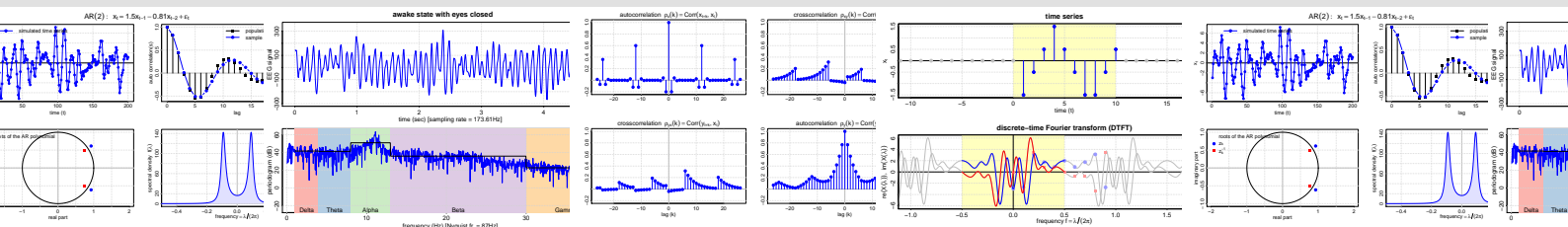


# STATIONÄRE PROZESSE UND ZEITREIHENANALYSE VO 105.453

SCHERRER WOLFGANG



INSTITUT FÜR STOCHASTIK UND WIRTSCHAFTSMATHEMATIK  
TU WIEN

Lernbehelf für die Vorlesung "Stationäre Prozesse und Zeitreihenanalyse" (VO 105.453)

Diese Unterlagen sind kein vollständiges Skriptum und decken den Inhalt der Vorlesung nur teilweise ab!  
Ich danke Hr. BSc Oliver Leodolter für das Korrekturlesen.

*Version 1.0, gedruckt am 1. März 2021*

# *Inhaltsverzeichnis*

<i>1</i>	<i>Einführung und Grundlagen</i>	<i>5</i>
<i>2</i>	<i>Stationäre Prozesse</i>	<i>7</i>
<i>3</i>	<i>Prognose</i>	<i>19</i>
<i>4</i>	<i>Spektral Darstellung</i>	<i>33</i>
<i>5</i>	<i>Lineare, dynamische Filter</i>	<i>51</i>
<i>6</i>	<i>Autoregressive Prozesse</i>	<i>63</i>
<i>7</i>	<i>ARMA Prozesse</i>	<i>89</i>
<i>8</i>	<i>Mehrgleichungs-Modelle</i>	<i>99</i>
<i>9</i>	<i>Literaturverzeichnis</i>	<i>111</i>
<i>10</i>	<i>Index</i>	<i>113</i>



# 1

## Einführung und Grundlagen

**Definition 1.1.** Eine *Zeitreihe* ist eine (endliche) Folge von zeitlich geordneten Beobachtungen/Messwerten.

$$((t_k, x_k) \mid k = 1, 2, \dots, T)$$

$t_k \in \mathbb{R}$ : Zeitpunkte,  $t_1 < t_2 < \dots < t_T$   
 $x_k \in \mathbb{R}^n$ : Messwert(e) zum Zeitpunkt  $t_k$

**Definition 1.2.** Ein *stochastischer Prozess*  $(X_t \mid t \in \mathbb{T})$  ist eine Familie von reell-wertigen Zufallsvektoren (Zufallsvariablen), die auf einem (gemeinsamen) Wahrscheinlichkeitsraum  $(\Omega, \mathcal{A}, P)$  definiert sind.

Die Indexmenge  $\mathbb{T}$  wird meist als *Zeit* interpretiert. Einen stochastischen Prozess kann man als

- eine Familie von Zufallsvektoren (Zufallsvariablen)

$$t \mapsto \begin{pmatrix} X_t(\omega) : \Omega \rightarrow \mathbb{R}^n \\ \omega \mapsto X_t(\omega) \end{pmatrix}$$

- als eine "Zufallsfunktion"

$$\omega \mapsto \begin{pmatrix} X_\times(\omega) : \mathbb{T} \rightarrow \mathbb{R}^n \\ t \mapsto X_t(\omega) \end{pmatrix}$$

- oder als eine Abbildung von  $(\Omega \times \mathbb{T})$  nach  $\mathbb{R}^n$

$$\begin{aligned} X : (\Omega \times \mathbb{T}) &\longrightarrow \mathbb{R}^n \\ (\omega, t) &\longmapsto X_t(\omega) \end{aligned}$$

interpretieren.

- Die Abbildung  $t \mapsto X_t(\omega)$ , für fixes  $\omega$ , nennt man *Trajektorie* oder *Pfad*.
- Ist der Prozess  $(X_t \mid t \in \mathbb{Z})$  integrierbar, d.h. gilt  $\mathbf{E}|X_{it}| < \infty$  für alle  $i = 1, \dots, n$  und  $t \in \mathbb{Z}$ , dann nennt man  $\mu(t) = \mathbf{E}X_t$  die *Mittelwertfunktion* des Prozesses  $(X_t)$ .

Grundlegende Konzepte und Begriffe (Zeitreihen, stochastische Prozesse, Mittelwertfunktion, Autokovarianzfunktion), Beispiele, Anwendungen, Daten, ...

skalare Zeitreihen:  $x_k \in \mathbb{R}$  ( $n = 1$ )  
 multivariate Zeitreihen:  $x_k \in \mathbb{R}^n$  ( $n > 1$ )

Wir betrachten in dieser Vorlesung nur *äquidistante* Zeitpunkte ( $t_k = t_0 + \Delta k$ ). O.E.d.A setzen wir  $t_0 = 0$  und  $\Delta = 1$ . Damit folgt  $t_k = k$  und wir schreiben kurz

$$(x_t \mid t = 1, \dots, T)$$

Time	Germany	Austria
1988Q1	248115.6	25681.6
1988Q2	257852.0	27531.2
1988Q3	266906.4	29184.7
1988Q4	286152.0	29303.6
1989Q1	266020.3	27801.9
⋮	⋮	⋮
2013Q3	703580.0	80082.2

Tabelle 1.1: Nominales Bruttoinlandsprodukt in Österreich und Deutschland, Quartalsdaten (1988Q1-2013Q3),  $T = 103$  Beobachtungen, Quelle: Eurostat.

skalare Prozesse:  $n = 1$   
 multivariate Prozesse:  $n > 1$

zeitstetige Prozesse:

$$\mathbb{T} = \mathbb{R}, \mathbb{T} = \mathbb{R}_+, \mathbb{T} = [0, 1], \dots$$

zeitdiskrete Prozesse:

$$\mathbb{T} = \mathbb{Z}, \mathbb{T} = \mathbb{N}, \dots$$

In dieser Vorlesung betrachten wir fast ausschließlich Prozesse auf  $\mathbb{Z}$ , d.h.  $(X_t \mid t \in \mathbb{Z})$  oder kurz  $(X_t)$ .

- Ist der Prozess  $(X_t | t \in \mathbb{Z})$  quadratisch integrierbar, d.h. gilt  $\mathbf{E}X_{it}^2 < \infty$  für alle  $i = 1, \dots, n$  und  $t \in \mathbb{Z}$ , dann nennt man  $\gamma(t, s) = \mathbf{Cov}(X_t, X_s) = \mathbf{E}(X_t - \mu(t))(X_s - \mu(s))'$  die Autokovarianzfunktion des Prozesses  $(X_t)$ .

Stochastische Prozesse sind (*stochastische*) Modelle für Zeitreihen. Das heißt, wir nehmen (meistens) an, dass die beobachtete Zeitreihe von einem stochastischem Prozess "erzeugt" wurde:

DGP: data generating process

- $(x_t = X_t(\omega) \in \mathbb{R}^n | t = 1, \dots, T)$  Zeitreihe

Die beobachtete Zeitreihe ist also ein endlicher Teil einer Trajektorie des zugrundeliegenden Prozesses.

Im Folgenden werden wir *nicht* mehr zwischen Zufallsvariablen/Zufallsvektoren und Realisationen/Beobachtungen unterscheiden. D.h.  $x_t$  steht sowohl für einen Zufallsvektor als auch für eine beobachtete Realisation dieses Zufallsvektors.

### 1.1 Notation

$A \in \mathbb{R}^{m \times n}$	$A'$	transponierte Matrix
$A \in \mathbb{C}^{m \times n}$	$A^*$	hermitesche transponierte Matrix
$A \in \mathbb{C}^{n \times n}$	$A \geq 0$ ( $A > 0$ )	$A$ ist positiv semidefinit (bzw. positiv definit)
$A, B \in \mathbb{C}^{n \times n}$	$A \geq B$ ( $A > B$ )	$(A - B)$ ist positiv semidefinit (bzw. positiv definit)
$A \in \mathbb{C}^{m \times n}$	$\ A\ $	eine beliebige Matrix-Norm, z.B. $\ A\ _F = \sqrt{\text{tr}(A'A)}$
$A \in \mathbb{C}^{n \times n}$	$\rho(A) = \max_i \{ \lambda_i(A) \}$	Spektralradius, maximaler Betrag der Eigenwerte von $A$
$A \in \mathbb{C}^{m \times n}$	$\text{vec}(A) = (a_{11}, \dots, a_{m1}, a_{12}, \dots, a_{m2}, a_{13}, \dots, a_{mn})'$	
$A \in \mathbb{C}^{n \times n}$	$\text{vech}(A) = (a_{11}, a_{21}, \dots, a_{n1}, a_{22}, \dots, a_{n2}, a_{33}, \dots, a_{nn})' \in \mathbb{R}^{(n(n+1)/2) \times 1}$	
$A \in \mathbb{C}^{m \times n}, B \in \mathbb{C}^{p \times q}$	$A \otimes B \in \mathbb{C}^{(mp) \times (nq)}$	Kronecker Produkt
$x \in \mathbb{R}^n$	$\ x\  = \sqrt{x'x}$	
$x \in \mathbb{C}^n$	$\ x\  = \sqrt{x^*x}$	
$\iota_n \in \mathbb{R}^{n \times 1}$	$\iota_n = (1, \dots, 1)' \in \mathbb{R}^{n \times 1}$	
<b>E, Var, Cov</b>		Erwartungswert, Varianz (Varianz-Kovarianz Matrix), Kovarianz (-Matrix)
$\mathcal{A} \subseteq \mathbb{R}$	$\chi_{\mathcal{A}}(\cdot)$	Indikatorfunktion
$x, y \in \mathbb{L}_2(\Omega, \mathcal{A}, P)$	$\langle x, y \rangle = \mathbf{E}xy$ $\ x\  = \sqrt{\mathbf{E}x^2}$	
$\mathbb{M} \subset \mathbb{L}_2(\Omega, \mathcal{A}, P)$	$\mathbf{P}_{\mathbb{M}}$	Orthogonalprojektion auf den Unterraum $\mathbb{M} = \overline{\mathbb{M}}$

## 2

# Stationäre Prozesse

### 2.1 Grundlagen

**Definition 2.1.** Ein stochastischer Prozess  $(x_t | t \in \mathbb{Z})$  ist *strikt stationär*, wenn die gemeinsame Verteilung von  $(x_{t_1+k}, \dots, x_{t_s+k})$  für alle endlichen Teilmengen  $\{t_1, \dots, t_s\} \subset \mathbb{Z}$  und für alle  $k \in \mathbb{Z}$  unabhängig von  $k$  ist.

**Definition 2.2.** Ein stochastischer Prozess  $(x_t | t \in \mathbb{Z})$  ist *schwach stationär*, wenn für alle  $t, s \in \mathbb{Z}$

1.  $\mathbb{E}x_t'x_t < \infty$
2.  $\mathbb{E}x_t = \mathbb{E}x_0$
3.  $\mathbb{E}x_t x_s' = \mathbb{E}x_{t-s} x_0'$  (oder äquivalent  $(\mathbf{Cov}(x_t, x_s) = \mathbf{Cov}(x_{t-s}, x_0))$ )

In dieser Vorlesung behandeln wir fast ausschließlich schwach stationäre Prozesse. Daher heißt "stationär" im Folgenden immer "schwach stationär" (außer es ist explizit von strikt stationären Prozessen die Rede.)

**Definition 2.3.** Die *Autokovarianzfunktion* (auto covariance function ACF) eines (schwach) stationären Prozesses  $(x_t | t \in \mathbb{Z})$  ist die Funktion

$$\begin{aligned} \gamma: \mathbb{Z} &\longrightarrow \mathbb{R}^{n \times n} \\ k &\longmapsto \gamma(k) = \mathbf{Cov}(x_{t+k}, x_t) \end{aligned} \quad \boxed{2.1}$$

und die *Autokorrelationsfunktion* (autocorrelation function ACF) ist

$$\begin{aligned} \rho: \mathbb{Z} &\longrightarrow \mathbb{R}^{n \times n} \\ k &\longmapsto \rho(k) = \mathbf{Corr}(x_{t+k}, x_t) \end{aligned} \quad \boxed{2.2}$$

Das  $(i, j)$ -te Element von  $\rho(k)$  ist die Korrelation von  $x_{i,t+k}$  und  $x_{j,t}$

$$\rho_{ij}(k) = \mathbf{Corr}(x_{i,t+k}, x_{j,t}) = \frac{\gamma_{ij}(k)}{\sqrt{\gamma_{ii}(0)\gamma_{jj}(0)}}$$

und  $\rho_{ii}(k) = \frac{\gamma_{ii}(k)}{\gamma_{ii}(0)}$ . Die Korrelationen sind beschränkt  $|\rho_{ij}(k)| \leq 1$  und es gilt natürlich  $\rho_{ii}(0) = 1$ .

strikt/schwach stationäre Prozesse, Autokovarianzfunktion (ACF), Beispiele (weißes Rauschen, MA(q) Prozesse), Konvergenz im quadratischen Mittel, (Riesz-Fisher Theorem, Stetigkeit der ersten beiden Momente), MA( $\infty$ ) Prozesse, Beispiele für nicht stationäre Prozesse (deterministischer Trend, random walk, integrierte Prozesse), Schätzung des Mittels und der Autokovarianzfunktion, Test auf weißes Rauschen

Das bedeutet natürlich insbesondere, dass alle Zufallsvektoren  $x_t$  dieselbe Verteilung haben.

Es gilt  $\gamma(k) = \gamma(-k)'$  und  $\gamma(0) = \gamma(0)'$  und für einen skalaren Prozess ( $n = 1$ ) folgt  $\gamma(k) = \gamma(-k)$ .

Für quadratische integrierbare Zufallsvariable  $x, y$  gilt die Cauchy-Schwarz Ungleichung:

$$|\mathbf{Cov}(x, y)| \leq \sqrt{\mathbf{Var}(x)\mathbf{Var}(y)}$$

und Gleichheit gilt dann und nur dann, wenn  $y$  und  $x$  linear abhängig sind, d.h. wenn  $\alpha, \beta \in \mathbb{R}$  existieren, sodass

Wir betrachten oft “gestapelte” Zufallsvektoren der Form  $x_t^k := (x_t', x_{t-1}', \dots, x_{t+1-k}')'$ . Die Kovarianzmatrix dieser Zufallsvektoren ist

$$\Gamma_k := \text{Var}(x_{t-1}^k) = (\text{Cov}(x_{t-i}, x_{t-j}))_{i,j=1,\dots,k} = (\gamma(j-i))_{i,j=1,\dots,k}$$

$$\Gamma_k = \begin{pmatrix} \gamma(0) & \gamma(1) & \gamma(2) & \cdots & \gamma(k-1) \\ \gamma(-1) & \gamma(0) & \gamma(1) & \cdots & \gamma(k-2) \\ \gamma(-2) & \gamma(-1) & \gamma(0) & \cdots & \gamma(k-3) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \gamma(1-k) & \gamma(2-k) & \gamma(3-k) & \cdots & \gamma(0) \end{pmatrix} \in \mathbb{R}^{nk \times nk}$$

2.3

Diese (Block-Toeplitz) Matrizen sind (Varianz-) Kovarianz Matrizen und daher immer positiv semidefinit (und symmetrisch).

**Definition 2.4.** Eine Folge  $(a(k) \in \mathbb{R}^{n \times n} \mid k \in \mathbb{Z})$  heißt *positiv semidefinit*, wenn sie symmetrisch ist  $(a(k) = a(-k)')$  und wenn die Block-Toeplitz Matrizen  $A_p = [a(j-i)]_{i,j=1,\dots,p}$  für all  $p \in \mathbb{N}$  positiv semidefinit sind.

**Theorem 2.5.** Eine Folge  $(a(k) \in \mathbb{R}^{n \times n} \mid k \in \mathbb{Z})$  ist dann und nur dann die Autokovarianzfunktion eines (schwach) stationären Prozesses, wenn sie positiv semidefinit ist.

Dass die ACF eines stationären Prozesses positiv semidefinit ist, folgt unmittelbar aus dem oben gesagten. Die andere Richtung folgt aus dem Existenzsatz von Kolmogorov (siehe z.B. BROCKWELL&DAVIS<sup>1</sup> (Theorem 1.2.1))

Für einen “gestapelten”, stationären Prozess der Form  $x_t = (y_t', z_t')'$  wird die Autokovarianzfunktion entsprechend partitioniert als

$$\gamma_x(k) = \begin{pmatrix} \gamma_y(k) & \gamma_{yz}(k) \\ \gamma_{zy}(k) & \gamma_z(k) \end{pmatrix}$$

Die Diagonalblöcke  $\gamma_y(\cdot)$  und  $\gamma_z(\cdot)$  sind die Autokovarianzfunktionen der Prozesse  $(y_t)$  bzw.  $(z_t)$  und  $\gamma_{yz}(\cdot)$  ist die Kreuzkovarianzfunktion zwischen den Prozessen  $(y_t)$  und  $(z_t)$ . Analog kann man auch die Autokorrelationsfunktion partitionieren (und interpretieren).

## 2.2 Konvergenz im quadratischen Mittel

**Definition 2.6.** Eine Folge von Zufallsvektoren  $(x_k \mid k \in \mathbb{N})$  konvergiert im quadratischen Mittel, wenn

$$\mathbb{E}x_0'x_0 < \infty, \quad \mathbb{E}x_k'x_k < \infty \quad \forall k \in \mathbb{N}$$

und

$$\lim_{k \rightarrow \infty} \mathbb{E}(x_k - x_0)'(x_k - x_0) = 0$$

Wir schreiben

$$x_0 = \text{l.i.m}_{k \rightarrow \infty} x_k \quad \text{oder} \quad x_k \xrightarrow{m.s.} x_0$$

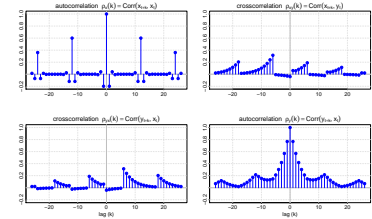


Abbildung 2.1: Autokorrelationsfunktion eines AR(13) Prozesses.

<sup>1</sup> P.J. Brockwell and R.A. Davis. *Time Series: Theory and Methods*. Springer series in statistics. Springer-Verlag, New York, 2nd edition, 1991



Der Grenzwert  $x_0$  ist fast sicher eindeutig, d.h. falls  $x_0 = \text{l.i.m } x_k$  und  $\tilde{x}_0 = \text{l.i.m } x_k$  dann folgt  $\mathbf{E}(x_0 - \tilde{x}_0)'(x_0 - \tilde{x}_0) = 0$  und daher  $x_0 = \tilde{x}_0$  f.s. Die ersten beiden Momente sind *stetig* bzgl. der Konvergenz im quadratischen Mittel, d.h. für zwei (l.i.m) konvergente Folgen  $(x_k)$  und  $(y_k)$  gilt

$$\begin{aligned}\mathbf{E}((\text{l.i.m } x_k)(\text{l.i.m } y_k)') &= \lim_{k \rightarrow \infty} (\mathbf{E}x_k y_k') \\ \mathbf{E}(\text{l.i.m } x_k) &= \lim_{k \rightarrow \infty} (\mathbf{E}x_k)\end{aligned}$$

Das folgende Theorem (siehe z.B. BROCKWELL&DAVIS<sup>2</sup> (Proposition 2.7.1)) zeigt, dass ein Grenzwert im quadratischen Mittel dann und nur dann existiert, wenn die Folge der Zufallsvektoren eine "Cauchy-Folge" ist. Dabei wird der "Abstand" zwischen  $x_k$  und  $x_l$  definiert durch  $\mathbf{E}(x_k - x_l)'(x_k - x_l)$ .

<sup>2</sup> P.J. Brockwell and R.A. Davis. *Time Series: Theory and Methods*. Springer series in statistics. Springer-Verlag, New York, 2nd edition, 1991

**Theorem 2.7.** (Riesz-Fisher Theorem) Sei  $(x_k | k \in \mathbb{N})$  eine Folge von quadratisch integrierbaren Zufallsvektoren (d.h.  $\mathbf{E}x_k'x_k < \infty$ ). Diese Folge besitzt dann und nur dann einen Grenzwert (im l.i.m Sinne) wenn für jedes  $\epsilon > 0$  ein  $k_0 \in \mathbb{N}$  existiert, sodass  $\mathbf{E}(x_k - x_l)'(x_k - x_l) \leq \epsilon$  für alle  $k, l \geq k_0$ .

### 2.3 Beispiele von stationären Prozessen

**Definition 2.8.** Ein Prozess  $(\epsilon_t | t \in \mathbb{Z})$  (oder  $(\epsilon_t | t \in \mathbb{N})$ ) ist ein IID Prozess, wenn die  $\epsilon_t$ 's identisch verteilt und voneinander unabhängig sind.

