, New York
- Kullback, S./Leibler, R. A. [1951]: *On Information and Sufficiency*. Annals of Math. Stat., Vol. 22, S. 79-86
- Landers, T. E./Lacoss, R. T. [1977]: *Some Geophysical Applications of Autoregressive Spectral Estimates*. IEEE Transactions on Geoscience Electronics, Vol. GE-15, S. 26-32
- Lambert, P. J./Poskitt, D. S. [1983]: *Stationary Processes in Time Series Analysis: The Mathematical Foundations*. Vandenhoeck & Ruprecht Göttingen
- Lamperti, J. [1977]: *Stochastic Processes*. Springer Verlag New York Heidelberg Berlin
- Lim, J. S. [1990]: *Two-dimensional Signal and Image Processing*. Prentice Hall, Englewood Cliffs, N. J.
- Linhart, H./Zucchini, W. [1986]: *Model Selection*. John Wiley & Sons, New York u.a.
- Lysne, D./Tjostheim, D. [1987]: *Loss of Spectral Peaks in Autoregressive Spectral Estimation*. Biometrika, Vol. 74, S. 200-206
- Makhoul, J. [1977]: *Stable and Efficient Lattice Methods for Linear Prediction*. IEEE Vol. ASSP-25, S. 423-428, Korrekturen: IEEE Vol. ASSP-26 [1978], S. 111
- Makhoul, J. [1981]: *On the Eigenvectors of Symmetric Toeplitz Matrices*. IEEE Vol. ASSP-29, S. 868-872, Korrekturen siehe Makhoul[1985]
- Makhoul, J. [1985]: *Author's Reply to "Comments and Corrections to 'On the Eigenvectors of Symmetric Toeplitz Matrices' "*. IEEE Vol. ASSP-33, S. 737-738
- Markel, J. D./Gray, A. H., Jr. [1976]: *Linear Prediction of Speech*. Springer Verlag, Berlin Heidelberg New York
- Marple, S. L. [1980]: *A new autoregressive spectrum analysis algorithmen*. IEEE Trans. Acoust. Speech Signal Process., Vol. ASSP-28, S. 441-454
- Marple, S. L. [1983]: *A Tale of Two Periodi(cities): The case for Spectral Estimation Performance Metrics*. Proc. 2th IEEE ASSP Spectrum Estimation Workshop, Tampa, S. 7-11
- Marple, S. L. [1987]: *Digital Spectral Analysis*. Prentice Hall, Englewood Cliffs, NJ 07632
- Marron, J. S. [1988]: *Automatic Smoothing Parameter Selection: A Survey*. Econometric Reviews, S. 65-85
- Matsuoka, T./Ulrych, T. J. [1984]: *Phase Estimation Using the Bispectrum*. Proc. of the IEEE, Vol. 72, S. 1403-1411

- MacDonald, G. J. [1990]: *Global Climate Change*. in: MacDonald/Sertorio [1990, S. 1 - 95]
- MacDonald, G. J./Sertorio, L. [1990]: *Global Climate Change and Ecosystem Change*. NATO ASI Series B: Physics, Vol. 240, Plenum Press, New York London
- Moskilde, E./Moskilde, L. [1991]: *Complexity, Chaos and Biological Evolution*. Proc. of a NATO Advanced Research Workshop on Complex Dynamics and Biological Evolution, August 6-10, 1990, Hindsgavl, Denmark, Plenum Press, New York London
- Narayan, R./Nityananda, R. [1983]: *Maximum-Entropy — Flexibility versus Fundamentalism*. in: Roberts [1983, S. 281-290]
- Newton, J. H. [1988]: *TIMESLAB: A Time Series Analysis Laboratory*. Wadsworth, Inc., Belmont, CA
- Nuttal, A. H. [1976]: Spectral Analysis of a Univariate Process with Bad Data Points, via Maximum Entropy and Linear Predictive Techniques. Naval Underwater Systems Center Technical Report TR-5303
- Oppenheim, A. V. (Ed.) [1978]: *Applications of Digital Signal Processing*. Prentice Hall Inc. , Englewood Cliffs, New Jersey
- Oppenheim, A. V. [1978a]: *Digital Processing of Speech*. in: Oppenheim (Ed.) [1978, S. 117-168]
- Oppenheim, A. V./Schafer, R. W. [1992]: *Zeitdiskrete Signalverarbeitung*. Oldenbourg Verlag, München Wien
- Pagano, M. [1972]: *An Algorithm for Fitting Autoregressive Schemes*. J. R. Statist. Soc. C (Applied Statistics), Vol. 21, S. 274-281
- Parzen, E. [1957]: *On Consistent Estimates of the Spectrum of a Stationary Time Series*. Ann. Math. Statist., Vol. 28, S. 329-348
- Parzen, E. [1982]: *Maximum Entropy Interpretation of Autoregressive Spectral Densities*. Statistic & Probability Letters, Vol. 1, S. 7-11
- Parzen, E. [1983]: *Autoregressive Spectral Estimation*. in: Brillinger/Krishnaiah (Eds.) [1983, S. 221-247]
- Parzen, E. [1992]: *Time Series, Statistics and Information*. in: Brillinger et al. (Eds.) [1992, Part I, S. 265-285]
- Prank, K./Harms, H./Kayser, Chr./Brabant, G./Olsen, L. F./Hesch, R. D. [1990]: *The dynamic code: information transfer in hormonal systems*. in: Moskilde/Moskilde (Eds.) [1991, S. 95-118]

- Priestley, M. B. [1981]: *Spectral Analysis and Time Series. Vol. 1 und Vol. 2.* Academic Press, London
- Pukkila, T. M. [1988]: *An Improved Estimation Method for Univariate Autoregressive Models.* Journal of Multivariate Analysis, Vol. 27, S. 422-433
- Press, W. H./Teukolsky, S. A./Vetterling, W. T./Flannery, B. P. [1992]: *Numerical Recipes in C.* 2nd ed., Cambridge University Press.
- Rényi, A. [1970]: *Probability Theory.* North-Holland, Amsterdam
- Reza, F. M. [1961]: *An Introduction to Information Theory.* McGraw Hill, New York
- Rice, J. R. [1983]: *Numerical Methods, Software and Analysis.* McGraw-Hill, Auckland u.a.
- Rissanen, J. [1990]: *Complexity of Models.* in: Zurek (Ed.) [1990, S. 117-125]
- Roberts, J. A. (Ed.) [1984]: *Indirect Imaging.* Proceedings of an International Symposium, Sydney Sept. 1983, Cambridge University Press, Cambridge u.a.
- Robinson, E. A. [1967]: *Multichannel Time Series Analysis with Digital Computer Programs.* Holden Day, San Francisco u.a.
- Robinson, E. A. [1983]: *Iterative Least-Squares Procedure for ARMA Spectral Estimation.* in: Haykin [1983], S. 127-153
- Robinson, P. M. [1991]: *Automatic Frequency Domain Inference on Semiparametric and Nonparametric Models.* Econometrica, Vol. 59, S. 1329-1363
- Sakamoto, Y./Ishigiro, M./Kitagawa, G. [1986]: *Akaike Information Criterion Statistics.* D. Reidel Publishing Company, Dordrecht u.a.
- Schmidt, Martin [1989]: *Nichtlineare Methoden der Spektralanalyse. Anwendungen in der kernmagnetischen Resonanz-Spektroskopie.* Diss., TU Berlin
- Schlittgen, R./Streitberg, B. H. J. [1984]: *Zeitreihenanalyse.* Oldenbourg Verlag, München Wien
- Soong, F. K./Sondhi, M. M. [1988]: *A Frequency-Weighted Itakura Spectral Distortion Measure and its Application to Speech Recognition in Noise.* IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. ASSP-36, S. 41-48
- Schwarz, G. [1978]: *Estimating the dimension of a model.* Ann. Stat., Vol. 6, S. 461-464
- Shannon, C. E. [1976]: *Die mathematische Theorie der Kommunikation.* in: Shannon/Weaver [1976, S. 41-143]

- Shannon, C. E./Weaver, W. [1976]: *Mathematische Grundlagen der Informationstheorie*. Oldenbourg Verlag, München Wien
- Shibata, R. [1976]: *Selection of the order of an autoregressive model by Akaike's information criterion*. *Biometrika*, Vol. 63, S. 117-126
- Shibata, R. [1980]: *Asymptotically Efficient Selection of the Order of the Model for Estimating Parameters of a Linear Process*. *Annals of Statistics*, Vol. 8, S. 147-164
- Shibata, R. [1981]: *An optimal autoregressive spectral estimate*. *Annals of Statistics*, Vol. 9, S. 300-306
- Shibata, R. [1985]: *Various Model Selection Techniques in Time Series Analysis*. in: Hannan/Krishnaiah/Rao [1985, S. 179-187]
- Shibata, R. [1989]: *Statistical aspects of model selection*. in: Willems (Ed.) [1989, S. 215-240]
- Shore, J. E./Johnson, R. W. [1983]: *Axiomatic Derivation of the Principle of Maximum Entropy and the Principle of Minimum Cross Entropy*. *IEEE Trans. Inform. Theory*, Vol. IT-26, S. 26-37
- Shumway, R. H. [1988]: *Applied Statistical Time Series Analysis*. Prentice Hall, Englewood Cliffs, New Jersey
- Skilling, J./Bryan, R. K. [1984]: *Maximum entropy image reconstruction: general algorithm*. *Monthly Notices of the Royal Astronomical Society*, Vol. 211, S. 111-124
- Skilling, J./Gull, S. F. [1985]: *Algorithms and Applications*. in: Smith/Grandy (Eds.) [1985, S. 83-132]
- Smith, C. R./Grandy, W. T. (Eds.) [1985]: *Maximum-Entropy and Bayesian Methods in Inverse Problems*. D. Reidel, Dordrecht, Boston, Lancaster
- Steinberg, H.-W./Gasser, T./Franke, J. [1985]: *Fitting Autoregressive Models to EEG Time Series: An Empirical Comparison of Estimates of the Order*. *IEEE ASSP*, Vol. 33, S. 143-155
- Stine, R. A./Shaman, P. [1990]: *Bias of Autoregressive Spectral Estimators*. *JASA* Vol. 85, S. 1091-1098
- Stone, M. [1974]: *Cross-validatory Choice and Assessment of Statistical Prediction*. *Journal of the Royal Statistical Society, B*, S. 111-147, (mit Diskussion)
- Suguiwa, N. [1978]: *Further analysis of the data by Akaike's information criterion and the finite corrections*. *Commun. Statist. - Theor. Meth.*, Vol. A7(1), S. 13-26