IID Prozesse sind strikt stationär. Sind die  $\epsilon_t$ 's auch quadratisch integrierbar, dann ist  $(\epsilon_t)$  auch schwach stationär. Wir verwenden die Notation  $(\epsilon_t) \sim \text{IID}(\Sigma)$  für einen IID Prozess mit Erwartungswert  $\mathbf{E}\epsilon_t = 0$  und Varianz  $\mathbf{Var}(\epsilon_t) = \Sigma$ .

**Definition 2.9.** Ein ( $n$ -dimensionaler) Prozess  $(\epsilon_t | t \in \mathbb{Z})$  ist ein *weißes Rauschen* (white noise), wenn für alle  $t, s \in \mathbb{Z}, t \neq s$

1.  $\mathbf{E}\epsilon_t'\epsilon_t < \infty$
2.  $\mathbf{E}\epsilon_t = 0 \in \mathbb{R}^n$
3.  $\mathbf{E}\epsilon_t\epsilon_t' = \mathbf{E}\epsilon_0\epsilon_0' = \Sigma \in \mathbb{R}^{n \times n}$
4.  $\mathbf{E}\epsilon_t\epsilon_s' = 0 \in \mathbb{R}^{n \times n}$

Wir verwenden oft die Notation  $(\epsilon_t) \sim \text{WN}(\Sigma)$  für einen white noise Prozess mit Varianz  $\mathbf{E}\epsilon_t\epsilon_t' = \Sigma$ . Klarerweise ist ein white noise Prozess schwach stationär.

White noise Prozesse haben "kein Gedächtnis", da  $\gamma(k) = 0$  für  $k \neq 0$ .

**Definition 2.10.** Sei  $(\epsilon_t) \sim \text{WN}(\Sigma)$  eine white noise Prozess,  $b_0, b_1, \dots, b_q \in \mathbb{R}^{m \times n}$  und  $b_0\Sigma b_q' \neq 0$ . Dann nennt man

$$(x_t = b_0\epsilon_t + \dots + b_q\epsilon_{t-q} | t \in \mathbb{Z})$$

einen  $\text{MA}(q)$  Prozess (*moving average* Prozess der Ordnung  $q$ ).

Ein MA(q) Prozess ist schwach stationär mit Mittelwertfunktion

$$\mathbf{E}x_t = \mathbf{E}(b_0\epsilon_t + \dots + b_q\epsilon_{t-q}) = 0$$

und Autokovarianzfunktion

$$\begin{aligned}\gamma(k) &= \mathbf{Cov}(x_{t+k}, x_t) \\ &= \mathbf{E}(b_0\epsilon_{t+k} + \dots + b_q\epsilon_{t+k-q})(b_0\epsilon_t + \dots + b_q\epsilon_{t-q})' \\ &= \sum_{i,j=0}^q b_i \mathbf{E}\epsilon_{t+k-i}\epsilon_{t-j}' b_j' \\ &= \sum_{j=0}^q b_{j+k} \Sigma b_j'\end{aligned}$$

2.4

In der Doppelsumme in der dritten Zeile fallen alle Terme weg, bis auf die Terme für die  $t+k-i = t-j$  gilt. In der letzten Zeile haben wir der Einfachheit halber,  $b_j = 0$  für  $j < 0$  oder  $j > q$  gesetzt. Man kann zeigen, dass ein stationärer Prozess dann und nur dann ein MA(q) Prozess ist, wenn die Autokovarianzen  $\gamma(k)$  für alle lags  $|k| > q$  gleich null sind. Weiters kann man sich o.E.d.A. auf den Fall  $m = n$  und  $b_0 = I_n$  beschränken.

**Definition 2.11.** Ein MA( $\infty$ ) Prozess  $(x_t | t \in \mathbb{Z})$  ist ein Prozess der Form

$$x_t = \sum_{j=-\infty}^{\infty} b_j \epsilon_{t-j}$$

wobei  $(\epsilon_t | t \in \mathbb{Z}) \sim \text{WN}(\Sigma)$  ein  $n$ -dimensionales weißes Rauschen ist und die Folge  $(b_j \in \mathbb{R}^{m \times n} | j \in \mathbb{Z})$  quadratisch summierbar<sup>3</sup> ist, d.h.

$$\sum_{j=-\infty}^{\infty} \|b_j\|^2 < \infty$$

**Satz 2.12.** Die unendliche Summe  $\sum_{j=-\infty}^{\infty} b_j \epsilon_{t-j}$  existiert (als l.i.m Grenzwert der Partialsummen) dann und nur dann, wenn  $\sum_{j=-\infty}^{\infty} \text{tr}(b_j \Sigma b_j') < \infty$ . Für  $\Sigma > 0$  ist diese Bedingung äquivalent zu  $\sum_j \|b_j\|^2 < \infty$ . MA( $\infty$ ) Prozesse sind schwach stationär mit

$$\mathbf{E}x_t = 0$$

und

$$\gamma(k) = \mathbf{Cov}(x_{t+k}, x_t) = \sum_{j=-\infty}^{\infty} b_{j+k} \Sigma b_j'$$

2.5

Einen MA( $\infty$ ) Prozess  $(x_t = \sum_{j \geq 0} b_j \epsilon_{t-j})$  nennt man kausal.

Es gibt noch eine ganze Reihe von Beispielen/Klassen von stationären Prozessen, die etwas später behandelt werden:

- harmonische Prozesse, siehe Abschnitt 4
- AR Prozesse, siehe Abschnitt 6.
- ARMA, ARX, ARMAX, ... Prozesse, siehe Abschnitt 7.

White noise Prozesse werden vor allem als "Bausteine" für komplexere Prozesse verwendet.

MA(q) Prozesse haben ein "endliches Gedächtnis", da  $\gamma(k) = 0$  für  $|k| > q$ .

<sup>3</sup> Hier und im Folgenden bezeichnet  $\|A\|$  eine beliebige Matrix Norm, z.B. die Frobeniusnorm  $\|A\|_F^2 = \text{tr}(A'A) = \sum_{ij} a_{ij}^2$  oder die Spektralnorm  $\|A\|_2 = \max\{\lambda_i(A'A)\}$ . Diese beiden Matrix Normen sind submultiplikativ, d.h. es gilt  $\|AB\| \leq \|A\| \|B\|$ .

MA( $\infty$ ) Prozesse haben ein "schwindendes Gedächtnis", da  $\gamma(k) \rightarrow 0$  for  $|k| \rightarrow \infty$ .

## 2.4 Beispiele für nicht stationäre Prozesse

Ein einfaches Bsp. für einen nicht stationären Prozess ist

$$x_t = \mu_t + u_t$$

wobei  $(u_t)$  stationär (mit Mittelwert gleich Null) ist und  $\mu_t \neq \text{const}$  eine deterministische Funktion der Zeit bezeichnet. Im simpelsten Fall ist  $\mu_t = \alpha + \beta t$  ein linearer *Trend*.

Ein *random walk* (Irrfahrt) ist ein Prozess der Form

$$(x_t = \sum_{k=1}^t \epsilon_k \mid t \in \mathbb{N})$$

wobei  $(\epsilon_t) \sim \text{WN}(\sigma^2)$  ein (skalares) weißes Rauschen ist. Ein *random walk mit Drift* ist definiert durch

$$(x_t = \alpha + \beta t + \sum_{k=1}^t \epsilon_k \mid t \in \mathbb{N})$$

Die Momente von  $x_t$  sind

$$\mathbb{E}x_t = \alpha + \beta t$$

$$\text{Cov}(x_t, x_s) = \min(t, s)\sigma^2$$

und daher ist  $(x_t)$  nicht stationär. Der Prozess  $(x_t)$  ist eine Lösung der Differenzengleichung

$$x_t - x_{t-1} = \beta + \epsilon_t, \quad t = 1, 2, 3, \dots$$

wenn man  $x_0 = \alpha$  setzt.

Etwas allgemeiner, betrachtet man oft Prozesse, die Lösungen der Differenzengleichung

$$x_t = x_{t-1} + u_t, \quad t = 1, 2, 3, \dots$$

sind, wobei  $(u_t)$  einen stationären Prozess bezeichnet. Die Lösungen dieser Differenzengleichung sind

$$x_t = x_0 + \sum_{j=1}^t u_j$$

Typischerweise ist  $(x_t)$  nicht stationär<sup>4</sup>. Die ersten Differenzen,  $\Delta x_t = (x_t - x_{t-1}) = u_t$  sind aber per Konstruktion stationär. Prozesse mit dieser Eigenschaft nennt man integriert der Ordnung eins ( $I(1)$  Prozess). Entsprechend nennt man einen stationären Prozess manchmal auch integriert der Ordnung Null, bzw.  $I(0)$  Prozess.

<sup>4</sup> Beachten Sie aber folgendes Beispiel:  $u_t = v_t - v_{t-1}$ , wobei  $(v_t)$  stationär ist. Der Prozess

$$x_t = x_0 + \sum_{j=1}^t u_j = x_0 + v_t - v_0$$

ist *stationär*, wenn man als Startwert  $x_0 = v_0$  wählt.

## 2.5 Explorative Zeitreihenanalyse

Siehe Handouts (StatPrZRA.handouts.pdf).

## 2.6 Schätzung des Mittelwerts

Das Stichprobenmittel

$$\bar{x}_T = \frac{1}{T} \sum_{t=1}^T x_t$$

ist ein *unverzerrter* Schätzer des Erwartungswertes  $\mathbf{E}x_t = \mu$ , da  $\mathbf{E}\bar{x}_T = \mu$ . Die Varianz des Stichprobenmittels ist

$$\mathbf{Var}(\bar{x}_T) = \frac{1}{T^2} \sum_{t,s=1}^T \underbrace{\mathbf{Cov}(x_t, x_s)}_{\gamma(t-s)} = \frac{1}{T} \sum_{k=-T+1}^{T-1} \left(1 - \frac{|k|}{T}\right) \gamma(k) \quad \boxed{2.6}$$

Ein stationärer Prozess heißt Mittelwert-ergodisch<sup>5</sup> wenn das Stichprobenmittel  $\bar{x}_T$  gegen den Erwartungswert  $\mathbf{E}x_0 = \mu$  konvergiert. d.h. wenn "Mitteln in der Zeit" und "Mitteln im Ensemble" dasselbe Ergebnis liefern.

**Theorem 2.13.**

1. Wenn  $\lim_{k \rightarrow \infty} \gamma(k) = 0$  dann folgt  $\lim_{T \rightarrow \infty} \mathbf{Var}(\bar{x}_T) = 0$  und das Stichprobenmittel ist daher konsistent (im quadratischen Mittel und in Wahrscheinlichkeit).
2. Falls  $\sum_{k=-\infty}^{\infty} \|\gamma(k)\| < \infty$  dann gilt  $\lim_{T \rightarrow \infty} T \mathbf{Var}(\bar{x}_T) = \sum_{k=-\infty}^{\infty} \gamma(k)$ .
3. Für einen Prozess der Form  $x_t = \mu + \sum_{j=-\infty}^{\infty} b_j \epsilon_{t-j}$  wobei  $(\epsilon_t) \sim \text{WN}(\Sigma)$  und  $\sum_{j=-\infty}^{\infty} \|b_j\| < \infty$  folgt  $\sum_{k=-\infty}^{\infty} \|\gamma(k)\| < \infty$  und  $\sum_{k=-\infty}^{\infty} \gamma(k) = (\sum_{j=-\infty}^{\infty} b_j) \Sigma (\sum_{j=-\infty}^{\infty} b_j)'$ .
4. Gilt zusätzlich  $(\epsilon_t) \sim \text{IID}(\Sigma)$  und  $\det(\sum_{k=-\infty}^{\infty} \gamma(k)) \neq 0$  dann ist das Stichprobenmittel asymptotisch normalverteilt mit

$$\sqrt{T}(\bar{x}_T - \mu) \xrightarrow{d} N(0, V)$$

$$\text{wobei } V = \sum_{k=-\infty}^{\infty} \gamma(k) = (\sum_{j=-\infty}^{\infty} b_j) \Sigma (\sum_{j=-\infty}^{\infty} b_j)'$$

*Bemerkung.*

- Die wesentliche Bedingung für das obige Theorem ist, dass die Kovarianzen mit wachsendem Lag schnell genug gegen Null konvergieren, d.h. dass der Prozess "schnell genug vergisst"<sup>6</sup>.
- Falls  $\sum_k \gamma(k)$  singular ist, dann kann die Konvergenz schneller sein. Betrachten Sie z.B. den MA(1) Prozess  $x_t = \epsilon_t - \epsilon_{t-1}$ . Hier ist  $\sum_k \gamma(k) = -\sigma^2 + 2\sigma^2 - \sigma^2 = 0$  und  $\bar{x}_T = \frac{1}{T} \sum_{t=1}^T (\epsilon_t - \epsilon_{t-1}) = \frac{1}{T} (\epsilon_T - \epsilon_0)$ . Die Varianz des Stichprobenmittels  $\mathbf{Var}(\bar{x}_T) = \frac{2\sigma^2}{T^2}$  ist also von der Größenordnung  $O(T^{-2})$  statt  $O(T^{-1})$ .

*Beweis von Theorem 2.13.*

1. Konvergiert  $\gamma(k) \rightarrow 0$  dann existieren für jedes  $\epsilon > 0$  zwei Schranken  $0 < K < T$ , sodass  $\|\gamma(k)\| < \epsilon/4$  für  $|k| > K$  und  $\frac{1}{T} \sum_{|k| \leq K} \|\gamma(k)\| \leq \epsilon/2$  gilt. Damit folgt

$$\|\mathbf{Var}(\bar{x}_T)\| \leq \frac{1}{T} \sum_{k=-T+1}^{T-1} \|\gamma(k)\| = \frac{1}{T} \sum_{|k| \leq K} \|\gamma(k)\| + \frac{1}{T} \sum_{T > |k| > K} \|\gamma(k)\| \leq \frac{\epsilon}{2} + \frac{2(T-K+1)}{T} \frac{\epsilon}{4} \leq \epsilon$$

<sup>5</sup> Nicht jeder stationäre Prozess ist ergodisch. Betrachten Sie z.B. den stationären Prozess  $(x_t = z \mid t \in \mathbb{Z})$ , wobei  $z$  eine beliebige quadratisch integrierbare Zufallsvariable ist. Das Stichprobenmittel  $\bar{x}_T = \frac{1}{T} \sum_{t=1}^T z = z$  konvergiert *nicht* gegen den Erwartungswert  $\mathbf{E}x_t = \mathbf{E}z$  wenn  $z$  eine positive Varianz  $\mathbf{Var}(z) > 0$  besitzt!

<sup>6</sup> Für das obige Bsp.  $(x_t = z)$  ist diese Bedingung natürlich nicht erfüllt, da  $\gamma(k) = \mathbf{Var}(z) \equiv \text{const.}$

und daher auch  $\text{Var}(\bar{x}_T) \rightarrow 0$  wie gewünscht.

2. Wenn die Autokovarianzen absolut summierbar sind (d.h.  $\sum_k \|\gamma(k)\| < \infty$ ), dann kann man für jedes  $\epsilon > 0$  zwei Schranken  $0 < K < T < \infty$  finden, sodass  $\sum_{|k|>K} \|\gamma(k)\| \leq \epsilon/2$  und  $K/T \sum_{|k|\leq K} \|\gamma(k)\| \leq \epsilon/2$ . Daher folgt

$$\begin{aligned} \|T\text{Var}(\bar{x}_T) - \sum_{k=-\infty}^{\infty} \gamma(k)\| &= \left\| -\sum_{|k|<T} \frac{|k|}{T} \gamma(k) - \sum_{|k|\geq T} \gamma(k) \right\| \\ &\leq \sum_{|k|\leq K} \frac{|k|}{T} \|\gamma(k)\| + \sum_{K<|k|<T} \frac{|k|}{T} \|\gamma(k)\| + \sum_{|k|\geq T} \|\gamma(k)\| \\ &\leq \frac{K}{T} \sum_{|k|\leq K} \|\gamma(k)\| + \sum_{|k|>K} \|\gamma(k)\| \\ &\leq \epsilon \end{aligned}$$

und damit auch  $T\text{Var}(\bar{x}_T) \rightarrow \sum_k \gamma(k)$  für  $T \rightarrow \infty$ .

3. Folgt aus der Gleichung (2.5) für die ACF eines MA( $\infty$ ) Prozesses.
4. Zunächst betrachten wir eine Approximation des Prozesses durch (zweiseitige) MA(2q) Prozesse der Form  $x_t^q = \mu + \sum_{j=-q}^q b_j \epsilon_{t-j}$ . Das empirische Mittel von  $x_t^q$  lässt sich schreiben als

$$\bar{x}_T^q = \frac{1}{T} \sum_{t=1}^T x_t^q = \mu + \sum_{j=-q}^q b_j \bar{\epsilon}_{T,j} = \mu + \sum_{j=-q}^q b_j \bar{\epsilon}_T + \sum_{j=-q}^q b_j u_{T,j}$$

wobei

$$\bar{\epsilon}_{T,j} = \frac{1}{T} \sum_{t=1}^T \epsilon_{t-j} \text{ and } u_{T,j} = \bar{\epsilon}_{T,j} - \bar{\epsilon}_T = \begin{cases} T^{-1} \sum_{k=0}^{j-1} (\epsilon_{-k} - \epsilon_{T-k}) & \text{for } j > 0 \\ 0 & \text{for } j = 0 \\ T^{-1} \sum_{k=1}^{-j} (\epsilon_{T+k} - \epsilon_k) & \text{for } j < 0 \end{cases}$$

Die Varianz von  $\sqrt{T}u_{T,j}$  ist  $T^{-1}2j\Sigma$  und daher konvergiert  $\sqrt{T}u_{T,j}$  für  $T \rightarrow \infty$  gegen Null (im l.i.m Sinne und daher auch in Wahrscheinlichkeit). Mit dem CLT 2.19 folgt nun

$$\sqrt{T}(\bar{x}_T^q - \mu) = B_q \sqrt{T} \bar{\epsilon}_T + \sum_{j=-q}^q b_j \underbrace{\sqrt{T}u_{T,j}}_{\rightarrow 0} \xrightarrow{d} m_q := B_q e \sim N(0, V_q)$$

2.7

wobei  $B_q = \sum_{j=-q}^q b_j$  und  $V_q = B_q \Sigma B_q'$ . Damit haben wir das gewünschte Resultat für MA Prozesses *endlicher* Ordnung schon gezeigt.

Es ist leicht zu zeigen, dass  $V_q \rightarrow V = (\sum_{j=-\infty}^{\infty} b_j) \Sigma (\sum_{j=-\infty}^{\infty} b_j)'$  und damit

$$m_q \xrightarrow{d} m \sim N(0, V) \text{ für } q \rightarrow \infty$$

2.8

Der "Approximationsfehler"  $(x_t - x_t^q) = \sum_{|j|>q} b_j \epsilon_{t-j}$  konvergiert für wachsende Ordnung  $q$  gegen Null. Entsprechend konvergiert auch die Differenz der Stichprobenmittel  $(\bar{x}_T - \bar{x}_T^q)$  gegen Null. Konkret folgt aus den Punkten (2) und (3), dass

$$\lim_{q \rightarrow \infty} \lim_{T \rightarrow \infty} \text{Var}(\sqrt{T}(\bar{x}_T - \bar{x}_T^q)) = \lim_{q \rightarrow \infty} \left( \sum_{|j|>q} b_j \right) \Sigma \left( \sum_{|j|>q} b_j \right)' = 0$$

und aus der Chebychev Ungleichung (für jedes  $\epsilon > 0$ ), dass

$$\lim_{q \rightarrow \infty} \limsup_{T \rightarrow \infty} \mathbf{P}(\|\sqrt{T}(\bar{x}_T - \bar{x}_T^q)\| > \epsilon) = 0$$

2.9

Das Resultat folgt nun mit (2.7), (2.8) und (2.9) aus dem Satz 6.3.9 in BROCKWELL&DAVIS<sup>7</sup>. Die wesentliche Idee des Beweises steckt in der folgenden Skizze

$$\begin{array}{ccc} \sqrt{T}(\bar{x}_T^q - \mu) & \xrightarrow{T \rightarrow \infty} & m_q \\ \downarrow & & \downarrow \\ \sqrt{T}(\bar{x}_T - \mu) & \xrightarrow{T \rightarrow \infty} & m \end{array} \quad \text{für } q \rightarrow \infty$$

□

<sup>7</sup> P.J. Brockwell and R.A. Davis. *Time Series: Theory and Methods*. Springer series in statistics. Springer-Verlag, New York, 2nd edition, 1991

## 2.7 Schätzung der Autokovarianzfunktion

Die Autokovarianzfunktion  $\gamma(\cdot)$  wird mit

$$\hat{\gamma}_T(k) = \begin{cases} \frac{1}{T} \sum_{t=1}^{T-k} (x_{t+k} - \bar{x}_T)(x_t - \bar{x}_T)' & \text{für } 0 \leq k < T \\ 0 & \text{für } T \leq k \\ \hat{\gamma}_T(-k)' & \text{für } k < 0 \end{cases}$$

geschätzt. Die geschätzte Autokovarianzfunktion ist eine positiv semidefinite Folge<sup>8</sup> und damit ein "zulässiger" Schätzer für  $\gamma(\cdot)$ . Mit der geschätzten ACF  $\hat{\gamma}$  erhält man auch sofort einen Schätzer für die Autokorrelationsfunktion  $\rho(\cdot)$

$$\hat{\rho}_{ij,T}(k) = \frac{\hat{\gamma}_{ij,T}(k)}{\sqrt{\hat{\gamma}_{ii,T}(0)\hat{\gamma}_{jj,T}(0)}}$$

Insbesondere setzt man für skalare Prozesse (Zeitreihen)

$$\hat{\rho}_T(k) = \frac{\hat{\gamma}_T(k)}{\hat{\gamma}_T(0)}$$

Die Schätzer für  $\gamma$  (bzw.  $\rho$ ) ändern sich nicht, wenn man  $x_t$  (und entsprechend dann  $\bar{x}_T$ ) durch die Mittelwert-bereinigten Werte  $(x_t - \mu)$  (bzw.  $(\bar{x}_T - \mu)$ ) ersetzt. Daher können wir im Folgenden o.E.d.A annehmen, dass  $(x_t)$  ein Prozess mit Mittel  $\mu = \mathbf{E}x_t = 0$  ist. Für  $T > k \geq 0$  können wir den Schätzer  $\hat{\gamma}_T(k)$  umformen zu

$$\hat{\gamma}_T(k) = \frac{1}{T} \sum_{t=1}^T x_t x'_{t-k} + E_{T,k}$$

wobei

$$E_{T,k} = -\frac{1}{T} \sum_{t=1}^k x_t x'_{t-k} - \bar{x}_T (\bar{x}_T - u_{T,k}^+)' - (\bar{x}_T - u_{T,k}^-) \bar{x}_T' + \frac{T-k}{T} \bar{x}_T \bar{x}_T'$$

$$u_{T,k}^- = \frac{1}{T} \sum_{t=1}^k x_t \quad \text{and} \quad u_{T,k}^+ = \frac{1}{T} \sum_{t=T-k+1}^T x_t$$

Der Erwartungswert von  $\hat{\gamma}_T(k)$  ist

$$\mathbf{E}\hat{\gamma}_T(k) = \frac{1}{T} \sum_{t=1}^T \mathbf{E}x_t x'_{t-k} + \mathbf{E}E_{T,k} = \gamma(k) + \mathbf{E}E_{T,k}$$

und wir sehen, dass die Stichproben-Kovarianzfunktion im allgemeinen kein unverzerrter Schätzer ist. Aber sie ist zumindest ein asymptotisch unverzerrter Schätzer, wenn das Stichproben-Mittel  $\bar{x}_T$  konsistent (im quadratischen Mittel) ist:

<sup>8</sup> Ersetzt man in der Definition den Faktor  $1/T$  durch  $1/(T-K)$  dann ist die geschätzte ACF im allgemeinen nicht mehr positiv semidefinit.

**Satz 2.14.** Gilt  $\text{l.i.m}_{T \rightarrow \infty} \bar{x}_T = \mu = \mathbf{E}x_t$ , dann folgt  $\lim_{T \rightarrow \infty} \mathbf{E}\hat{\gamma}_T(k) = \gamma(k)$ .

*Beweis.* Wie oben angemerkt, nehmen wir an, dass  $\mathbf{E}x_t = 0$ . Aus  $\text{l.i.m } \bar{x}_T = 0$  folgt  $\lim \mathbf{E}\bar{x}_T \bar{x}_T' = \mathbf{E}(\text{l.i.m } \bar{x}_T)(\text{l.i.m } \bar{x}_T)' = 0$ . Es ist auch leicht zu sehen, dass  $\text{l.i.m } u_{T,k}^+ = \text{l.i.m } u_{T,k}^- = 0$  und damit auch  $\lim \mathbf{E}\bar{x}_T(u_{T,k}^+)' = \lim \mathbf{E}u_{T,k}^- \bar{x}_T' = 0$ . Zusammen mit  $\mathbf{E}\frac{1}{T} \sum_{t=1}^k x_t x_{t-k}' = \frac{k}{T} \gamma(k) \rightarrow 0$  folgt nun, dass  $\mathbf{E}E_{T,k} \rightarrow 0$ .  $\square$

Das nächste Theorem (siehe BROCKWELL&DAVIS<sup>9</sup>, Proposition 7.3.4, Theorem 7.2.1 und 7.2.2) zeigt, dass die Stichproben- Autokovarianzfunktion und Autokorrelationsfunktion unter recht allgemeinen Voraussetzungen konsistent und asymptotisch normalverteilt sind<sup>10</sup>.

**Theorem 2.15.** Sei  $(x_t)$  ein skalarer, stationärer Prozess der Form  $x_t = \mu + \sum_{j=-\infty}^{\infty} b_j \epsilon_{t-j}$ , wobei  $\sum_j |b_j| < \infty$ ,  $(\epsilon_t) \sim \text{IID}(0, \sigma^2)$  und  $\mathbf{E}\epsilon_t^4 = \kappa \sigma^4 < \infty$ . Dann sind die Schätzer  $\hat{\gamma}_T(\cdot)$  und  $\hat{\rho}_T(\cdot)$  (schwach) konsistent und asymptotisch normalverteilt:

$$\sqrt{T} \begin{pmatrix} \hat{\gamma}_T(0) - \gamma(0) \\ \vdots \\ \hat{\gamma}_T(k) - \gamma(k) \end{pmatrix} \xrightarrow{d} N(0, V) \text{ und } \sqrt{T} \begin{pmatrix} \hat{\rho}_T(1) - \rho(1) \\ \vdots \\ \hat{\rho}_T(k) - \rho(k) \end{pmatrix} \xrightarrow{d} N(0, W)$$

wobei die  $(i, j)$ -ten Elemente von  $V$  und  $W$  gegeben sind durch

$$v_{ij} = (\kappa - 3)\gamma(i)\gamma(j) + \sum_{k=-\infty}^{\infty} (\gamma(k)\gamma(k-i+j) + \gamma(k+j)\gamma(k-i))$$

$$w_{ij} = \sum_{k=-\infty}^{\infty} \left( \rho(k+i)\rho(k+j) + \rho(k-i)\rho(k+j) + 2\rho(i)\rho(j)\rho^2(k) - 2\rho(i)\rho(k)\rho(k+j) - 2\rho(j)\rho(k)\rho(k+i) \right)$$

Die Aussagen für die Autokorrelationsfunktion bleiben auch ohne die Annahme  $\mathbf{E}\epsilon_t^4 < \infty$  erhalten, wenn man dafür zusätzlich  $\sum_{j=-\infty}^{\infty} |j|b_j^2 < \infty$  verlangt.

Für den Spezialfall eines IID Prozesses ( $x_t = \mu + \epsilon_t$ ) vereinfacht sich die asymptotische Verteilung der Stichproben Autokorrelationsfunktion zu

$$\sqrt{T}(\hat{\rho}_T(1), \dots, \hat{\rho}_T(k))' \xrightarrow{d} N(0, I_k) \quad \boxed{2.10}$$

Die Standard Tests auf weißes Rauschen sind der *Box-Pierce Test*

$$Q_{BP} = T \sum_{k=1}^p \hat{\rho}_T^2(k) \quad \boxed{2.11}$$

und der *Ljung-Box Test*

$$Q_{LB} = T(T+2) \sum_{k=1}^p \frac{1}{T-k} \hat{\rho}_T^2(k) \quad \boxed{2.12}$$

Aus (2.10) folgt, dass beide Teststatistiken – unter der Nullhypothese, dass die beobachtete Zeitreihe von einem IID Prozess erzeugt wurde – gegen eine  $\chi_p^2$  Verteilung konvergieren

<sup>9</sup> P.J. Brockwell and R.A. Davis. *Time Series: Theory and Methods*. Springer series in statistics. Springer-Verlag, New York, 2nd edition, 1991

<sup>10</sup> Das Theorem ist für den skalaren Fall formuliert, aber analoge Resultate gelten auch für multivariate Prozesse.

Für endliche Stichproben verhält sich die LB-Teststatistik etwas besser, d.h. die Approximation durch eine Chi-Quadrat Verteilung ist etwas besser.

$$Q_{BP} \xrightarrow{d} \chi_p^2 \text{ und } Q_{LB} \xrightarrow{d} \chi_p^2$$

Der Beweis von Theorem 2.15 ist etwas mühsam, aber zumindest eine Skizze des Beweises für den Fall eines skalaren IID Prozesses ( $x_t = \epsilon_t$ )  $\sim$  IID( $\sigma^2$ ) sei hier angeführt.

1.  $\sqrt{T}E_{T,k} \xrightarrow{p} 0$ . Es genügt daher die asymptotischen Eigenschaften von  $\frac{1}{\sqrt{T}} \sum_{t=1}^T \epsilon_t \epsilon_{t-k}$  zu untersuchen. Genauer gesagt gilt  $\sqrt{T}(\hat{\rho}(s) - \tilde{\rho}_T(s)) \rightarrow 0$ , wobei  $\tilde{\rho}(s) = (\frac{1}{T} \sum_{t=1}^T \epsilon_t \epsilon_{t-s})(\frac{1}{T} \sum_{t=1}^T \epsilon_t^2)^{-1}$ . Daher haben  $\sqrt{T}(\hat{\rho}_T(1), \dots, \hat{\rho}_T(k))'$  und  $\sqrt{T}(\tilde{\rho}_T(1), \dots, \tilde{\rho}_T(k))'$  dieselbe asymptotische Verteilung.
2.  $\frac{1}{T} \sum_{t=1}^T \epsilon_t^2 \rightarrow \sigma^2$  da  $(\epsilon_t) \sim$  IID( $\sigma^2$ ), siehe Satz 2.17.
3.  $z_t = (\epsilon_t \epsilon_{t-1}, \epsilon_t \epsilon_{t-2}, \dots, \epsilon_t \epsilon_{t-k})'$  ist eine strikt stationäre, ergodische Martingaldifferenz.  $(z_t)$  ist eine Transformation des IID Prozesses  $(\epsilon_t)$  und ist daher strikt stationär und ergodisch. Es gilt

$$\mathbf{E}(\epsilon_t \epsilon_{t-j} | \epsilon_{t-1}, \dots, \epsilon_{1-k}) = \epsilon_{t-j} \mathbf{E}(\epsilon_t) = 0$$

und daher mit der Turmeigenschaft des bedingten Erwartungswertes

$$\mathbf{E}(z_t | z_{t-1}, \dots, z_1) = 0$$

Aus Theorem 2.19 folgt nun

$$\frac{1}{\sqrt{T}} \sum_{t=1}^T z_t \xrightarrow{d} z \sim N(0, V)$$

wobei  $V = \mathbf{Var}(z_t) = (\sigma^2)^2 I_k$ , da

$$\mathbf{Cov}(\epsilon_t \epsilon_{t-j}, \epsilon_t \epsilon_{t-l}) = \mathbf{E} \epsilon_t^2 \epsilon_{t-j} \epsilon_{t-l} = \begin{cases} 0 & \text{für } j \neq l \\ (\sigma^2)^2 & \text{für } j = l \end{cases}$$

4. Schließlich folgt nun

$$\sqrt{T}(\tilde{\rho}_T(1), \dots, \tilde{\rho}_T(k))' = \underbrace{\left(\frac{1}{T} \sum_{t=1}^T \epsilon_t^2\right)^{-1}}_{\rightarrow \sigma^2} \underbrace{\frac{1}{\sqrt{T}} \sum_{t=1}^T z_t}_{\rightarrow z} \xrightarrow{d} \sigma^{-2} z \sim N(0, I_k)$$

## 2.8 Anhang

Zusammenstellung von wichtigen Begriffen und Resultaten der Asymptotik. Siehe z.B. HAYASHI<sup>11</sup>.

**Definition 2.16.** Einen strikt stationären Prozess  $(x_t)$  nennt man *ergodisch*, wenn für beliebige reell-wertige, beschränkte Funktionen  $f$  und  $g$  folgendes gilt:

$$\lim_{n \rightarrow \infty} |\mathbf{E}[f(x_t, \dots, x_{t+k})g(x_{t+n}, \dots, x_{t+n+l})]| = |\mathbf{E}[f(x_t, \dots, x_{t+k})]| |\mathbf{E}[g(x_t, \dots, x_{t+l})]|$$

$\sqrt{T}E_{T,k} \xrightarrow{p} 0$  gilt auch für den allgemeinen Fall.

<sup>11</sup> Fumio Hayashi. *Econometrics*. Princeton Univ. Press, 2000



Diese Bedingung besagt, dass die Zufallsvariablen/Zufallsvektoren  $(x_t, \dots, x_{t+k})$  und  $(x_{t+n}, \dots, x_{t+n+l})$  "fast" unabhängig sind, wenn nur der "Abstand"  $n$  groß genug ist. Ergodische Prozesse sind aufgrund des folgenden Theorems sehr wichtig für die Statistik stochastischer Prozesse.

**Theorem 2.17.** Sei  $(x_t)$  ein ergodischer Prozess mit Mittel  $\mathbf{E}x_t = \mu$ , dann gilt

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T x_t = \mu \text{ f.s.}$$

Jede Transformation eines ergodischen Prozesses mit einer messbaren Funktion  $h(\cdot)$  liefert wieder einen ergodischen Prozess  $(y_t = h(x_t, x_{t-1}, \dots, x_{t-k}))$ . Daher können z.B. auch höhere Momente (wenn sie existieren) konsistent geschätzt werden. Für einen ergodischen Prozess mit  $\mathbf{E}x_t x'_t = \Sigma < \infty$  folgt z.B. unmittelbar

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T x_t x'_t = \Sigma \text{ f.s.}$$

Ist der Prozess  $(x_t)$  strikt stationär, dann ist auch der transformierte Prozess strikt stationär.

**Definition 2.18.** Ein Prozess  $(x_t \mid t \in \mathbb{N})$  ist eine Martingaldifferenz, wenn  $x_t$  integrierbar ist,  $\mathbf{E}x_t = 0$  für alle  $t$  und

Für eine Martingaldifferenz gilt  $\mathbf{E}x_t x'_s = 0$  für alle  $t \neq s$ .

$$\mathbf{E}(x_t \mid x_{t-1}, x_{t-2}, \dots, x_1) = 0 \text{ für } t > 1$$

Ist der Prozess  $(x_t)$  stationär, ergodisch und eine Martingaldifferenz, dann gilt folgender zentraler Grenzwertsatz.

**Theorem 2.19.** Sei  $(x_t \mid t \in \mathbb{N})$  eine strikt stationäre, ergodische Martingaldifferenz mit beschränkten zweiten Momenten  $\mathbf{E}x_t x'_t = \Sigma$ , dann gilt

$$\sqrt{T} \left[ \frac{1}{T} \sum_{t=1}^T x_t \right] \xrightarrow{d} N(0, \Sigma)$$

Jeder IID Prozess ist ergodisch. Daher gilt Theorem 2.17 insbesondere für integrierbare IID Prozesse. Ist  $(x_t \mid t \in \mathbb{N})$  ein integrierbarer IID Prozess, dann ist  $(x_t - \mu)$  eine Martingaldifferenz, wobei  $\mathbf{E}x_t = \mu$ . Wenn man zusätzlich noch verlangt, dass die  $x_t$ 's quadratisch integrierbar sind, also  $\mathbf{Var}(x_t) = \Sigma < \infty$  dann folgt aus Theorem 2.19 der klassische Grenzwertsatz für IID Prozesse:

$$\sqrt{T}(\bar{x}_T - \mu) \xrightarrow{d} N(0, \Sigma)$$



# 3

## Prognose

In diesem Abschnitt beschäftigen wir uns mit der Prognose, d.h. heißt wir wollen Vorhersagen über die Werte des Prozesses zu bestimmten Zeitpunkten in der Zukunft machen. Wir nehmen an, dass wir die Werte des Prozesses bis zum gegenwärtigen Zeitpunkt  $t$  kennen, und machen eine Prognose für  $x_{t+h}$  von der Form

$$\hat{x}_{t,h} = g(x_t, x_{t-1}, \dots),$$

d.h. wir approximieren den unbekannten, zukünftigen Wert  $x_{t+h}$  durch eine Funktion der beobachteten Werte bis zur Gegenwart  $t$ . Die Güte der Prognose messen wir mit dem *mittleren, quadratischen Prognosefehler* (MSE), d.h. mit

$$\mathbf{E}(x_{t+h} - \hat{x}_{t,h})'(x_{t+h} - \hat{x}_{t,h})$$

Wir beschränken uns auf *affine* Funktionen, d.h. auf Prognosen der Form

$$\hat{x}_{t,h} = g(x_t, x_{t-1}, \dots) = c_0 + c_1 x_t + c_2 x_{t-1} + \dots$$

Zunächst behandeln wir ein idealisiertes Problem, bei dem wir annehmen, dass wir die Autokovarianzfunktion und den Mittelwert des Prozesses exakt kennen. Wie man bei einer realistischen Problemstellung, bei der man z.B. nur eine Zeitreihe zur Verfügung hat, vorgehen kann, werden wir am Ende des Kapitels kurz diskutieren.

Das Prognoseproblem lässt sich besonders elegant behandeln, wenn man es als ein Approximations-Problem im Hilbertraum  $\mathbb{L}_2(\Omega, \mathcal{A}, P)$  der quadratisch integrierbaren Zufallsvariablen interpretiert. Im ersten Abschnitt wiederholen wir daher einige wichtige Resultate über Hilberträume (insbesondere den  $\mathbb{L}_2$ ) und definieren den sogenannten *Zeitbereich* eines stationären Prozesses.

### 3.1 Der Zeitbereich stationärer Prozesse

**Definition 3.1.** Ein reeller (komplexer) Hilbertraum ist ein reeller (komplexer) Vektorraum mit einem inneren Produkt, der bezüglich der vom inneren Produkt erzeugten Norm vollständig ist. Vollständigkeit bedeutet, dass eine Folge dann und nur dann konvergiert, wenn sie eine Cauchy-Folge ist.

Naive Prognose, einfache exponentielle Glättung, Prognosegüte, (empirische) Auswertung/Vergleich von Prognoseverfahren, (MSE) optimale lineare Prognose aus der endlichen, bzw. unendlichen Vergangenheit, Prognosegleichung, Wold-Zerlegung von stationären Prozessen, singuläre und reguläre Prozesse, komplexe/reelle Hilberträume, der  $\mathbb{L}_2$  (komplex, bzw. reell), Projektionssatz, Interpretation von stationären Prozessen im  $\mathbb{L}_2$ , der Zeitbereich eines stationären Prozesses

Für das innere Produkt  $\langle x, y \rangle$  und die dadurch erzeugte Norm  $\|x\|^2 = \langle x, x \rangle$  gelten folgende nützliche Beziehungen

$$\begin{aligned} \text{Cauchy-Schwarz} \quad & |\langle x, y \rangle| \leq \|x\| \|y\| \\ \text{Dreiecksungleichung} \quad & \|x + y\| \leq \|x\| + \|y\| \\ & \|x + y\| \geq \|x\| - \|y\| \\ \text{Pythagoras} \quad & \|x + y\|^2 = \|x\|^2 + \|y\|^2 \quad \text{wenn } \langle x, y \rangle = 0 \end{aligned}$$

**Theorem 3.2** (Projektionssatz). Sei  $\mathbb{H}$  ein Hilbertraum und  $\mathbb{M} \subseteq \mathbb{H}$  ein abgeschlossener Unterraum.  $\mathbb{M}$  ist also selbst wieder ein Hilbertraum. Für jedes Element  $x \in \mathbb{H}$  existiert eine eindeutige Zerlegung

$$x = \hat{x} + \hat{u}$$

wobei  $\hat{x} \in \mathbb{M}$  und  $\hat{u} \perp \mathbb{M}$ .  $\hat{x}$  nennt man die orthogonale Projektion von  $x$  auf  $\mathbb{M}$ .  $\hat{x}$  ist auch die (eindeutige) beste Approximation von  $x$  durch ein Element des Unterraums  $\mathbb{M}$ , d.h.

$$\|x - \hat{x}\| < \|x - \tilde{x}\| \quad \text{für alle } \tilde{x} \in \mathbb{M}, \tilde{x} \neq \hat{x}.$$

Die Abbildung, die  $x \in \mathbb{H}$  die orthogonale Projektion  $\hat{x} \in \mathbb{M}$  zuordnet, ist eine lineare Abbildung von  $\mathbb{H}$  nach  $\mathbb{M}$ . Diese Projektionsabbildung (Projektionsoperator) wird oft mit  $\mathbf{P}_{\mathbb{M}}$  oder auch kurz mit  $\mathbf{P}$  bezeichnet.

Sind  $\mathbb{M}_1 \subseteq \mathbb{H}$  und  $\mathbb{M}_2 \subseteq \mathbb{H}$  zwei zueinander orthogonale Hilberträume ( $\mathbb{M}_1 \perp \mathbb{M}_2$ ) und  $\mathbb{M} = \mathbb{M}_1 \oplus \mathbb{M}_2$ , dann folgt

$$\mathbf{P}_{\mathbb{M}} = \mathbf{P}_{\mathbb{M}_1} + \mathbf{P}_{\mathbb{M}_2}$$

Sei  $(\Omega, \mathcal{A}, P)$  ein Wahrscheinlichkeitsraum. Der Hilbertraum  $\mathbb{L}_2(\Omega, \mathcal{A}, P)$  der quadratisch integrierbaren Zufallsvariablen  $x : (\Omega, \mathcal{A}, P) \mapsto \mathbb{R}$  wird folgendermaßen konstruiert. Die Menge aller quadratisch integrierbaren Zufallsvariablen ( $\mathbb{E}x^2 < \infty$ ) bildet einen reellen Vektorraum. Man definiert ein "inneres Produkt" mit  $\langle x, y \rangle = \mathbb{E}xy$ . Dieses Produkt erfüllt fast alle Axiome eines inneren Produkts. Nur folgt aus  $\langle x, x \rangle = \mathbb{E}x^2 = 0$  nicht  $x = 0$ , sondern nur  $x = 0$  f.s. Daher betrachtet man Äquivalenzklassen von Zufallsvariablen, die bis auf eine Nullmenge übereinstimmen, also Mengen der Form  $[x] := \{\tilde{x} \mid \tilde{x} = x \text{ f.s.}\}$ . Die Menge aller Äquivalenzklassen ist ein reeller Vektorraum und die Abbildung  $\langle [x], [y] \rangle = \mathbb{E}xy$  ist ein inneres Produkt. Die Vollständigkeit folgt aus dem Riesz-Fisher Theorem 2.7.