- Sugimoto, S./Wada, T. [1988]: Spectral Expressions of Information Measures of Gaussian Time Series and Their Relation to AIC and CAT. *IEEE Transactions on Information Theory*, Vol. 34, S. 625-631
- Swingler, D. N. [1979]: *A Modified Burg Algorithm for Maximum Entropy Spectral Analysis*. *Proc. IEEE*, Vol. 67, S. 1368-1369
- Swingler, D. N. [1980]: *Frequency Errors in MEM Processing*. *IEEE Trans. Acoust. Speech Signal Process.*, Vol. ASSP-28, S. 257-259
- Therrien, C. W. [1992]: *Discrete Random Signals and Statistical Signal Processing*. Prentice Hall, Englewood Cliffs, NJ
- Thompson, A. M., Kay, J. W., Titterton, D. M. [1989]: *A Cautionary Note About Crossvalidatory Choice*. *J. Statist. Comput. Simul.*, Vol. 33, S. 199-216
- Titterton, D. M. [1985]: *Common Structure of Smoothing Techniques in Statistics*. *International Statistical Review*, Vol. 53, S. 141-170
- Tjøstheim, D./Paulsen, D. [1983]: *Bias of some commonly-used time-series estimates*. *Biometrika*, Vol. 70, S. 389-399, Correction [1984]: Vol. 71, S. 656
- Tufts, D. W./Kumaresan, R. [1982]: *Estimation of Frequencies of Multiple Sinusoids: Making Linear Prediction Perform Like Maximum Likelihood*. *Proc. IEEE*, Vol. 70, S. 975-989
- Tukey, J. W. [1967]: *An Introduction to the Calculation of Numerical Spectrum Analysis*. in: Harris [1967, S. 25-46]
- Ulrych, T. J./Bishop, Th. N. [1975]: *Maximum Entropy Spectral Analysis and Autoregressive Decomposition*. *Review of Geophysics and Space Physics*, Vol. 13, S. 183-200
- Ulrych, T. J./Clayton, R. W. [1976]: *Time Series Modeling and Maximum Entropy*. *Physics of the Earth and Planetary Interiors*, Vol. 12, S. 188-200
- Ulrych, T. J./Ooe, M.. [1983]: *Autoregressive and Mixed Autoregressive-Moving Models and Spectra*. in: Haykin [1983, S. 73-125]
- Whaba, G. [1980]: *Automatic Smoothing of the Log Periodogram*. *JASA*, Vol. 75, S. 122-132
- Whaba, G./Wold, S. [1975a]: *A Completely Automatic French Curve: Fitting Spline Functions by Cross Validation*. *Communications in Statistics*, Vol. 4(1), S. 1-17
- Whaba, G./Wold, S. [1975b]: *Periodic Splines for Spectral Density Estimation: The Use of Cross Validation for Determining the Degree of Smoothing*. *Communications in Statistics*, Vol. 4(2), S. 125-141

- Willems, J. C. (Ed.) [1989]: *From data to model*. Springer Verlag, Berlin Heidelberg
- White, H. [1982]: *Maximum-Likelihood Estimation of Misspecified Models*. *Econometrica*, Vol. 50, S. 1-25
- Wolters, J. [1991]: *Alternative Ansätze zur Messung der Persistenz in ökonomischen Zeitreihen*. *Allgemeines Statistisches Archiv*, Vol. 75, S. 184-196
- Wolters, J. [1992]: *Persistence and Seasonality in Output and Employment of the Federal Republic of Germany*. *Recherches Economique de Louvain*, Vol. 58, S. 421-439
- Yule, G. U. [1927]: *On a Method of Investigating Periodicities in Disturbed Series With Special Reference to Wolfer's Sunspot Numbers*. *Philos. Trans. Roy. Soc. London*, Ser. A, Vol. 226, S. 267-298
- Zurek, W. H. (Ed.) [1990]: *Complexity, Entropy and the Physics of Information*. Proceedings of the Santa Fe Institute Workshop, Addison-Wesley, Redwood City
- Zurek, W. H. [1990a]: *Algorithmic Information Content, Church-Turing Thesis, Physical Entropy, and Maxwell's Demon*. in: Zurek (Ed.) [1990, S. 73-89]