Im Folgenden werden wir (etwas schlampig) die Äquivalenzklassen mit einem (beliebigen) Element identifizieren und daher z.B. einfach  $x \in \mathbb{L}_2$  schreiben.

Die Norm-Konvergenz im  $\mathbb{L}_2$  entspricht der Konvergenz im quadratischem Mittel, d.h.  $\|x_n - x\| \rightarrow 0$  ist gleichbedeutend mit l.i.m  $x_n = x$ .

Viele statistische Begriffe/Konzepte haben eine Entsprechung im  $\mathbb{L}_2$  und bekommen damit eine "geometrische" Interpretation. Besonders einfach ist diese "Übersetzung" für Zufallsvariable mit Erwartungswert gleich Null:

Notation:

$x \perp y$  heißt  $\langle x, y \rangle = 0$

$x \perp \mathbb{M} \subseteq \mathbb{H}$  heißt  $\langle x, y \rangle = 0$  für alle  $y \in \mathbb{M}$

$\mathbb{N} \perp \mathbb{M}$  für  $\mathbb{M} \subseteq \mathbb{H}$  und  $\mathbb{N} \subseteq \mathbb{H}$  heißt  $\langle x, y \rangle = 0$  für alle  $x \in \mathbb{N}$  und  $y \in \mathbb{M}$

Wesentlich ist, dass Linearkombinationen von quadratisch integrierbaren Zufallsvariablen selbst wieder quadratisch integrierbar ist. D.h.  $\mathbb{E}(\alpha x + \beta y)^2 < \infty$  für alle quadratisch integrierbaren Zufallsvariablen  $x, y$  und reelle Zahlen  $\alpha, \beta \in \mathbb{R}$ .

- Die Norm von  $x$  ist gleich der Standardabweichung:  $\|x\| = \sqrt{\mathbf{Var}(x)}$
- Das innere Produkt ist gleich der Kovarianz:  $\langle x, y \rangle = \mathbf{Cov}(x, y)$
- Die Korrelation ist der Cosinus des "Winkels zwischen den beiden Zufallsvariablen":

$$\mathbf{Corr}(x, y) = \frac{\mathbf{Cov}(x, y)}{\sqrt{\mathbf{Var}(x)\mathbf{Var}(y)}} = \frac{\langle x, y \rangle}{\|x\|\|y\|} = \cos(\angle(x, y))$$

- Unkorreliert bedeutet orthogonal im  $\mathbb{L}_2$ :  $\mathbf{Cov}(x, y) = 0$  ist äquivalent zu  $x \perp y$
- Die Zufallsvariablen  $x_1, \dots, x_k \in \mathbb{L}_2$  sind dann und nur dann linear unabhängig im  $\mathbb{L}_2$ , wenn die Varianz-Kovarianzmatrix  $\mathbf{Var}((x_1, \dots, x_k)')$  positiv definit ist.

Entsprechend lassen sich auch Zufallsvariable mit einem Erwartungswert ungleich Null behandeln. Insbesondere gilt  $\mathbf{E}x = \mathbf{E}1x = \langle 1, x \rangle$  und  $\mathbf{E}x = 0$  ist äquivalent zu  $x \perp 1$ . Hier steht  $1 \in \mathbb{L}_2$  für eine Zufallsvariable, die nur den Wert  $1 \in \mathbb{R}$  annimmt.

Sei  $(x_t | t \in \mathbb{Z})$  ein stationärer Prozess. Die Zufallsvektoren  $x_t$  sind auf einem (gemeinsamen) Wahrscheinlichkeitsraum  $(\Omega, \mathcal{A}, P)$  definiert. Wir betrachten daher den entsprechenden Hilbertraum  $\mathbb{L}_2(\Omega, \mathcal{A}, P)$ . Die Zufallsvariablen  $x_{it}$  sind quadratisch integrierbar und daher Elemente dieses Hilbertraum  $\mathbb{L}_2(\Omega, \mathcal{A}, P)$ , d.h.  $x_{it} \in \mathbb{L}_2(\Omega, \mathcal{A}, P)$  für alle  $i = 1, \dots, n$  und  $t \in \mathbb{Z}$ .

Der Zeitbereich von  $(x_t)$  ist der von den  $x_{it}$ 's erzeugte (Teil-) Hilbertraum:

$$\mathbb{H}_x := \overline{\text{span}}\{x_{it} | i = 1, \dots, n, t \in \mathbb{Z}\} \subseteq \mathbb{L}_2(\Omega, \mathcal{A}, P)$$

Wir benötigen auch manchmal den Hilbertraum der komplexwertigen, quadratisch integrierbaren Zufallsvariablen  $\mathbb{L}_2^c(\Omega, \mathcal{A}, P)$  der analog konstruiert wird mit dem inneren Produkt  $\langle x, y \rangle = \mathbf{E}x\bar{y}$ .

**ZUFALLSVEKTOREN:** Im Folgenden werden wir uns oft mit Zufallsvektoren befassen, deren Komponenten Elemente von  $\mathbb{L}_2$  sind, insbesondere natürlich die  $x_t$ 's. Daher führen wir folgende Konventionen ein. Seien  $u, v, \dots$  Zufallsvektoren die auf dem Wahrscheinlichkeitsraum  $(\Omega, \mathcal{A}, P)$  definiert sind und  $\mathbb{L}_2 = \mathbb{L}_2(\Omega, \mathcal{A}, P)$  der entsprechende Hilbertraum der quadratisch integrierbaren Zufallsvariablen.

- $u = (u_1, \dots, u_p)' \in \mathbb{M} \subseteq \mathbb{L}_2$  heißt  $u_i \in \mathbb{M}$  für  $i = 1, \dots, p$
- $u = (u_1, \dots, u_p)' \perp v = (v_1, \dots, v_q)'$  heißt  $u_i \perp v_j$  für alle  $i = 1, \dots, p$  und  $j = 1, \dots, q$ . Diese Bedingung ist äquivalent zu  $\mathbf{E}uv' = 0 \in \mathbb{R}^{p \times q}$ .
- $u \perp \mathbb{M} \subseteq \mathbb{L}_2$  heißt  $u_i \perp \mathbb{M}$  für  $i = 1, \dots, p$

- Die Projektion von  $u = (u_1, \dots, u_p)$  auf einen (Teil-) Hilbertraum  $M \subseteq \mathbb{L}_2$  definiert man Komponenten-weise

$$\mathbf{P}_M u = \begin{pmatrix} \mathbf{P}_M u_1 \\ \mathbf{P}_M u_2 \\ \vdots \\ \mathbf{P}_M u_p \end{pmatrix} \quad \boxed{3.1}$$

Aufgrund der Linearität der Projektion gilt für jede Matrix  $A \in \mathbb{R}^{q \times p}$

$$\mathbf{P}_M(Au) = A(\mathbf{P}_M u)$$

- Für Unterräume der Form  $\text{span}\{x_{it} \mid i = 1, \dots, n; t \in I \subseteq \mathbb{Z}\}$  schreiben wir einfach  $\text{span}\{x_t \mid t \in I \subseteq \mathbb{Z}\}$ . Also z.B.

$$\mathbb{H}_x = \overline{\text{span}}\{x_t \mid t \in \mathbb{Z}\}$$

### 3.2 Prognose aus der endlichen Vergangenheit

Wir wollen nun die optimale  $h$ -Schrittprognose<sup>1</sup> aus der endlichen Vergangenheit konstruieren. Das heißt, wir haben folgendes Optimierungsproblem zu lösen:

<sup>1</sup>  $h$  ist der Prognosehorizont

$$\begin{aligned} \mathbf{E}(x_{t+h} - \check{x})'(x_{t+h} - \check{x}) &\longrightarrow \min \\ \check{x} &= c_0 + c_1 x_t + \dots + c_k x_{t+1-k} \end{aligned}$$

Zunächst zeigen wir, dass der Erwartungswert des (optimalen) Prognosefehlers gleich Null sein muss. Sei  $\check{x} = c_0 + c_1 x_t + \dots + c_k x_{t+1-k}$  eine beliebige, nicht unbedingt optimale Prognose und  $\check{u} = x_{t+h} - \check{x}$  der entsprechende Prognosefehler. Der MSE der Prognose  $\check{x} = \check{x} + \mathbf{E}\check{u}$  (mit Fehler  $\check{u} = x_{t+h} - \check{x} = \check{u} - \mathbf{E}\check{u}$ ) ist gleich

$$\mathbf{E}(\check{u}'\check{u}) = \text{tr}(\mathbf{E}\check{u}\check{u}') = \text{tr}(\mathbf{Var}(\check{u})) \leq \text{tr}(\mathbf{Var}(\check{u}) + \mathbf{E}\check{u}\mathbf{E}\check{u}') = \text{tr}(\mathbf{E}\check{u}\check{u}') = \mathbf{E}\check{u}'\check{u}$$

Daher ist die Prognose  $\check{\check{x}} = \check{x} + \mathbf{E}\check{u}$  strikt besser, wenn  $\mathbf{E}\check{u} \neq 0$  gilt.

Sei nun  $\hat{x}_{t,h,k} = c_0 + c_1 x_t + \dots + c_k x_{t+1-k}$  die optimale Prognose für  $x_{t+h}$  und  $\hat{u}_{t,h,k} = x_{t+h} - \hat{x}_{t,h,k}$  der entsprechende Prognosefehler. Aus der Bedingung  $\mathbf{E}\hat{u}_{t,h,k} = 0$  folgt mit  $\mu = \mathbf{E}x_s$

$$c_0 = (I - c_1 - \dots - c_k)\mu$$

$$\hat{x}_{t,h,k} = \mu + c_1(x_t - \mu) + \dots + c_k(x_{t+1-k} - \mu)$$

Für  $\mu = 0$  ist die optimale Prognose also *linear* (d.h.  $c_0 = 0$ ). Man kann sich auch leicht überzeugen, dass  $(\hat{x}_{t,h,k} - \mu) = c_1(x_t - \mu) + \dots + c_k(x_{t+1-k} - \mu)$  die optimale Prognose für den Mittelwert bereinigten Prozess  $(x_t - \mu)$  ist. Daher kann die Prognose für den Fall  $\mu \neq 0$  auch folgendermaßen berechnet werden. Zunächst bestimmt man die optimale (lineare) Prognose für den Mittelwert bereinigten Prozess und addiert dann einfach noch den Erwartungswert  $\mu$ .

Im Folgenden nehmen wir daher o.E.d.A an, dass der betrachtete Prozess Erwartungswert gleich Null hat. Wir können uns dann auch immer auf lineare Prognosen ( $c_0 = 0$ ) beschränken.

Das Prognoseproblem lautet nun

$$\begin{aligned} E(x_{t+h} - \check{x})'(x_{t+h} - \check{x}) &\longrightarrow \min \\ \check{x} &= c_1 x_t + \dots + c_k x_{t+1-k} \end{aligned}$$

Dieses Problem lässt sich in  $n$  unabhängige Probleme zerlegen:

$$\begin{aligned} E(x_{i,t+h} - \check{x}_i)^2 &\longrightarrow \min \\ \check{x}_i &= c_{i1} x_t + \dots + c_{ik} x_{t+1-k} \end{aligned}$$

wobei  $c_{ij} \in \mathbb{R}^{1 \times n}$  die  $i$ -te Zeile von  $c_j \in \mathbb{R}^{n \times n}$  bezeichnet. Äquivalent dazu ist folgendes Problem im  $\mathbb{L}_2(\Omega, \mathcal{A}, P)$

$$\begin{aligned} \|x_{i,t+h} - \check{x}_i\| &\longrightarrow \min \\ \check{x}_i &\in \text{span}\{x_t, \dots, x_{t+1-k}\} =: \mathbb{M} \subset \mathbb{L}_2 \end{aligned}$$

Die Lösung folgt unmittelbar aus dem Projektionssatz 3.2

$$\hat{x}_{i,t,h,k} = \mathbf{P}_{\mathbb{M}} x_{i,t+h} \quad \boxed{3.2}$$

Das heißt, die optimale Prognose für  $x_{i,t+h}$  ist die Projektion von  $x_{i,t+h}$  auf den Unterraum  $\mathbb{M}$ . Um die Projektion bzw. Prognose zu berechnen benötigen wir noch die Koeffizienten  $c_{i1}, \dots, c_{ik}$ . Laut Projektionssatz ist der Fehler  $\hat{u}_{i,t,h,k} = x_{i,t+h} - \hat{x}_{i,t,h,k}$  orthogonal auf den Unterraum  $\mathbb{M}$ , d.h.

$$\begin{aligned} 0 &= \langle \hat{u}_{i,t,h,k}, x_{j,t+1-l} \rangle = \mathbf{Cov}(\hat{u}_{i,t,h,k}, x_{j,t+1-l}) \\ &\text{für alle } j = 1, \dots, n \text{ und } l = 1, \dots, k \end{aligned}$$

Mithilfe des gestapelten Vektors  $x_t^k = (x_t', x_{t-1}', \dots, x_{t+1-k}')'$  kann man diese Gleichungen (für alle  $i = 1, \dots, n$ ) in folgendes Gleichungssystem zusammenfassen:

$$\mathbf{Cov}(\hat{u}_{t,h,k}, x_t^k) = 0$$

bzw. mit  $\hat{u}_{t,h,k} = x_{t+h} - \hat{x}_{t,h,k} = x_{t+h} - (c_1, \dots, c_k) x_t^k$

$$\mathbf{Cov}(x_{t+h}, x_t^k) = (c_1, \dots, c_k) \mathbf{Var}(x_t^k)$$

Schließlich erhalten wir folgende "Prognosegleichungen" zur Bestimmung der Koeffizienten  $c_1, \dots, c_k$ :

$$(\gamma(h), \dots, \gamma(h+k-1)) = (c_1, \dots, c_k) \Gamma_k \quad \boxed{3.3}$$

Die Block-Töplitz Matrix  $\Gamma_k = \mathbf{Var}(x_t^k)$  ist die Varianz-Kovarianzmatrix von  $x_t^k$ , siehe auch Gleichung (2.3). Die Varianz des Prognosefehlers kann folgendermaßen bestimmt werden. Da  $\hat{x}_{j,t,h,k} \in \mathbb{M}$  folgt  $\langle \hat{u}_{i,t,h,k}, \hat{x}_{j,t,h,k} \rangle = \mathbf{Cov}(\hat{u}_{i,t,h,k}, \hat{x}_{j,t,h,k}) = 0$  und daher  $\mathbf{Var}(x_{t+h}) = \mathbf{Var}(\hat{x}_{t,h,k} + \hat{u}_{t,h,k}) = \mathbf{Var}(\hat{x}_{t,h,k}) + \mathbf{Var}(\hat{u}_{t,h,k})$ . Somit haben wir

$$\begin{aligned} \Sigma_{h,k} &:= E \hat{u}_{t,h,k} \hat{u}_{t,h,k}' = \mathbf{Var}(x_{t+h}) - \mathbf{Var}(\hat{x}_{t,h,k}) = \\ &= \gamma(0) - (c_1, \dots, c_k) \Gamma_k (c_1, \dots, c_k)' \quad \boxed{3.4} \end{aligned}$$

Die Koeffizienten hängen von  $h$  und  $k$ , aber nicht von  $t$ , ab. Der Unterraum  $\mathbb{M}$  hängt von  $t$  und  $k$  ab. Diese Abhängigkeit wird der Einfachheit halber *nicht* durch eine entsprechende Notation gekennzeichnet.

Aus dem Projektionssatz können wir folgende Schlüsse ziehen: Die (optimale) Prognose  $\hat{x}_{t,h,k}$  und damit auch der entsprechende Prognosefehler  $\hat{u}_{t,h,k}$  (und dessen Varianz  $\Sigma_{h,k}$ ) sind (fast sicher) eindeutig. Die Prognosegleichungen (3.3) sind immer lösbar, auch wenn die Matrix  $\Gamma_k$  singular ist.

Wenn  $\Gamma_k > 0$  positiv definit ist, dann hat (3.3) eine eindeutige Lösung

$$(c_1, \dots, c_k) = (\gamma(h), \dots, \gamma(h+k-1))\Gamma_k^{-1} \quad 3.5$$

und die Prognosefehler-Varianz lässt sich berechnen mit

$$\Sigma_{h,k} = \gamma(0) - (\gamma(h), \dots, \gamma(h+k-1))\Gamma_k^{-1}(\gamma(h), \dots, \gamma(h+k-1))' \quad 3.6$$

Falls  $\Gamma_k$  singular ist, dann existieren unendlich viele Lösungen. Die Komponenten  $\{x_{1t}, \dots, x_{nt}, x_{1,t-1}, \dots, x_{n,t+1-k}\}$  sind linear abhängig und daher ist  $\{x_{1t}, \dots, x_{nt}, x_{1,t-1}, \dots, x_{n,t+1-k}\}$  keine Basis für  $\mathbb{M}$ .

Die Projektion eines Zufallsvektors definiert man Komponentenweise, siehe (3.1). Die optimale Prognose ist damit einfach

$$\hat{x}_{t,h} = \mathbf{P}_{\mathbb{M}}x_{t+h}$$

Aufgrund der Linearität der Projektion können wir auch sofort die (MSE) optimale Prognose für beliebige Linearkombinationen  $cx_{t+h}$ ,  $c \in \mathbb{R}^{1 \times n}$  angeben:

$$\mathbf{P}_{\mathbb{M}}(cx_{t+h}) = c(\mathbf{P}_{\mathbb{M}}x_{t+h}) = c\hat{x}_{t,h}$$

Daher folgt auch

$$\Sigma_{h,k} \leq \mathbf{E}(x_{t+h} - \tilde{x})(x_{t+h} - \tilde{x})' \quad 3.7$$

für alle  $\tilde{x}$  der Form  $\tilde{x} = \tilde{c}_0 + \tilde{c}_1x_t + \dots + \tilde{c}_kx_{t+1-k}$ ,  $\tilde{c}_0 \in \mathbb{R}^{n \times 1}$ ,  $\tilde{c}_i \in \mathbb{R}^{n \times n}$   $i = 1, \dots, k$ . Eine weitere Folgerung ist

$$\Sigma_{h,k} \leq \Sigma_{h,k-1} \quad 3.8$$

d.h. die Prognose wird i.A. besser (zumindest kann sie nicht schlechter werden) je mehr Information zur Prognose zur Verfügung steht. (Diese Ungleichung folgt aus (3.7), wenn man  $\tilde{x} = \hat{x}_{t,h,k-1}$  setzt.)

Wenn  $\Sigma_{h,k}$  singular ist, dann gibt es Linear-Kombinationen  $cx_{t+h}$ , die perfekt (d.h. ohne Fehler) prognostiziert werden. In diesem Fall ist die (Block-) Töplitz Matrix  $\Gamma_{k+h}$  auch singular. Allgemein gilt

$$\text{rk}(\Gamma_{k+1}) = \text{rk}(\Gamma_k) + \text{rk}(\Sigma_{1,k}) \quad 3.9$$

und daher auch

$$(\Gamma_{k+1} > 0) \iff ((\Gamma_k > 0) \text{ und } (\Sigma_{1,k} > 0)) \quad 3.10$$

$$(\det(\Gamma_k) = 0) \implies (\det(\Sigma_{1,k-1}) = 0) \implies (\det(\Sigma_{1,k}) = 0) \quad 3.11$$

In der Praxis haben wir typischerweise nur eine Zeitreihe zur Verfügung. Daher muss in einem ersten Schritt der Erwartungswert und die Autokovarianzfunktion geschätzt werden. Das kann z.B. mit den Schätzern von Abschnitt 2 gemacht werden, oder man schätzt zunächst ein Modell für den zugrunde liegenden Prozess (z.B. ein AR oder ARMA Modell) und konstruiert dann die Schätzer für den Erwartungswert und die Autokovarianzfunktion mit diesem Modell.

Natürlich verschlechtert sich die Prognose, wenn wir nur eine Schätzung für  $\mu$  und  $\gamma(\cdot)$  zur Verfügung haben. Insbesondere gilt die Ungleichung (3.8) für "reale Prognosen" nicht mehr für jedes beliebige  $k$ . D.h. wenn man  $k$  (bzw. die Ordnung des AR oder ARMA Modells) zu groß wählt, dann erhält man schlechte Prognosen.



### 3.3 Prognose aus der unendlichen Vergangenheit

In diesem Abschnitt betrachten wir den Grenzwert der Prognosen  $\hat{x}_{t,h,k}$  für  $k \rightarrow \infty$ , d.h. für den Fall dass wir mehr und mehr Informationen aus der Vergangenheit benutzen. Wir betrachten wieder nur den Fall von Prozessen mit Erwartungswert  $\mathbf{E}x_t = 0$ . Alle Resultate lassen sich aber ohne weiteres auf den allgemeinen Fall übertragen. Es gilt folgender Satz

**Satz 3.3.** Die Folge  $(\hat{x}_{t,h,k} \mid k \in \mathbb{N})$  konvergiert (im quadratischen Mittel, bzw. im  $\mathbb{L}_2$ ) und der Grenzwert ist die Projektion von  $x_{t+h}$  auf den Raum  $\mathbb{H}_x(t) = \overline{\text{span}}\{x_{is} \mid 1 \leq i \leq n, s \leq t\} = \overline{\text{span}}\{x_s \mid s \leq t\}$

$$\hat{x}_{t,h} = \text{l.i.m}_{k \rightarrow \infty} \hat{x}_{t,h,k} = \mathbf{P}_{\mathbb{H}_x(t)} x_{t+h}$$

Die Varianz des entsprechenden Fehlers  $\hat{u}_{t,h} = x_{t+h} - \hat{x}_{t,h}$  ist

$$\Sigma_h := \mathbf{Var}(\hat{u}_{t,h}) = \mathbf{E}\hat{u}_{t,h}\hat{u}_{t,h}' = \lim_{k \rightarrow \infty} \Sigma_{h,k}$$

*Beweis.* Die Folge der Prognosefehler Varianzen  $\Sigma_{h,k}$  ist monoton nicht wachsend und nach unten beschränkt<sup>2</sup>. Daher existiert der Grenzwert  $\Sigma_h = \lim_k \Sigma_{h,k}$  und für jedes  $\epsilon > 0$  existiert ein  $k_0 \in \mathbb{N}$  sodass  $0 \leq \Sigma_{h,k} - \Sigma_{h,l} \leq \epsilon I_n$  für alle  $k_0 \leq k \leq l$  gilt. Wir betrachten nun die entsprechenden Prognosen aus  $k$  bzw.  $l$  vergangenen Werten

$$^2 0 \leq \Sigma_{h,k+1} \leq \Sigma_{h,k}$$

$$x_{t+h} = \hat{x}_{t,h,l} + \hat{u}_{t,h,l} = \hat{x}_{t,h,k} + (\hat{x}_{t,h,l} - \hat{x}_{t,h,k}) + \hat{u}_{t,h,l}$$

und die Prognosefehler

$$\hat{u}_{t,h,k} = (\hat{x}_{t,h,l} - \hat{x}_{t,h,k}) + \hat{u}_{t,h,l}$$

Da  $\hat{u}_{t,h,l}$  zu  $x_t, x_{t-1}, \dots, x_{t+1-l}$  unkorreliert (orthogonal) ist, folgt weiter

$$\mathbf{Var}(\hat{u}_{t,h,k}) = \mathbf{Var}((\hat{x}_{t,h,l} - \hat{x}_{t,h,k}) + \hat{u}_{t,h,l}) = \mathbf{Var}(\hat{x}_{t,h,l} - \hat{x}_{t,h,k}) + \mathbf{Var}(\hat{u}_{t,h,l})$$

und daher

$$\mathbf{Var}(\hat{x}_{t,h,l} - \hat{x}_{t,h,k}) = \Sigma_{h,k} - \Sigma_{h,l} \leq \epsilon I$$

Die Folge  $(\hat{x}_{t,h,k} \mid k \in \mathbb{N})$  ist also eine Cauchy Folge und daher konvergent.

Da  $\hat{x}_{t,h,k} \in \mathbb{H}_x(t)$  für alle  $k$ , folgt  $\hat{x}_{t,h} = \text{l.i.m}_k \hat{x}_{t,h,k} \in \mathbb{H}_x(t)$  und der Fehler  $\hat{u}_{t,h}$  ist orthogonal auf alle Elemente von  $\mathbb{H}_x(t)$ , da

$$\mathbf{E}\hat{u}_{t,h}x_s' = \mathbf{E}(\text{l.i.m}_k \hat{u}_{t,h,k})x_s' = \lim_k \underbrace{\mathbf{E}\hat{u}_{t,h,k}x_s'}_{=0 \text{ für } k > (t-s)} = 0$$

Diese beiden Bedingungen charakterisieren aber genau die Projektion von  $x_{t+h}$  auf  $\mathbb{H}_x(t)$ . Das heißt  $\hat{x}_{t,h} = \mathbf{P}_{\mathbb{H}_x(t)} x_{t+h}$ .  $\square$

Mit der Ein-Schrittprognose aus der unendlichen Vergangenheit erhalten wir eine Zerlegung des Prozesses der Form

$$x_t = \hat{x}_{t-1,1} + \hat{u}_{t-1,1}$$

wobei  $\hat{x}_{t-1,1}$  der Teil von  $x_t$  ist, der aus der Vergangenheit bestimmt ist, und  $\hat{u}_{t-1,1}$  ist der "nicht vorhersehbare" Anteil. Daher nennt man die Ein-Schritt Prognosefehler aus der unendlichen Vergangenheit die *Innovationen* des Prozesses.

**Satz 3.4.** Die Innovationen eines stationären Prozesses sind ein *white noise* Prozess.

*Beweis.* Klarerweise sind die Innovationen (schwach) stationär und der Erwartungswert ist gleich Null ( $\mathbb{E}\hat{u}_{t-1,1} = 0$ ). Es gilt  $\hat{u}_{t-1,1} \in \mathbb{H}_x(t)$  und  $\hat{u}_{t-1,1} \perp \mathbb{H}_x(t-1)$ . Daher gilt  $\hat{u}_{t-1,1} \perp \hat{u}_{s-1,1} \in \mathbb{H}_x(s) \subseteq \mathbb{H}_x(t-1)$  für alle  $s \leq t-1$  und damit also auch  $\hat{u}_{t,1} \perp \hat{u}_{s,1}$  für alle  $s \neq t$ .  $\square$

Die Ungleichung

$$\Sigma_{h+1} \geq \Sigma_h$$

3.12

folgt unmittelbar aus  $\mathbb{H}_x(t-1) \subseteq \mathbb{H}_x(t)$  und  $\hat{x}_{t-1,h+1} = \mathbf{P}_{\mathbb{H}_x(t-1)}x_{t+h}$  und  $\hat{x}_{t,h} = \mathbf{P}_{\mathbb{H}_x(t)}x_{t+h}$ .

### 3.4 Reguläre und singuläre Prozesse und die Wold Zerlegung

**Definition 3.5.** Ein stationärer Prozess  $(x_t)$  ist

- *regulär* (regular, purely non-deterministic), wenn  $\lim_{h \rightarrow \infty} \hat{x}_{t,h} = 0$  (und daher  $\lim_{h \rightarrow \infty} \Sigma_h = \mathbb{E}x_t x_t'$ )
- *singulär* (singular, deterministic), wenn  $\Sigma_h = 0$  für ein  $h > 0$  (und daher auch für alle  $h > 0$ ).

Der Erwartungswert eines regulären Prozesses muss Null sein:  $\mathbb{E}x_t = 0$ .

Einen singulären Prozess nennt man auch *deterministisch*, weil die Zukunft aus der Vergangenheit bestimmt ist.

**Beispiel 3.6.**

1. MA( $q$ ) Prozesse sind regulär, da  $\hat{x}_{t,h} = 0$  für  $h > q$ .
2. Ein stationärer Prozess der Form  $(x_t = z \mid t \in \mathbb{Z})$  ist singulär, da  $\hat{x}_{t,1} = 1x_t = z = x_{t+1}$ . Man kann auch allgemeiner zeigen, dass harmonische Prozesse (siehe Definition 4.3) singulär sind.

**Theorem 3.7** (Wold Zerlegung).

1. Jeder stationäre Prozess  $(x_t)$  besitzt eine eindeutige Zerlegung  $x_t = y_t + z_t$  mit folgenden Eigenschaften:
  - (a)  $(y_t)$  ist regulär und  $(z_t)$  ist singulär.
  - (b) Die Prozesse  $(y_t)$  und  $(z_t)$  sind zueinander orthogonal ( $\mathbb{E}y_t z_s' = 0$  für alle  $t, s \in \mathbb{Z}$ ).
  - (c)  $y_t \in \text{span}\{1\} + \mathbb{H}_x(t)$  und  $z_t \in \text{span}\{1\} + \mathbb{H}_x(t)$ .
2. Der reguläre Prozess  $(y_t)$  besitzt eine kausale MA( $\infty$ ) Darstellung

$$y_t = \sum_{j \geq 0} b_j \epsilon_{t-j}, \quad b_0 = I, \quad \sum_{j \geq 0} \|b_j\|^2 < \infty$$

wobei  $(\epsilon_t)$  eine white noise Prozess ist. Es gilt  $\mathbb{H}_\epsilon(t) = \mathbb{H}_y(t)$  und die  $\epsilon_t$ 's sind sowohl die Innovationen von  $(x_t)$  als auch von  $(y_t)$ .

Für den Beweis dieses Theorems benötigen wir folgendes Lemma.

**Lemma 3.8.** Sei  $(\epsilon_t) \sim \text{WN}(\Sigma)$  ein white noise Prozess. Die Projektion von  $x \in \mathbb{L}_2$  auf  $\mathbb{H}_\epsilon(t) = \overline{\text{span}}\{\epsilon_s \mid s \leq t\}$  lässt sich schreiben als

$$\mathbf{P}_{\mathbb{H}_\epsilon(t)}x = \sum_{j \geq 0} b_j \epsilon_{t-j} \text{ wobei } \sum_{j \geq 0} \|b_j\|^2 < \infty$$

*Beweis.* Wir betrachten die Projektion von  $x$  auf  $\mathbb{H}_\epsilon(t-k:t) := \text{span}\{\epsilon_{t-j} \mid 0 \leq j \leq k\}$

$$\hat{x}^k := \mathbf{P}_{\mathbb{H}_\epsilon(t-k:t)}x = \sum_{j=0}^k \mathbf{P}_{\text{span}\{\epsilon_{t-j}\}}x = \sum_{j=0}^k b_j \epsilon_{t-j}$$

Hier haben wir  $\epsilon_t \perp \epsilon_s$  und damit  $\text{span}\{\epsilon_t\} \perp \text{span}\{\epsilon_s\}$  für alle  $t \neq s$  verwendet. Man kann leicht zeigen (siehe auch Beweis von Satz 3.3), dass die Folge  $(\hat{x}^k \mid k \in \mathbb{N})$  konvergiert und dass der Grenzwert gleich der Projektion von  $x$  auf  $\mathbb{H}_\epsilon(t)$  ist. Es gilt also  $\mathbf{P}_{\mathbb{H}_\epsilon(t)}x = \text{l.i.m.}_{k \rightarrow \infty} \hat{x}^k = \text{l.i.m.}_{k \rightarrow \infty} \sum_{j=0}^k b_j \epsilon_{t-j} = \sum_{j \geq 0} b_j \epsilon_{t-j}$  wie behauptet. Weiters folgt

$$\sum_{j=0}^k b_j \Sigma b_j' = \|\hat{x}^k\|^2 \leq \|x\|^2$$

und daher  $\sum_{j \geq 0} b_j \Sigma b_j' < \infty$ . Für  $\Sigma > 0$  folgt daher auch  $\sum_{j \geq 0} \|b_j\|^2 < \infty$ . Falls  $\Sigma$  singular ist, dann sind die Koeffizienten  $b_j$  nicht eindeutig, aber man kann sie so wählen, dass sie quadratisch summierbar sind.  $\square$

*Beweis von Theorem 3.7.* Zunächst definieren wir die Innovationen

$$\epsilon_t = x_t - \mathbf{P}_{\mathbb{H}_x(t-1)}x_t \quad (3.13)$$

des Prozesses  $(x_t)$  und merken an, dass

$$\begin{aligned} \epsilon_t &\in \mathbb{H}_x(t) \\ \mathbb{H}_\epsilon(t) &\subseteq \mathbb{H}_x(t) \\ \epsilon_t &\perp \mathbb{H}_x(t-1) \end{aligned} \quad (3.14)$$

Die Prozesse  $(y_t)$  und  $(z_t)$  definieren wir wie folgt

$$y_t = \mathbf{P}_{\mathbb{H}_\epsilon(t)}x_t \quad (3.15)$$

$$z_t = x_t - y_t \quad (3.16)$$

Der Prozess  $(\epsilon_t)$  ist ein white noise Prozess (siehe Satz 3.4) und daher besitzt  $y_t \in \mathbb{H}_\epsilon(t)$  (nach dem obigen Lemma) eine Darstellung als  $\text{MA}(\infty)$  Prozess

$$y_t = \sum_{j \geq 0} b_j \epsilon_{t-j} \quad (3.17)$$

Die Koeffizienten  $b_j$  sind quadratisch summierbar und wegen

$$b_0 \epsilon_t = \mathbf{P}_{\text{span}\{\epsilon_t\}}x_t = \mathbf{P}_{\text{span}\{\epsilon_t\}}\epsilon_t + \mathbf{P}_{\text{span}\{\epsilon_t\}}\mathbf{P}_{\mathbb{H}_x(t-1)}x_t = \epsilon_t$$

Im skalaren Fall  $(\epsilon_t) \sim \text{WN}(\sigma^2 = 1)$  ist  $\{\epsilon_{t-j} \mid j \geq 0\}$  eine ortho-normal Basis von  $\mathbb{H}_\epsilon(t)$  und die Koeffizienten lassen sich besonders einfach berechnen

$$b_j = \langle x, \epsilon_{t-j} \rangle = \mathbf{E}x\epsilon_{t-j}$$

Um den Beweis etwas zu vereinfachen, nehmen wir hier an, dass  $\mathbf{E}x_t = 0$  gilt. Für den allgemeinen Fall betrachtet man einfach die Wold-Zerlegung  $\tilde{x}_t = \tilde{y}_t + \tilde{z}_t$  des Mittelwertbereinigten Prozesses  $\tilde{x}_t = x_t - \mathbf{E}x_t$  und setzt dann  $x_t = y_t + z_t$  wobei  $y_t = \tilde{y}_t$  und  $z_t = \tilde{z}_t + \mathbf{E}x_t$ .

können wir o.E.d.A.  $b_0 = I$  setzen. Wir sehen auch, dass

$$\begin{aligned} y_t &\in \mathbb{H}_\epsilon(t) \subseteq \mathbb{H}_x(t) \\ \mathbb{H}_y(t) &\subseteq \mathbb{H}_\epsilon(t) \subseteq \mathbb{H}_x(t) \\ z_t &\in \mathbb{H}_x(t) \\ \mathbb{H}_z(t) &\subseteq \mathbb{H}_x(t) \\ z_t &\perp \mathbb{H}_\epsilon(t) \end{aligned}$$

Die Prozesse  $(z_t)$  und  $(\epsilon_t)$  sind orthogonal zueinander, da

$$z_t \perp \epsilon_s \text{ für alle } s, t \in \mathbb{Z}$$

Für  $t \geq s$  folgt diese Behauptung aus  $z_t \perp \mathbb{H}_\epsilon(t)$ ,  $\epsilon_s \in \mathbb{H}_\epsilon(t)$  und für  $s > t$  aus  $\epsilon_s \perp \mathbb{H}_x(t)$ ,  $z_t \in \mathbb{H}_x(t)$ . Da  $\mathbb{H}_y(t) \subseteq \mathbb{H}_\epsilon(t)$  sind auch die Prozesse  $(y_t)$  und  $(z_t)$  orthogonal zueinander. Es gilt  $x_t = y_t + z_t$  und daher  $\mathbb{H}_x(t) \subseteq \mathbb{H}_y(t) \oplus \mathbb{H}_z(t)$ . Andererseits haben wir aber auch  $\mathbb{H}_y(t) \oplus \mathbb{H}_z(t) \subseteq \mathbb{H}_x(t)$  wegen  $\mathbb{H}_y(t) \subseteq \mathbb{H}_x(t)$  und  $\mathbb{H}_z(t) \subseteq \mathbb{H}_x(t)$ . Der Unterraum  $\mathbb{H}_x(t)$  ist also die Summe von zwei orthogonalen Unterräumen

$$\mathbb{H}_x(t) = \mathbb{H}_y(t) \oplus \mathbb{H}_z(t)$$

Wegen  $\mathbb{H}_\epsilon(t) \subseteq \mathbb{H}_x(t) = \mathbb{H}_y(t) \oplus \mathbb{H}_z(t)$ ,  $\mathbb{H}_\epsilon(t) \perp \mathbb{H}_z(t)$  und  $\mathbb{H}_y(t) \subseteq \mathbb{H}_\epsilon(t)$  folgt auch

$$\mathbb{H}_\epsilon(t) = \mathbb{H}_y(t)$$

Wir betrachten nun die Prognose von  $y_{t+h}$  aus der unendlichen Vergangenheit. Dazu zerlegen wir  $y_{t+h}$  in

$$y_{t+h} = \underbrace{\sum_{j=0}^{h-1} b_j \epsilon_{t+h-j}}_{\perp \mathbb{H}_\epsilon(t) = \mathbb{H}_y(t)} + \underbrace{\sum_{j \geq h} b_j \epsilon_{t+h-j}}_{\in \mathbb{H}_\epsilon(t) = \mathbb{H}_y(t)}$$

Der zweite Teil ist in  $\mathbb{H}_y(t) = \mathbb{H}_\epsilon(t)$  enthalten und der erste ist orthogonal auf diesen Raum. Daher ist die Prognose gleich

$$\hat{y}_{t,h} = \mathbf{P}_{\mathbb{H}_y(t)} y_{t+h} = \sum_{j \geq h} b_j \epsilon_{t+h-j} \quad \boxed{3.18}$$

und wir sehen, dass der Prozess  $(y_t)$  regulär ist, da

$$\text{l.i.m}_{h \rightarrow \infty} \hat{y}_{t,h} = \text{l.i.m}_{h \rightarrow \infty} \sum_{j \geq h} b_j \epsilon_{t+h-j} = 0$$

Der Fehler der Ein-Schrittprognose für  $y_{t+1}$  ist  $b_0 \epsilon_{t+1} = \epsilon_{t+1}$ . Das heißt, die  $\epsilon_t$ 's sind auch die Innovationen von  $(y_t)$ .

Aufgrund der Orthogonalitätsbeziehung  $\epsilon_t \perp \mathbb{H}_x(t-1)$ , siehe (3.14) und (3.13), können wir den Raum  $\mathbb{H}_x(t)$  auch folgendermaßen in eine Summe von orthogonalen Räumen zerlegen

$$\begin{aligned} \mathbb{H}_x(t) &= \text{span}\{\epsilon_t\} \oplus \mathbb{H}_x(t-1) \\ &= \text{span}\{\epsilon_t\} \oplus \text{span}\{\epsilon_{t-1}\} \oplus \mathbb{H}_x(t-2) \\ &\vdots \\ &= \text{span}\{\epsilon_t\} \oplus \cdots \oplus \text{span}\{\epsilon_{t+1-k}\} \oplus \mathbb{H}_x(t-k) \end{aligned}$$

$z_t$  ist orthogonal auf  $\mathbb{H}_\epsilon(t)$ , d.h. auf alle  $\epsilon_s$ ,  $s \leq t$ . Daher folgt aus der obigen Zerlegung von  $\mathbb{H}_x(t)$  (zusammen mit  $z_t \in \mathbb{H}_x(t)$ ), dass

$$z_t \in \mathbb{H}_x(s) \text{ für alle } s \leq t$$

Insbesondere gilt  $z_{t+1} \in \mathbb{H}_x(t) = \mathbb{H}_y(t) \oplus \mathbb{H}_z(t)$  und wegen  $z_{t+1} \perp \mathbb{H}_y(t)$  auch  $z_{t+1} \in \mathbb{H}_z(t)$ . Das bedeutet aber

$$\hat{z}_{t,1} = \mathbf{P}_{\mathbb{H}_z(t)} z_{t+1} = z_{t+1}$$

und wir haben somit gezeigt, dass  $(z_t)$  ein singulärer Prozess ist.

Es bleibt nur noch die Eindeutigkeit dieser Wold Darstellung zu beweisen. Sei also  $x_t = y_t + z_t$  eine (beliebige) Zerlegung, die die Bedingungen (a)-(c) erfüllt. Damit folgt  $\mathbb{H}_x(t) = \mathbb{H}_y(t) \oplus \mathbb{H}_z(t)$ ,  $\mathbb{H}_y(t) \perp \mathbb{H}_z(t)$  und daher

$$\begin{aligned} \text{l.i.m}_{s \rightarrow -\infty} \mathbf{P}_{\mathbb{H}_x(s)} x_t &= \text{l.i.m}(\mathbf{P}_{\mathbb{H}_y(s)} + \mathbf{P}_{\mathbb{H}_z(s)})(y_t + z_t) \\ &= \underbrace{\text{l.i.m} \mathbf{P}_{\mathbb{H}_y(s)} y_t}_{=0} + \underbrace{\text{l.i.m} \mathbf{P}_{\mathbb{H}_z(s)} z_t}_{=z_t} \\ &= z_t \end{aligned}$$

Das heißt  $z_t$  (und damit natürlich auch  $y_t$ ) ist eindeutig.  $\square$

Folgendes Korollar ist eine unmittelbare Folgerung des obigen Theorems bzw. dessen Beweises.

**Korollar 3.9.** *Ein stationärer Prozess  $(x_t)$  ist dann und nur dann regulär, wenn er eine kausale  $MA(\infty)$  Darstellung  $x_t = \sum_{j \geq 0} b_j \epsilon_{t-j}$  mit  $b_0 = I$  und  $\mathbb{H}_x(t) = \mathbb{H}_\epsilon(t)$  besitzt. Die  $\epsilon_t$ 's sind die Innovationen des Prozesses  $(x_t)$  und für die  $h$ -Schrittprognose gilt*

$$\begin{aligned} \hat{x}_{t,h} &= \sum_{j \geq h} b_j \epsilon_{t+h-j} \\ \hat{u}_{t,h} &= \sum_{j=0}^{h-1} b_j \epsilon_{t+h-j} \\ \Sigma_h &= \sum_{j=0}^{h-1} b_j \Sigma b_j' \end{aligned}$$

Für eine ausführlichere Diskussion, siehe z.B. BROCKWELL&DAVIS<sup>3</sup> (section 5.7).

Diese Überlegungen zeigen auch, dass die  $h$ -Schrittprognose für  $x_{t+h}$  (aus der unendlichen Vergangenheit) gegeben ist durch

$$\hat{x}_{t,h} = \hat{y}_{t,h} + z_{t+h}$$

Für ARMA Prozesse, siehe Abschnitt 7, werden wir zeigen, wie man diese  $MA(\infty)$  Darstellung bestimmen kann.

<sup>3</sup> P.J. Brockwell and R.A. Davis. *Time Series: Theory and Methods*. Springer series in statistics. Springer-Verlag, New York, 2nd edition, 1991

### 3.5 Einfache, heuristische Prognoseverfahren

In diesem Abschnitt diskutieren wir zwei einfache, heuristische Prognoseverfahren, die oft als "benchmark" dienen.

Die *naïve Prognose*

$$\hat{x}_{t,h} = x_t$$

nimmt einfach den letzten, beobachteten Wert als Prognose für alle zukünftigen Werte. Man kann zeigen, dass diese Prognose (MSE) optimal ist für einen random walk (ohne Drift).

Die einfache exponentielle Glättung (single exponential smoothing) ist durch folgende Rekursionsgleichungen definiert

$$\begin{aligned} m_t &= m_{t-1} + \alpha(x_t - m_{t-1}) = \alpha x_t + (1 - \alpha)m_{t-1} \\ &= (1 - \alpha)^t m_0 + \alpha \sum_{j=0}^{t-1} (1 - \alpha)^j x_{t-j} \text{ für } t = 1, 2, \dots \\ \hat{x}_{t,h} &= m_t \end{aligned}$$

Die Prognose hängt von zwei Tuning-Parametern, dem “Glättungsfaktor”  $0 < \alpha \leq 1$  und dem Startwert  $m_0$  ab. (Oft setzt man einfach  $m_0 = x_1$  und damit auch  $m_1 = x_1$ .) Die grundlegende Idee ist es, den Prozess (die Zeitreihe) in eine Trendkomponente und einen “zufälligen” Rest zu zerlegen. Die Trendkomponente wird mit  $m_t$  also einem gewichtetem Mittel aller vergangener Werte geschätzt. Die Gewichte

$\alpha(1 - \alpha)^j$  nehmen mit wachsendem “lag”  $j$  ab, das heißt je weiter ein Wert zurückliegt, desto weniger geht er in die Schätzung des Trends ein. Die geschätzte Trendkomponente ist dann die Prognose für alle zukünftigen Werte. Man kann zeigen, dass dieses Prognoseverfahren für “verrauschte random walk” Prozesse, d.h. für Prozesse der Form  $x_t = (\sum_{j=1}^t \epsilon_j) + \eta_t$ ,  $\epsilon_t \sim \text{WN}(\sigma_\epsilon^2)$ ,  $(\eta_t) \sim \text{WN}(\sigma_\eta^2)$ ,  $\mathbb{E}\epsilon_t\eta_s = 0$  asymptotisch optimal ist. Asymptotisch optimal bedeutet  $\text{l.i.m}(\hat{x}_{t,h} - \hat{x}_{t,h,t}) \rightarrow 0$ , wobei  $\hat{x}_{t,h,t}$  die (MSE) optimale  $h$ -Schrittprognose für  $x_{t+h}$  aus den Werten  $x_1, \dots, x_t$  bezeichnet.

### 3.6 Empirischer Vergleich von Prognoseverfahren

Um einen fairen Vergleich zwischen verschiedenen Prognoseverfahren zu bekommen, ist es wichtig die Daten in einen “Schätz-Datensatz”  $t = 1, \dots, T_1$  und einen “Validierungs-Datensatz”  $t = T_1 + 1, T_1 + 2, \dots, T_1 + T_2 = T$  zu zerlegen. Die ersten  $T_1$  Beobachtungen dienen zur Bestimmung der Tuning-Parameter des Prognoseverfahrens (also z.B. zur Schätzung der ACF  $\gamma$  für die (MSE) optimale, lineare Prognose, oder für die Bestimmung des Glättungsparameters  $\alpha$  für die exponentielle Glättung). Der Validierungs-Datensatz dient dann zum Vergleich der Prognoseverfahren. Dazu verwendet man eine Kostenfunktion  $c(x_t, \hat{x}_t)$ , die die Güte der Prognose  $\hat{x}_t$  quantifiziert. Oft verwendete (aggregierte) Güte-Maße sind

MSE (mean squared error)	$\frac{1}{T_2} \sum_{t=T_1+1}^{T_1+T_2} (x_t - \hat{x}_t)'(x_t - \hat{x}_t)$
RMSE (root mean squared error)	$\sqrt{\text{MSE}}$
MAE (mean absolute error)	$\frac{1}{T_2} \sum_{t=T_1+1}^{T_1+T_2} \sum_{i=1}^n  x_{it} - \hat{x}_{it} $
MAPE (mean absolute percentage error)	$\frac{1}{nT_2} \sum_{t=T_1+1}^{T_1+T_2} \sum_{i=1}^n \frac{ x_{it} - \hat{x}_{it} }{ x_{it} }$

Die Prognose  $\hat{x}_t$  wird mit dem Prognoseverfahren berechnet, das bewertet werden soll. Grundsätzlich sind dabei folgende zwei Varianten möglich

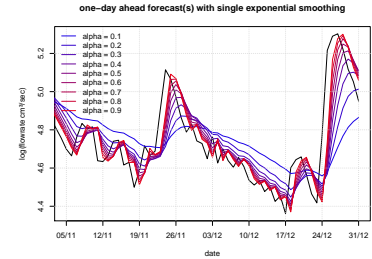


Abbildung 3.1: Effekt des Glättungsparameters  $\alpha$  auf die Prognose

Ein kleines  $\alpha \approx 0$  bedeutet starke Glättung, da die Gewichte langsam abklingen und daher auch noch weit zurückliegende Werte für die Trendschätzung einen erheblichen Beitrag liefern. Umgekehrt bedeutet ein großes  $\alpha \approx 1$ , dass die Gewichte schnell abklingen und daher vor allem die Werte in der nahen Vergangenheit die Trendschätzung bestimmen. Der geschätzte Trend reagiert daher schneller auf Änderungen und ist “rauer”.

Für  $\alpha = 1$  erhält man die naive Prognose.

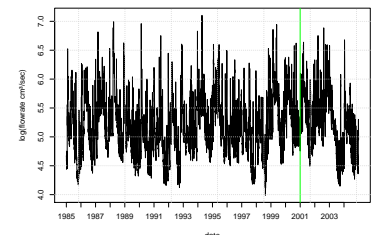


Abbildung 3.2: logarithmierte Durchflussraten der Donau bei Donauwörth

1.  $\hat{x}_t$  ist die  $(t - T_1)$ -Schritt Prognose für  $x_t$  basierend auf den Beobachtungen bis zum Zeitpunkt  $T_1$ , oder
2.  $\hat{x}_t$  ist die 1-Schritt Prognose für  $x_t$  basierend auf den Beobachtungen bis zum Zeitpunkt  $t - 1$ .

Man berechnet nun für alle in Frage kommenden Prognoseverfahren das Gütekriterium (z.B. RMSE) auf dem Validierungs-Datensatz und wählt dann das Verfahren, das am besten abschneidet.

Wählt man das Prognose-Verfahren aufgrund des MSE auf dem Schätz-Datensatz, dann wird man typischerweise das "komplexeste" Verfahren auswählen. Das heißt, man würde z.B. für die in Abschnitt 3.2 diskutierte Prognose, aufgrund der Ungleichung  $\Sigma_{h,k} \geq \Sigma_{h,k+1}$ , immer ein möglichst großes  $k$  wählen. Diese Ungleichung ist aber nur gültig, wenn man die wahre Autokovarianz zur Bestimmung der Prognose kennt. Wenn man die ACF erst schätzen muss, dann verschlechtert sich die Prognose und das insbesondere für große  $k$ , wo viele Koeffizienten  $c_0, c_1, \dots, c_k$  geschätzt werden müssen.

Sei  $\hat{x}_{t,h,k} = \hat{c}(1, (x_t^k)')'$  die Prognose, die mit der geschätzten ACF berechnet wurde und  $\hat{x}_{t,h,k} = c(1, (x_t^k)')'$  die optimale Prognose (die mit der wahren ACF berechnet wurde). Es gilt

$$\begin{aligned} x_{t+h} - \hat{x}_{t,h,k} &= (x_{t+h} - \hat{x}_{t,h,k}) + (\hat{x}_{t,h,k} - \hat{x}_{t,h,k}) \\ &= \hat{u}_{t,h,k} + (c - \hat{c})(1, (x_t^k)')' \end{aligned}$$

Die Varianz von  $\hat{u}_{t,h,k}$  wird mit wachsendem  $k$  kleiner (da  $\Sigma_{h,k+1} \leq \Sigma_{h,k}$ ). Aber die Varianz des zweiten Terms  $(c - \hat{c})(1, (x_t^k)')'$  wird mit wachsendem  $k$  eher größer, da immer mehr Koeffizienten geschätzt werden müssen.

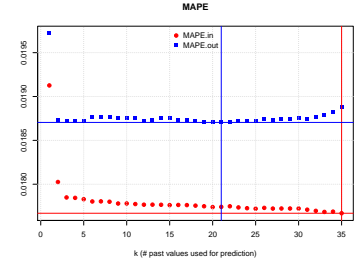


Abbildung 3.3: MAPE Werte der Prognosefehler (der 1-Schrittprognose für die Durchflussmengen) auf dem Schätzzeitraum und auf dem Validierungszeitraum.





## 4

# Spektral Darstellung

In diesem Kapitel werden wir zeigen, dass sich jeder stationäre Prozess als Summe (genauer Integral) von harmonischen Schwingungen mit (zufälligen und unkorrelierten Amplituden) darstellen lässt. Das heißt, man kann den Prozess beliebig genau durch

$$x_t \approx \sum_{k=1}^K [a_k \cos(\lambda_k t) + b_k \sin(\lambda_k t)]$$

approximieren. Diese Spektraldarstellung ist ein zentrales Resultat der Theorie stationärer Prozesse, das sowohl viele methodische Anwendungen hat, als auch für die Interpretation von Prozessen extrem wichtig ist. Besonders nützlich ist diese Spektraldarstellung z.B. für die Analyse von linearen, dynamischen Transformationen von stationären Prozessen, die im Kapitel 5 behandelt wird.

Die Spektraldarstellung des Prozesses liefert auch eine entsprechende Fourierdarstellung der Autokovarianzfunktion. Die Fourier-Transformation der Autokovarianzfunktion (wenn sie existiert) ist die sogenannte spektrale Dichte des Prozesses, die ein wichtiges Instrument zur Analyse und Interpretation von stationären Prozessen ist.

Wir werden in diesem Kapitel nur Prozesse mit Mittel Null betrachten, d.h. wir setzen immer  $\mathbb{E}x_t = 0$ .

### 4.1 Fourier-Transformation

#### Zeitdiskrete Fourier-Transformation

**Definition 4.1.** Sei  $(x_t | t \in \mathbb{Z}) \in \mathbb{C}^{\mathbb{Z}}$  eine absolut summierbare Folge von komplexen Zahlen, d.h.  $\sum_{t=-\infty}^{\infty} |x_t| < \infty$ . Die *zeitdiskrete Fourier-Transformation* (discrete time fouriertransform, DTFT) ist die Funktion

$$X(\lambda) = \sum_{t=-\infty}^{\infty} x_t \exp(-i\lambda t), \quad \lambda \in \mathbb{R} \quad 4.1$$

Die inverse zeitdiskrete Fourier-Transformation (iDTFT) ist definiert durch

$$x_t = \frac{1}{2\pi} \int_{-\pi}^{\pi} X(\lambda) \exp(i\lambda t) d\lambda, \quad t \in \mathbb{Z} \quad 4.2$$

Fourier Transformation, Harmonische Prozesse, Zyklen, saisonale Komponenten, stochastisches Integral, spektral Darstellung von stationären Prozessen, spektrale Verteilung, spektrale Dichte, Zusammenhang mit der ACF, Interpretation der spektralen Verteilung/Dichte

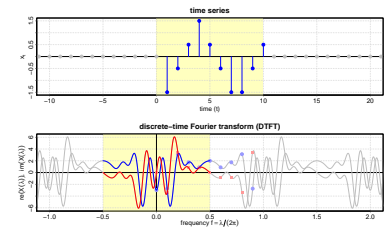


Abbildung 4.1: zeitdiskrete Fourier-Transformation (DTFT)

Einige wichtige Eigenschaften der DTFT sind:

1. Da die Folge  $(x_t)$  absolut summierbar ist, folgt  $\sum_{t=-\infty}^{\infty} |x_t \exp(-i\lambda t)| = \sum_{t=-\infty}^{\infty} |x_t| < \infty$ . Daher konvergiert die Folge der Partialsummen

$$\sum_{t=-T}^T x_t \exp(-i\lambda t) \rightarrow \sum_{t=-\infty}^{\infty} x_t \exp(-i\lambda t)$$

gleichmäßig gegen eine stetige Funktion.  $X(\lambda)$  ist wohldefiniert und stetig.

2. Die Fourier-Transformation  $X(\lambda)$  ist periodisch mit Periode  $2\pi$  ( $X(\lambda) = X(\lambda + 2\pi)$ ). Es genügt daher  $X(\lambda)$  auf dem Intervall  $[-\pi, \pi]$  zu betrachten.
3. Für reelle Folgen ( $x_t \in \mathbb{R}$ ) gilt:  $X(\lambda) = \overline{X(-\lambda)}$ .
4. Die inverse zeitdiskrete Fourier-Transformation liefert wieder die ursprüngliche Folge:

$$(x_t) \xrightarrow{\text{DTFT}} X(\lambda) \xrightarrow{\text{iDTFT}} (x_t)$$

Eine reelle Folge ( $x_t \in \mathbb{R} \mid t \in \mathbb{Z}$ ) kann also mittels der iDTFT (4.2) als Überlagerung von harmonischen Schwingungen mit (Kreis-) Frequenzen  $\lambda \in [0, \pi]$  dargestellt werden.

5. Sei  $Y(\lambda)$  die DTFT von  $(y_t = x_{t-s} \mid t \in \mathbb{Z})$  dann gilt

$$Y(\lambda) = \exp(-i\lambda s) X(\lambda) \quad \boxed{4.3}$$

6. Identität von Parseval:  $(x_t), (y_t)$  seien zwei (skalare, absolut summierbare) Folgen und  $X(\cdot), Y(\cdot)$  deren DTFT's, dann gilt

$$\sum_{t=-\infty}^{\infty} x_t \overline{y_t} = \frac{1}{2\pi} \int_{-\pi}^{\pi} X(\lambda) \overline{Y(\lambda)} d\lambda \quad \boxed{4.4}$$

$$\sum_{t=-\infty}^{\infty} |x_t|^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} |X(\lambda)|^2 d\lambda \quad \boxed{4.5}$$

Diese Eigenschaften sind mit den "Orthogonalitätsbeziehungen" der Funktionen  $\exp(i \cdot t)$  relativ einfach zu beweisen:

Beweis von Punkt 3.:

$$X(\lambda) = \sum_{t=-\infty}^{\infty} x_t \exp(-i\lambda t) = \sum_{t=-\infty}^{\infty} x_t \overline{\exp(-i(-\lambda)t)} = \overline{X(-\lambda)}$$

Beweis von Punkt 4.:

$$\begin{aligned} \frac{1}{2\pi} \int_{-\pi}^{\pi} X(\lambda) \exp(i\lambda t) d\lambda &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \left[ \sum_{s=-\infty}^{\infty} x_s \exp(-i\lambda s) \right] \exp(i\lambda t) d\lambda \\ &= \sum_{s=-\infty}^{\infty} x_s \left[ \frac{1}{2\pi} \int_{-\pi}^{\pi} \exp(i\lambda(t-s)) d\lambda \right] = x_t \end{aligned}$$

Beweis von Punkt 5.:

$$Y(\lambda) = \sum_{t=-\infty}^{\infty} x_{t-s} \exp(-i\lambda t) = \sum_{t=-\infty}^{\infty} x_t \exp(-i\lambda(t+s)) = \exp(-i\lambda s) X(\lambda)$$

Der einfacheren Notation wegen, betrachten wir nur den skalaren Fall. Die Verallgemeinerung auf den multivariaten Fall ist ohne weiters möglich.

Für reelle Folgen ( $x_t \in \mathbb{R} \mid t \in \mathbb{Z}$ ) gilt

$$\begin{aligned} x_t &= \frac{1}{2\pi} \int_{-\pi}^0 X(\lambda) e^{i\lambda t} d\lambda + \frac{1}{2\pi} \int_0^{\pi} X(\lambda) e^{i\lambda t} d\lambda \\ &= \frac{1}{\pi} \Re \left[ \int_0^{\pi} X(\lambda) e^{i\lambda t} d\lambda \right] \\ &= \frac{1}{\pi} \int_0^{\pi} [a(\lambda) \cos(\lambda t) - b(\lambda) \sin(\lambda t)] d\lambda \end{aligned}$$

wobei

$$X(\lambda) = a(\lambda) + ib(\lambda)$$

Die Funktionen  $\{\exp(i \cdot t) \mid t \in \mathbb{Z}\}$  bilden ein "orthonormales" Funktionensystem

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \exp(i\lambda t) \overline{\exp(i\lambda s)} d\lambda = \begin{cases} 1 & \text{für } t = s \\ 0 & \text{sonst.} \end{cases}$$

da

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \exp(i\lambda t) d\lambda = \begin{cases} 1 & \text{für } t = 0 \\ 0 & \text{sonst.} \end{cases}$$

Beweis von Punkt 6.:

$$\begin{aligned} \frac{1}{2\pi} \int_{-\pi}^{\pi} X(\lambda) \overline{Y(\lambda)} d\lambda &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \left[ \sum_{t=-\infty}^{\infty} x_t \exp(-i\lambda t) \right] \left[ \sum_{s=-\infty}^{\infty} \overline{y_s \exp(-i\lambda s)} \right] \\ &= \sum_{t,s=-\infty}^{\infty} x_t \overline{y_s} \left[ \frac{1}{2\pi} \int_{-\pi}^{\pi} \exp(i\lambda(s-t)) d\lambda \right] = \sum_{t=-\infty}^{\infty} x_t \overline{y_t} \end{aligned}$$

### Periodogramm

**Definition 4.2.** Das *Periodogramm* einer (skalaren) Zeitreihe  $(x_t \mid t = 1, \dots, T)$  mit Mittelwert  $\bar{x}$  ist:

$$I(\lambda) = T^{-1} |\tilde{X}(\lambda)|^2 \quad \text{für } \lambda \in [-\pi, \pi] \quad 4.6$$

wobei  $\tilde{X}(\lambda)$  die DTFT der *Mittelwert bereinigten* und auf ganz  $\mathbb{Z}$  mit Nullen fortgesetzten Zeitreihe ist, d.h.

$$\tilde{X}(\lambda) = \sum_{t=1}^T (x_t - \bar{x}) \exp(-i\lambda t), \quad \lambda \in [-\pi, \pi] \quad 4.7$$

Eigenschaften des Periodogramms:

$$I(\lambda) \geq 0 \quad 4.8$$

$$I(0) = 0 \quad 4.9$$

$$I(\lambda) = I(-\lambda) \quad 4.10$$

$$I(\lambda) = \sum_{s=-T+1}^{T-1} \hat{\gamma}(s) \exp(-i\lambda s) \quad 4.11$$

$$\hat{\gamma}(s) = \frac{1}{2\pi} \int_{-\pi}^{\pi} I(\lambda) \exp(i\lambda s) d\lambda \quad 4.12$$

$$\hat{\gamma}(0) = \frac{1}{2\pi} \int_{-\pi}^{\pi} I(\lambda) d\lambda \quad 4.13$$

Die ersten drei Punkte (4.8)-(4.10) sind eine unmittelbare Folge der Definition des Periodogramms und der Eigenschaften der DTFT. Die Gleichung (4.12) besagt, dass die Autokovarianzfunktion die inverse zeitdiskrete Fourier-Transformation des Periodogramms ist. Daher ist umgekehrt das Periodogramm die zeitdiskrete Fourier-Transformation der Autokovarianzfunktion, siehe (4.11). Wir müssen also nur die Gleichung (4.12) beweisen. Sei  $\tilde{x}_t = (x_t - \bar{x})$  für  $t = 1, \dots, T$  und  $\tilde{x}_t = 0$  sonst. Aus der Gleichung (4.3) folgt, dass  $\tilde{X}(\lambda) \exp(i\lambda s)$  die DTFT der Folge  $(\tilde{y}_t = \tilde{x}_{t+s} \mid t \in \mathbb{Z})$  ist. Mit Parseval (4.4) erhalten wir daher:

$$\begin{aligned} \hat{\gamma}(s) &= \frac{1}{T} \sum_{t=1}^T \tilde{x}_{t+s} \tilde{x}_t = \frac{1}{T} \sum_{t=-\infty}^{\infty} \tilde{y}_t \tilde{x}_t \\ &= \frac{1}{T} \frac{1}{2\pi} \int_{-\pi}^{\pi} \tilde{X}(\lambda) \exp(i\lambda s) \overline{\tilde{X}(\lambda)} d\lambda \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} I(\lambda) \exp(i\lambda s) d\lambda \end{aligned}$$

Interpretation des Periodogramms:

- Die ACF  $\hat{\gamma}(s)$  und das Periodogramm sind äquivalent zueinander. Sie enthalten dieselbe Information, stellen sie aber unterschiedlich dar.
- Die iDTFT stellt die (Mittelwert-bereinigte, mit Nullen fortgesetzte) Zeitreihe als Summe (genauer Integral) von harmonischen Schwingungen mit Amplituden  $\tilde{X}(\lambda)$  und (Kreis-) Frequenzen  $\lambda \in [0, \pi]$  dar. Aus dem Periodogramm ( $I(\lambda) = (1/T)|\tilde{X}(\lambda)|^2$ ) kann man ablesen, welche Frequenzen besonders wichtig sind, bzw. welche Frequenzen nur wenig beitragen.
- Die (empirische) Varianz der Zeitreihe erhält man aus dem Integral über das Periodogramm, siehe (4.13). Das Integral

$$\frac{1}{2\pi} \left\{ \int_{-\lambda_2}^{-\lambda_1} I(\lambda) d\lambda + \int_{\lambda_1}^{\lambda_2} I(\lambda) d\lambda \right\} = \frac{1}{\pi} \int_{\lambda_1}^{\lambda_2} I(\lambda) d\lambda$$

ist der Anteil der Varianz, der durch die Schwingungen im Frequenzband  $0 \leq \lambda_1 \leq \lambda_2 \leq \pi$  erklärt wird. Zyklische Komponenten kann man also im Periodogramm durch entsprechende "Spitzen" des Periodogramms erkennen.

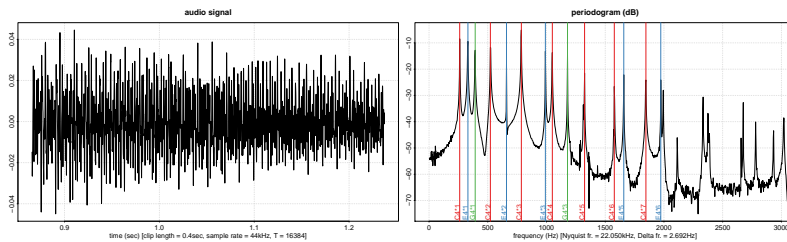


Abbildung 4.2: Akkord Analyse

## 4.2 Harmonische Prozesse

**Definition 4.3.** Ein (skalarer) *harmonischer Prozess* ist ein Prozess der Form

$$x_t = \sum_{k=1}^K z_k \exp(i\lambda_k t), \quad \text{für } t \in \mathbb{Z} \quad \boxed{4.14}$$

wobei  $-\pi < \lambda_1 < \lambda_2 < \dots < \lambda_K \leq \pi$  und die  $z_k$ 's komplexwertige Zufallsvariable sind.

**Satz 4.4.** Ein harmonischer Prozess (4.14) ist dann und nur dann ein reellwertiger, schwach stationärer Prozess, wenn folgende Bedingungen erfüllt sind:

1. Die Zufallsvariablen  $z_k$  sind quadratisch integrierbar (d.h.  $\mathbb{E}|z_k|^2 < \infty$  für  $k = 1, \dots, K$ )
2.  $\mathbb{E}z_k = 0$  wenn  $\lambda_k \neq 0$
3.  $\mathbb{E}z_k \bar{z}_l = 0$  für alle  $k \neq l$
4. Für jede Frequenz  $\pi > \lambda_k > 0$  existiert eine Frequenz  $\lambda_l$  mit  $\lambda_l = -\lambda_k$  und  $z_l = \bar{z}_k$ . Falls  $\lambda_k = 0$  oder  $\lambda_k = \pi$  dann muss  $z_k = \bar{z}_k$  gelten.

Auch in diesem Abschnitt betrachten wir nur den skalaren Fall.

*Beweis.* Wir beweisen hier nur eine Richtung und nehmen an, dass der harmonische Prozess  $(x_t = \sum_{k=1}^K z_k \exp(i\lambda_k t))$  die obigen Bedingungen erfüllt. Aus Punkt 4 folgt

$$x_t = \sum_{k=1}^K z_k \exp(i\lambda_k t) = \sum_{\{k|\lambda_k=0\}} z_k \underbrace{\exp(i0t)}_{=1} + \sum_{\{k|\lambda_k=\pi\}} z_k \underbrace{\exp(i\pi t)}_{=(-1)^t} + \sum_{\{k|0<\lambda_k<\pi\}} \underbrace{z_k \exp(i\lambda_k t) + \bar{z}_k \exp(-i\lambda_k t)}_{=2\Re(z_k \exp(i\lambda_k t))}$$

und daher ist  $x_t$  also eine reell-wertige Zufallsvariable<sup>1</sup>. Punkt 1 impliziert, dass  $x_t$  quadratisch integrierbar ist und mit Punkt 2 folgt

$$\mathbf{E}x_t = \sum_{k=1}^K \mathbf{E}z_k \exp(i\lambda_k t) = \sum_{\{k|\lambda_k=0\}} \mathbf{E}z_k \exp(i0t) \equiv \text{const}$$

Der Erwartungswert von  $x_t$  ist also, wie gewünscht, unabhängig von  $t$ . Die Autokovarianzen

$$\mathbf{E}x_{t+m}x_t = \mathbf{Cov}(x_{t+m}, x_t) + (\mathbf{E}x_t)^2 = \sum_{k,l=1}^K \mathbf{E}z_k \bar{z}_l \exp(i\lambda_k(t+m)) \exp(-i\lambda_l t) = \sum_{k=1}^K \mathbf{E}|z_k|^2 \exp(i\lambda_k m)$$

hängen nur vom "lag"  $m$  ab und somit haben wir gezeigt, dass  $(x_t)$  ein reell-wertiger, schwach stationärer Prozess ist.  $\square$

Wir nehmen<sup>2</sup> nun an, dass  $\mathbf{E}x_t = 0$  und definieren einen Prozess  $(z(\lambda) \mid \lambda \in [-\pi, \pi])$  durch

$$z(\lambda) = \sum_{\{k|\lambda_k \leq \lambda\}} z_k$$

und eine Funktion  $F : [-\pi, \pi] \rightarrow \mathbb{R}$  mit

$$F(\lambda) = \mathbf{E}|z(\lambda)|^2 = \sum_{\{k|\lambda_k \leq \lambda\}} \mathbf{E}|z_k|^2$$

Die Funktion  $F(\cdot)$  ist eine monoton nicht fallende, rechts-stetige Treppenfunktion (mit  $F(-\pi) = 0$  und  $F(\pi) = \mathbf{E}|z(\pi)|^2 = \mathbf{E}x_t^2 < \infty$ ) und definiert daher ein diskretes Maß auf dem Intervall  $[-\pi, \pi]$ . Man kann sich leicht überzeugen, dass man die Autokovarianzfunktion  $\gamma(k)$  des Prozesses  $(x_t)$  folgendermaßen darstellen kann:

$$\mathbf{E}x_{t+k}x_t = \gamma(k) = \int_{-\pi}^{\pi} \exp(i\lambda k) dF(\lambda) \quad (4.15)$$

Auch der Prozess selbst besitzt eine entsprechende Fourier-Darstellung

$$x_t = \int_{-\pi}^{\pi} \exp(i\lambda t) dz(\lambda) \quad (4.16)$$

Wie dieses *stochastische Integral* (4.16) zu interpretieren ist, werden wir im nächsten Abschnitt genauer diskutieren.

Harmonische Prozesse haben in der Praxis wenig Bedeutung, aber für die Theorie stationärer Prozesse sind sie enorm wichtig, da sich jeder stationäre Prozess durch harmonische Prozesse (beliebig genau) approximieren lässt. Man kann auch die sogenannte Spektraldarstellung (4.16) für beliebige stationäre Prozesse definieren und entsprechend besitzt auch die Autokovarianzfunktion jedes stationären Prozesses eine Darstellung, wie in Gleichung (4.15).

Wesentlich für die andere Richtung des Beweises ist die Unabhängigkeit der Funktionen  $\exp(i\lambda_k t)$ : Aus  $\sum_k z_k \exp(i\lambda_k t) = 0 \forall t \in \mathbb{Z}$  folgt  $z_k = 0$  für alle  $k$ .

<sup>1</sup> Die Index-Mengen dieser Summen können natürlich auch leer sein, wenn z.B. keine Frequenz  $\lambda_k$  gleich null ist. In diesem Fall wird die Summe gleich Null gesetzt.

<sup>2</sup> Das heißt  $\mathbf{E}z_k = 0$  für alle  $k = 1, \dots, K$ .

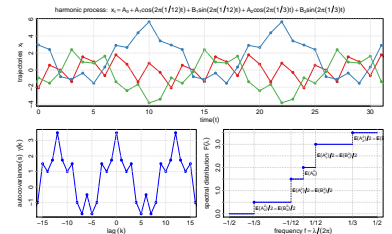


Abbildung 4.3: Trajektorien, ACF und spektrale Verteilung des harmonischen Prozesses:

$$x_t = A_0 + A_1 \cos(\lambda_1 t) + B_1 \sin(\lambda_1 t) + A_2 \cos(\lambda_2 t) + B_2 \sin(\lambda_2 t)$$

Harmonische Prozesse sind singulär!

### 4.3 Das stochastische Integral

**Definition 4.5.** Einen stochastischen Prozess  $(z(\lambda) \mid \lambda \in [-\pi, \pi])$  mit komplexwertigen Zufallsvektoren  $z(\lambda) : \Omega \rightarrow \mathbb{C}^n$  nennt man *Prozess mit orthogonalen Inkrementen*, wenn

1.  $z(-\pi) = 0$  f.s
2.  $\mathbf{E} z(\lambda)^* z(\lambda) < \infty$  für alle  $\lambda \in [-\pi, \pi]$  (der Prozess ist quadratisch integrierbar)
3.  $\mathbf{E} z(\lambda) = 0$  (die Erwartungswerte sind null)
4. l.i.m. $_{\epsilon \downarrow 0} z(\lambda + \epsilon) = z(\lambda)$  für  $\lambda \in [-\pi, \pi)$  (der Prozess ist rechts stetig)
5.  $\mathbf{E} \{ (z(\lambda_4) - z(\lambda_3)) (z(\lambda_2) - z(\lambda_1))^* \} = 0$  für alle  $-\pi \leq \lambda_1 \leq \lambda_2 \leq \lambda_3 \leq \lambda_4 \leq \pi$  (die Inkremente des Prozesses sind orthogonal)

Der Prozess  $z$  ist auf einer kontinuierlichen Indexmenge  $[-\pi, \pi]$  definiert und die Indizes  $\lambda$  werden wir immer als (Kreis-) Frequenzen interpretieren. Da wir hier nur Prozesse mit Mittelwert gleich Null betrachten, setzen wir Annahme 3. Die Komponenten  $z_k(\lambda)$  von  $z(\lambda)$  sind Elemente des (komplexen) Hilbertraums  $\mathbb{L}_2^c(\Omega, \mathcal{A}, P)$ .

Wir definieren eine zugehörige Funktion  $F : [-\pi, \pi] \rightarrow \mathbb{C}^{n \times n}$  mit  $F(\lambda) = \mathbf{E} z(\lambda) z(\lambda)^*$ . Man kann leicht zeigen, dass:

$$\begin{aligned} F(-\pi) &= 0 \\ F(\lambda) &\geq 0 \\ F(\lambda_2) - F(\lambda_1) &= \mathbf{E} \{ (z(\lambda_2) - z(\lambda_1)) (z(\lambda_2) - z(\lambda_1))^* \} \geq 0 \text{ für } \lambda_1 \leq \lambda_2 \\ F(\lambda+) = \lim_{\epsilon \downarrow 0} F(\lambda + \epsilon) &= F(\lambda) \end{aligned}$$

4.17

$F$  ist also eine positiv semidefinite, monoton wachsende, rechts-stetige (Matrix-) Funktion auf  $[-\pi, \pi]$ . (Genauer gesagt, ist  $F$  bzgl. der Löwner Ordnung monoton wachsend, d.h. die Inkremente  $F(\lambda_2) - F(\lambda_1)$  sind für alle  $\lambda_1 \leq \lambda_2$  positiv semidefinit.)

#### Der skalare Fall

Die Funktion  $F$  ist in diesem Fall eine reellwertige, nicht negative, rechts-stetige und monoton wachsende Funktion. D.h.  $F$  ist eine Verteilungsfunktion auf  $[-\pi, \pi]$ . Wir betrachten nun den (komplexen) Hilbertraum  $\mathbb{L}_2^c([-\pi, \pi], \mathcal{B}, F)$  der (bzgl.  $F$ ) quadratisch integrierbaren Funktionen auf  $[-\pi, \pi]$  mit dem inneren Produkt und der Norm:

$$\langle a, b \rangle_F := \int_{-\pi}^{\pi} a(\lambda) \bar{b}(\lambda) dF(\lambda) \text{ und } \|a\|_F^2 := \langle a, a \rangle_F = \int_{-\pi}^{\pi} |a(\lambda)|^2 dF(\lambda)$$

$\mathcal{B}$  bezeichnet die Borel-Sigma Algebra auf  $[-\pi, \pi]$ .

Das *stochastische Integral* bzgl.  $z(\lambda)$  wird zunächst für Treppenfunktion definiert und dann (schrittweise) auf alle Funktionen in  $\mathbb{L}_2^c([-\pi, \pi], \mathcal{B}, F)$  erweitert. Zunächst betrachten wir also die Menge

Genauer gesagt, betrachtet man Äquivalenzklassen von Funktionen bzgl. der Äquivalenzrelation

$$(a \sim b) \iff \int |a(\lambda) - b(\lambda)|^2 dF(\lambda) = 0$$

$\mathbb{M}_2^{\text{step}} \subset \mathbb{L}_2^c([-\pi, \pi], \mathcal{B}, F)$  aller (links-stetigen) Treppenfunktionen der Form

$$a(\lambda) = \sum_{k=1}^m a_k \chi_{(\lambda_k, \lambda_{k+1}]}(\lambda)$$

$$a_k \in \mathbb{C}, k = 1, \dots, m; \quad -\pi = \lambda_1 < \lambda_2 < \dots < \lambda_m < \lambda_{m+1} = \pi$$

$\chi_A(\cdot)$  ist die Indikatorfunktion der Menge  $A \subseteq [-\pi, \pi]$

Für Funktionen dieser Form definiert man nun

$$I(a) = \int_{-\pi}^{\pi} a(\lambda) dz(\lambda) := \sum_{k=1}^m a_k (z(\lambda_{k+1}) - z(\lambda_k))$$

Der Wert des stochastischen Integrals ist unabhängig von der Darstellung der Funktion  $a$  (d.h. von der Wahl der Frequenzen  $\lambda_i$ ), das stochastische Integral ist also wohl definiert.

Das stochastische Integral ist eine komplexwertige und quadratisch integrierbare Zufallsvariable, d.h.  $I(a) \in \mathbb{L}_2^c(\Omega, \mathcal{A}, P)$ . Der Erwartungswert von  $I(a)$  ist gleich Null

$$\mathbb{E}I(a) = \sum_{k=1}^n a_k \mathbb{E}(z(\lambda_{k+1}) - z(\lambda_k)) = 0$$

und für die Kovarianz zwischen  $I(a)$  und  $I(b)$  (für zwei Funktionen  $a, b \in \mathbb{M}_2^{\text{step}}$ , die o.E.d.A. auf einem gemeinsamen Frequenz-Raster definiert sind) gilt aufgrund der Orthogonalität der Inkremente und der Eigenschaften von  $F$

$$\begin{aligned} \mathbb{E}I(a)\overline{I(b)} &= \mathbb{E} \left( \sum_{k=1}^m a_k (z(\lambda_{k+1}) - z(\lambda_k)) \right) \overline{\left( \sum_{k=1}^m b_k (z(\lambda_{k+1}) - z(\lambda_k)) \right)} \\ &= \sum_{j,k=1}^m a_j \overline{b_k} \mathbb{E} (z(\lambda_{j+1}) - z(\lambda_j)) \overline{(z(\lambda_{k+1}) - z(\lambda_k))} \\ &= \sum_{k=1}^m a_k \overline{b_k} (F(\lambda_{k+1}) - F(\lambda_k)) = \int_{-\pi}^{\pi} a(\lambda) \overline{b(\lambda)} dF(\lambda) \end{aligned}$$

Das heißt

$$\langle I(a), I(b) \rangle = \langle a, b \rangle_F$$

4.18

Sei nun  $\mathbb{M}_2 = \overline{\mathbb{M}_2^{\text{step}}}$  der Abschluss von  $\mathbb{M}_2^{\text{step}}$  in  $\mathbb{L}_2^c([-\pi, \pi], \mathcal{B}, F)$ . Für jede Funktion  $a \in \mathbb{M}_2$  existiert eine Folge  $a_n \in \mathbb{M}_2^{\text{step}}$ , die gegen  $a$  konvergiert, d.h.  $\|a - a_n\|_F \rightarrow 0$ . Das stochastische Integral von  $a$  wird als Grenzwert (im quadratischen Mittel) der Integrale  $I(a_n)$  definiert:

$$\int_{-\pi}^{\pi} a(\lambda) dF(\lambda) = I(a) := \lim_{n \rightarrow \infty} I(a_n)$$

Wir müssen natürlich noch zeigen, dass diese Definition Sinn macht, d.h. wir müssen zeigen, dass der Grenzwert existiert und eindeutig ist. Die Folge  $\{a_n\}$  ist eine Cauchy Folge in  $\mathbb{L}_2^c([-\pi, \pi], \mathcal{B}, F)$  und auf Grund der Isometrie (4.18) gilt  $\|I(a_n) - I(a_m)\| = \|a_n - a_m\|_F$ . Daher ist auch  $\{I(a_n)\}$  eine Cauchy Folge (in  $\mathbb{L}_2^c(\Omega, \mathcal{A}, P)$ ) und besitzt also einen Grenzwert. Um die Eindeutigkeit zu zeigen, nehmen wir an, dass es zwei Folgen  $\{a_n\}$  und  $\{b_n\}$  gibt, die gegen  $a$

konvergieren. Dann konvergiert auch die Folge  $a_1, b_1, a_2, b_2, \dots$  gegen  $a$  und es muss daher auch die Folge  $I(a_1), I(b_1), I(a_2), I(b_2), \dots$  einen Grenzwert besitzen. Das impliziert nun wie gewünscht, dass  $\text{l.i.m } I(a_n) = \text{l.i.m } I(b_n)$ .

Damit folgt nun für  $a, b \in \mathbb{M}_2$ :

$$\mathbf{EI}(a) = \mathbf{E} \text{l.i.m}_{n \rightarrow \infty} I(a_n) = \lim_{n \rightarrow \infty} \underbrace{\mathbf{EI}(a_n)}_{=0} = 0$$

$$\langle I(a), I(b) \rangle = \lim_{n \rightarrow \infty} \langle I(a_n), I(b_n) \rangle = \lim_{n \rightarrow \infty} \langle a_n, b_n \rangle_F = \langle a, b \rangle_F$$

4.19

Man kann zeigen, dass das stochastische Integral für alle Funktionen aus  $\mathbb{L}_2^c([-\pi, \pi], \mathcal{B}, F)$  definiert ist. Das heißt

$$\mathbb{M}_2 = \mathbb{L}_2^c([-\pi, \pi], \mathcal{B}, F)$$

Insbesondere enthält

$$\mathbb{M}_2 = \mathbb{L}_2^c([-\pi, \pi], \mathcal{B}, F)$$

alle beschränkten und alle stückweise stetigen Funktionen.

### Der multivariate Fall

Der Fall  $n > 1$  kann im Prinzip ganz analog behandelt werden. Die Matrix-Funktion  $F(\cdot)$  ist nach (4.17) positiv semidefinit, monoton wachsend und rechtsstetig. Die Diagonalelemente  $F_{jj}(\cdot)$  sind also Verteilungsfunktionen von Maßen  $\mu_F^{jj}$ . Sei  $F_0 = \text{tr}(F)$  die Spur von  $F$ , dann ist  $F_0$  die Verteilungsfunktion der Summe dieser Maße, d.h. die Verteilungsfunktion von  $\mu_F^0 = (\mu_F^{11} + \dots + \mu_F^{nn})$ . Da die Inkremente  $\Delta F = F(\lambda_2) - F(\lambda_1) \geq 0$  für alle  $-\pi \leq \lambda_1 \leq \lambda_2 \leq \pi$  positiv semidefinit sind, folgt für  $k \neq j$

$$|\Re(\Delta F_{kj})| \leq \frac{1}{2} \Delta F_0 \text{ und } |\Im(\Delta F_{kj})| \leq \frac{1}{2} \Delta F_0$$

Daher sind sowohl der Realteil als auch der Imaginärteil von  $F_{kj}(\cdot)$  Funktionen beschränkter Variation und definieren signierte Maße auf  $[-\pi, \pi]$ . Wir können also  $F_{kj}(\cdot)$  als Verteilungsfunktion eines komplexwertigen Maßes  $\mu_F^{kj} = (\mu_F^{kj, \Re+} - \mu_F^{kj, \Re-}) + i(\mu_F^{kj, \Im+} - \mu_F^{kj, \Im-})$  interpretieren. Die Maße  $\mu_F^{kj, \Re+}, \dots, \mu_F^{kj, \Im-}$  sind "gewöhnliche" (endliche) Maßfunktionen. Die Funktion  $F(\cdot)$  kann also als Matrix von Verteilungsfunktionen  $F = (F_{kj})$  oder als Verteilungsfunktion eines Matrizenwertigen Maßes  $\mu_F$  interpretiert werden.

$$(b, 1) \begin{pmatrix} \Delta F_{kk} & \Delta F_{kj} \\ \Delta F_{jk} & \Delta F_{jj} \end{pmatrix} (\bar{b}, 1) =$$

$$= |b|^2 \Delta F_{kk} + \Delta F_{jj} + 2\Re(b \Delta F_{kj}) \geq 0$$

Setzt man nun  $b \pm 1$  und  $b = \pm i$  so folgt die Behauptung.

$$\mu_F : \mathcal{B} \longrightarrow \mathbb{C}_+^{n \times n} := \{M \in \mathbb{C}^{n \times n}, M \geq 0\} \subset \mathbb{C}^{n \times n}$$

Insbesondere gilt für Intervalle  $(-\pi \leq \lambda_1 \leq \lambda_2 \leq \pi)$

$$\mu_F((\lambda_1, \lambda_2]) = (F(\lambda_2) - F(\lambda_1)) \geq 0$$

Die Maße  $\mu_F^{kj}$  sind alle bzgl.  $\mu_F^0$  absolut stetig. D.h. Für  $A \in \mathcal{B}$  mit  $\mu_F^0(A) = 0$  folgt  $\mu_F^{kj}(A) = 0$ . Für Intervalle  $A = (\lambda_1, \lambda_2]$  folgt diese Behauptung wieder unmittelbar aus der Bedingung  $\mu_F((\lambda_1, \lambda_2]) = (F(\lambda_2) - F(\lambda_1)) \geq 0$ . Es existiert daher eine Matrizenwertige Funktion

$$M_F : [-\pi, \pi] \longrightarrow \mathbb{C}^{n \times n}$$



mit  $\text{tr}(M(\lambda)) = 1$ , sodass für alle Mengen  $A \in \mathcal{B}$

$$\mu_F(A) = \int_{-\pi}^{\pi} \chi_A(\lambda) dF(\lambda) = \int_{-\pi}^{\pi} \chi_A(\lambda) M_F(\lambda) dF_0(\lambda)$$

Hier bezeichnet  $\chi_A(\cdot)$  die entsprechende Indikator-Funktion für die Menge  $A$ .

Wir betrachten nun (Borel-messbare, Zeilen-) Funktionen

$$a : [-\pi, \pi] \longrightarrow \mathbb{C}^{1 \times n}$$

die in folgendem Sinne (bzgl.  $F$ ) quadratisch integrierbar sind:

$$\int_{-\pi}^{\pi} [a(\lambda) M_F(\lambda) a^*(\lambda)] dF_0(\lambda) < \infty$$

Bildet man wie üblich Äquivalenzklassen von Funktionen ( $a \sim b$  wenn  $\int_{-\pi}^{\pi} [(a(\lambda) - b(\lambda)) M_F(\lambda) (a(\lambda) - b(\lambda))^*] dF_0(\lambda) = 0$ ) dann erhält man einen komplexen Hilbertraum  $\mathbb{L}_2^c([-\pi, \pi], \mathcal{B}, F)$  mit dem inneren Produkt

$$\langle a, b \rangle_F := \int_{-\pi}^{\pi} [a(\lambda) M_F(\lambda) b(\lambda)^*] dF_0(\lambda)$$

und der Norm

$$\|a\|_F^2 := \langle a, a \rangle_F = \int_{-\pi}^{\pi} [a(\lambda) M_F(\lambda) a(\lambda)^*] dF_0(\lambda)$$

Die Konstruktion des stochastischen Integrals kann nach diesen Vorbereitungen nun eins zu eins vom skalaren Fall übernommen werden.

1. Für Stufenfunktionen der Form

$$a(\lambda) = \sum_{k=1}^m a_k \chi_{(\lambda_k, \lambda_{k+1}]}(\lambda)$$

$$a_k \in \mathbb{C}^{1 \times n}, k = 1, \dots, m; \quad -\pi = \lambda_1 < \lambda_2 < \dots < \lambda_m < \lambda_{m+1} = \pi$$

definiert man

$$\int_{-\pi}^{\pi} a(\lambda) dz(\lambda) = I(a) := \sum_{k=1}^m a_k (z(\lambda_{k+1}) - z(\lambda_k)) \in \mathbb{L}_2^c(\Omega, \mathcal{A}, P)$$

2. Mit den Eigenschaften des Prozesses  $z(\lambda)$  folgt für Stufenfunktionen  $a$  und  $b$

$$\mathbf{E}I(a) = \sum_{k=1}^m a_k \mathbf{E}(z(\lambda_{k+1}) - z(\lambda_k)) = 0$$

und

$$\begin{aligned} \langle I(a), I(b) \rangle &= \mathbf{E}I(a) \overline{I(b)} = \mathbf{E} \left( \sum_{k=1}^m a_k (z(\lambda_{k+1}) - z(\lambda_k)) \right) \overline{\left( \sum_{l=1}^m b_l (z(\lambda_{l+1}) - z(\lambda_l)) \right)} = \\ &= \sum_{k,l=1}^m a_k \mathbf{E} (z(\lambda_{k+1}) - z(\lambda_k)) (z(\lambda_{l+1}) - z(\lambda_l))^* b_l^* \\ &= \sum_{k=1}^m a_k (F(\lambda_{k+1}) - F(\lambda_k)) b_k^* \\ &= \int_{-\pi}^{\pi} [a(\lambda) M_F(\lambda) b(\lambda)^*] dF_0(\lambda) = \langle a, b \rangle_F \end{aligned}$$

3. Sei  $\mathbb{M}_2^{\text{step}} \subseteq \mathbb{L}_2^c([-\pi, \pi], \mathcal{B}, F)$  die Menge der Stufenfunktionen und  $\mathbb{M}_2 = \overline{\mathbb{M}_2^{\text{step}}}$  der Abschluss von  $\mathbb{M}_2^{\text{step}}$ . Für Funktionen  $a \in \mathbb{M}_2$  existiert eine Folge  $\{a_m \in \mathbb{M}_2^{\text{step}}\}$ , die gegen  $a$  konvergiert, d.h.  $\|a - a_m\|_F \rightarrow 0$ . Das stochastische Integral von  $a$  wird als Grenzwert (im quadratischen Mittel) der Integrale  $I(a_m)$  definiert:

$$\int_{-\pi}^{\pi} a(\lambda) dz(\lambda) = I(a) := \lim_{m \rightarrow \infty} I(a_m)$$

Der Grenzwert existiert und ist unabhängig von der Folge  $\{a_m\}$ . Das stochastische Integral ist also wohldefiniert.

4. Für  $a, b \in \mathbb{M}_2$  gilt:

$$\mathbf{E}I(a) = \mathbf{E} \lim_m I(a_m) = \lim_m \mathbf{E}I(a_m) = 0$$

$$\langle I(a), I(b) \rangle = \lim_m \langle I(a_m), I(b_m) \rangle = \lim_m \langle a_m, b_m \rangle_F = \langle a, b \rangle_F$$

4.20

5. Man kann wieder zeigen, dass  $\mathbb{M}_2 = \mathbb{L}_2^c([-\pi, \pi], \mathcal{B}, F)$  gilt. Das heißt, das stochastische Integral ist für alle (bzgl.  $F$ ) quadratisch integrierbaren (Zeilen-) Funktionen definiert. Insbesondere enthält  $\mathbb{L}_2^c([-\pi, \pi], \mathcal{B}, F)$  alle (stückweise) stetigen und alle beschränkten Funktionen.

In den nächsten Abschnitten benötigen wir noch das stochastische Integral von Matrizen Funktionen  $a : [-\pi, \pi] \rightarrow \mathbb{C}^{m \times n}$ , das wir einfach Zeilen-weise definieren:

$$\int_{-\pi}^{\pi} a(\lambda) dz(\lambda) = I(a) := \begin{pmatrix} I(a_i) \\ \vdots \\ I(a_m) \end{pmatrix}$$

wobei  $a_i \in \mathbb{L}_2([-\pi, \pi], \mathcal{B}, F)$  die Zeilen von  $a$  bezeichnen. Die Isometrie (4.20) ergibt dann für zwei Matrizen Funktionen  $a, b$

$$\mathbf{E}(I(a)I(b)^*) = \left[ \mathbf{E}I(a_i)\overline{I(b_j)} \right]_{i,j} = \left[ \int_{-\pi}^{\pi} [a_i(\lambda)M_F(\lambda)b_j(\lambda)^*] dF_0(\lambda) \right]_{i,j} = \int_{-\pi}^{\pi} [a(\lambda)M(\lambda)b(\lambda)^*] dF_0(\lambda)$$

Für Funktionen der Form  $a(\lambda) = a_0(\lambda)I_n$ , wobei  $a_0$  eine skalare Funktion ist, schreiben wir kurz

$$\int_{-\pi}^{\pi} a(\lambda) dz(\lambda) = \int_{-\pi}^{\pi} a_0(\lambda) dz(\lambda)$$

#### 4.4 Spektraldarstellung von stationären Prozessen

**Theorem 4.6.** Für jeden stationären Prozess  $(x_t | t \in \mathbb{Z})$  mit Erwartungswert gleich Null, existiert ein Prozess  $(z(\lambda) | \lambda \in [-\pi, \pi])$  mit orthogonalen Inkrementen, sodass für alle  $t \in \mathbb{Z}$

$$x_t = I(e^{i\lambda t} I_n) = \int_{-\pi}^{\pi} e^{i\lambda t} dz(\lambda) \text{ f.s.}$$

4.21

gilt. Der sogenannte Spektral-Prozess  $(z(\lambda))$  ist f.s. eindeutig durch  $(x_t)$  bestimmt.

Insbesondere gilt

$$x_0 = \int_{-\pi}^{\pi} e^{i\lambda 0} dz(\lambda) = z(\pi) \text{ f.s.}$$

Dieses Resultat zeigt, dass jeder stationäre Prozess beliebig genau durch einen harmonischen Prozess approximiert<sup>3</sup> werden kann:

$$x_t = \lim_{m \rightarrow \infty} x_t^m$$

wobei  $(x_t^m)$  ein harmonischer Prozess ist:

$$x_t^m := \sum_{k=1}^m e^{i\lambda_k^m t} (z(\lambda_{k+1}^m) - z(\lambda_k^m)); \quad \lambda_k^m = -\pi + \frac{2\pi(k-1)}{m}$$

Die Funktion  $F : [-\pi, \pi] \rightarrow \mathbb{C}^{n \times n}$ ,  $F(\lambda) = \mathbf{E}z(\lambda)z(\lambda)^*$  nennt man die *spektrale Verteilungsfunktion* von  $(x_t)$ . Wenn das durch  $F$  definierte Maß absolut stetig bzgl. des Lebesgue Maßes  $\mu$  ist, dann existiert eine Funktion  $f : [-\pi, \pi] \rightarrow \mathbb{C}^{n \times n}$  s.d.

$$F(\lambda) = \int_{-\pi}^{\lambda} f(v) dv,$$

Diese Funktion  $f$  nennt man die *spektrale Dichte* des Prozesses  $(x_t)$ . Die Verteilung  $F$  ist eindeutig bestimmt, die Dichte (wenn sie existiert) ist allerdings nur  $\mu$ -fast überall eindeutig. Das heißt, zwei Funktionen  $f(\lambda)$  und  $\tilde{f}(\lambda)$  erfüllen die Bedingung  $\int_{-\pi}^{\lambda} f(v) dv = \int_{-\pi}^{\lambda} \tilde{f}(v) dv = F(\lambda)$  für alle  $\lambda \in [-\pi, \pi]$ , dann und nur dann, wenn  $\mu\{\lambda \in [-\pi, \pi] \mid f(\lambda) \neq \tilde{f}(\lambda)\} = 0$ .

*Bemerkung 4.7.* Das stochastische Integral (bzgl. des Spektralprozesses  $z(\lambda)$ ) ist eine Isometrie zwischen dem sogenannten *Frequenzbereich* des Prozesses

$$\mathbb{L}_2^c([-\pi, \pi], \mathcal{B}, F)$$

und dem *Zeitbereich* des Prozesses

$$\mathbb{H}_x^c = \overline{\text{span}}\{x_{it} \mid i = 1, \dots, n, t \in \mathbb{Z}\} \subseteq \mathbb{L}_2^c(\Omega, \mathcal{A}, P)$$

Für dieses Resultat muss man “nur” noch zeigen, dass die Menge der komplexen Exponentialfunktionen  $\{u_k e^{i\lambda t} \mid k = 1, \dots, n, t \in \mathbb{Z}\}$  dicht in  $\mathbb{L}_2^c([-\pi, \pi], \mathcal{B}, F)$  sind. Die  $u_k \in \mathbb{C}^{1 \times n}$  bezeichnen hier die Einheitsvektoren.

#### Die Spektraldarstellung der Auto-Kovarianzfunktion

Die Autokovarianzfunktion des Prozesses  $(x_t)$  und die spektrale Verteilung (Dichte) sind folgendermaßen miteinander verknüpft

$$\begin{aligned} \gamma(s) &= \mathbf{E}x_s x_0^* = \mathbf{E}I(e^{i\lambda s} I_n) I(e^{i\lambda 0} I_n)^* = \int_{-\pi}^{\pi} \left[ (e^{i\lambda s} I_n) M_F(\lambda) (e^{i\lambda 0} I_n)^* \right] dF_0(\lambda) = \\ &= \int_{-\pi}^{\pi} e^{i\lambda s} M_F(\lambda) dF_0(\lambda) = \int_{-\pi}^{\pi} e^{i\lambda s} dF(\lambda) \end{aligned} \quad \boxed{4.22}$$

Wenn die spektrale Dichte  $f$  des Prozesses existiert, dann folgt weiters

$$\gamma(s) = \int_{-\pi}^{\pi} e^{i\lambda s} f(\lambda) d\lambda. \quad \boxed{4.23}$$

<sup>3</sup> Die Konvergenz ist punktweise aber i.A. nicht gleichmäßig. D.h. für jedes  $t \in \mathbb{Z}$  gilt  $\|x_{kt} - x_{kt}^m\| \rightarrow 0$  für  $m \rightarrow \infty$  aber  $\sup_{t \in \mathbb{Z}} \max_{k=1, \dots, n} \|x_{kt} - x_{kt}^m\|$  konvergiert i.A. nicht gegen Null.

Man kann auch zeigen, dass diese Gleichungen die spektrale Dichte (bzw. Verteilung) charakterisieren. D.h. wenn eine Funktion  $f$  die Gleichung (4.23) (für alle  $s \in \mathbb{Z}$ ) erfüllt, dann ist  $f$  die spektrale Dichte des Prozesses  $(x_t)$ .

Nicht jeder stationäre Prozess besitzt eine spektrale Dichte. Eine *hinreichende* Bedingung für die Existenz der spektralen Dichte ist

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

Wenn die Autokovarianzfunktion absolut summierbar ist, d.h. wenn  $\gamma(s)$  für  $s \rightarrow \infty$  schnell genug gegen Null konvergiert, dann existiert die spektrale Dichte. In diesem Fall kann man die Dichte mit

$$f(\lambda) = \frac{1}{2\pi} \sum_{s=-\infty}^{\infty} e^{-i\lambda s} \gamma(s). \quad \boxed{4.24}$$

berechnen. Die Autokovarianzfunktion  $\gamma$  und die spektrale Dichte sind also durch die zeitdiskreten Fourier-Transformation miteinander verknüpft.

*Bemerkung 4.8.* Jeder reguläre Prozess besitzt eine spektrale Dichte, siehe Abschnitt 3.4.

*Bemerkung 4.9.* Die Spektraldarstellung lässt sich auch für Prozesse mit Erwartungswert  $\mathbb{E}x_t \neq 0$  definieren. Allerdings hat die spektrale Verteilung  $F$  in diesem Fall eine Unstetigkeitsstelle an der Stelle  $\lambda = 0$  und es existiert also keine spektrale Dichte. Außerdem beschreibt  $F$  dann die nicht-zentrierten zweiten Momente  $\mathbb{E}x_s x_0' = \gamma(s) + \mathbb{E}x_0 (\mathbb{E}x_0)'$  und nicht die Autokovarianz-Funktion.

Eine Charakterisierung von spektralen Verteilungen bzw. spektralen Dichten gibt das folgende Theorem.

**Theorem 4.10.**

1. Eine Funktion  $F : [-\pi, \pi] \rightarrow \mathbb{C}^{n \times n}$  ist dann und nur dann die spektrale Verteilung eines stationären Prozesses, wenn

$$\begin{aligned} F(-\pi) &= 0; \quad F(\pi) < \infty \\ \lim_{\epsilon \downarrow 0} F(\lambda + \epsilon) &= F(\lambda) && (\text{rechts-stetig}) \\ (F(\lambda_2) - F(\lambda_1)) &\geq 0 \text{ für } \lambda_1 \leq \lambda_2 && (F \text{ ist monoton wachsend}) \end{aligned} \quad \boxed{4.25}$$

2. Eine Funktion  $f : [-\pi, \pi] \rightarrow \mathbb{C}^{n \times n}$  ist dann und nur dann die spektrale Verteilung eines stationären Prozesses, wenn

$$\begin{aligned} f(\lambda) &\geq 0 && \mu\text{-f.ü.} \\ \int_{-\pi}^{\pi} f(\lambda) d\lambda &< \infty \end{aligned} \quad \boxed{4.26}$$

*Beweis.* Wir zeigen nur eine Richtung. Die Eigenschaften der spektralen Verteilungsfunktion folgen direkt aus (4.17). Für die Dichte  $f$  müssen wir zeigen, dass sie f.ü. positiv semidefinit ist. Das folgt aus der Bedingung  $\int_{\lambda_1}^{\lambda_2} f(\lambda) d\lambda = F(\lambda_1) - F(\lambda_2) \geq 0$  für alle  $-\pi \leq \lambda_1 \leq \lambda_2 \leq \pi$ .

Für die andere Richtung kann man zeigen, dass die Folge  $\gamma(s) = \int_{-\pi}^{\pi} e^{i\lambda s} dF(\lambda)$  positiv semidefinit ist, wenn die Bedingungen (4.25) erfüllt sind. Es existiert also ein stationärer Prozess mit Autokovarianzfunktion  $\gamma(s)$ , dessen spektrale Verteilung daher gleich  $F$  ist. Siehe Theorem 2.5.  $\square$

Für reelle Prozesse  $(x_t)$  besitzen die spektrale Verteilung und die spektrale Dichte folgende Symmetrie Eigenschaften

$$F(\lambda_2) - F(\lambda_1) = (F((-\lambda_1)-) - F((-\lambda_2)-))' \quad \boxed{4.27}$$

$$f(\lambda) = f(-\lambda)' \quad \boxed{4.28}$$

Es genügt also das Spektrum auf  $[0, \pi]$  zu betrachten. Für den Fall einer absolut summierbaren Autokovarianzfunktion kann man (4.28) leicht zeigen:

$$f(\lambda) = \frac{1}{2\pi} \sum_{s=-\infty}^{\infty} e^{-i\lambda s} \gamma(s) \stackrel{s \rightarrow -s}{=} \frac{1}{2\pi} \sum_{s=-\infty}^{\infty} e^{i\lambda s} \gamma(-s) \stackrel{\gamma(-s) = \gamma(s)'}{=} \frac{1}{2\pi} \sum_{s=-\infty}^{\infty} e^{i\lambda s} \gamma(s)' = f(-\lambda)'$$

Für skalare Prozess gilt also insbesondere

$$f(\lambda) = f(-\lambda)$$

Betrachtet man einen gestapelten Prozess  $(z_t = (x_t', y_t')')$  und partitioniert man die spektrale Dichte (falls sie existiert) entsprechend in

$$f_z = \begin{pmatrix} f_x & f_{xy} \\ f_{yx} & f_y \end{pmatrix}$$

dann nennt man  $f_x, f_y$  die Auto-Spektren der Prozesse  $(x_t)$  bzw.  $(y_t)$  und  $f_{yx}, f_{xy}$  sind die sogenannten Kreuz-Spektren zwischen  $(y_t)$  und  $(x_t)$  bzw. zwischen  $(x_t)$  und  $(y_t)$ .

Für einen  $n$ -dimensionalen Prozess  $(x_t)$  mit spektraler Dichte  $f = (f_{ij})_{i,j=1,\dots,n}$  ist  $f_{ii}$  das Auto-Spektrum des Komponenten-Prozesses  $(x_{it})$  und  $f_{ij}$  ist das Kreuz-Spektrum zwischen den beiden (skalaren) Prozessen  $(x_{it})$  und  $(x_{jt})$ .

### Interpretation des Spektrums

Die Spektraldarstellung (4.21) zerlegt den Prozess in eine Summe (genauer Integral) von harmonischen Schwingungen. Wir betrachten jetzt ein Frequenz-Band  $(\lambda_1, \lambda_2) \subset [0, \pi]$  und wollen den Beitrag der zugehörigen Schwingungen zum Prozess genauer quantifizieren. Wir definieren die zwei Indikatorfunktionen

$$\chi_1(\lambda) = \chi_{[-\lambda_2, -\lambda_1) \cup (\lambda_1, \lambda_2]}(\lambda) \text{ und } \chi_2(\lambda) = 1 - \chi_1(\lambda)$$

und zerlegen  $x_t$  entsprechend in zwei Teile

$$x_t = \int_{-\pi}^{\pi} e^{i\lambda t} dz(\lambda) = \underbrace{\int_{-\pi}^{\pi} \chi_1(\lambda) e^{i\lambda t} dz(\lambda)}_{=: x_t^1} + \underbrace{\int_{-\pi}^{\pi} \chi_2(\lambda) e^{i\lambda t} dz(\lambda)}_{=: x_t^2}$$

Um den Prozess in *reelle* Komponenten zu zerlegen, müssen wir auch das "gespiegelte" Frequenzband  $[-\lambda_2, -\lambda_1)$  mit berücksichtigen.

$x_t^1$  ist der Anteil des Prozesses, der von den Schwingungen mit Frequenzen im Band  $(\lambda_1, \lambda_2]$  erzeugt wird und  $x_t^2$  ist der "Rest". Die beiden Komponenten sind orthogonal zueinander, da sich die entsprechenden Frequenzbereiche nicht überlappen ( $\chi_1 \chi_2 = 0$ ). Es gilt (siehe auch Gleichung (4.27))

$$\begin{aligned} \mathbf{E}x_t^1(x_t^2)^* &= \int_{-\pi}^{\pi} \underbrace{\chi_1(\lambda)e^{i\lambda t} \overline{(\chi_2(\lambda)e^{i\lambda t})}}_{=0} dF(\lambda) = 0 \\ \mathbf{E}x_t^1(x_t^1)^* &= \int_{-\pi}^{\pi} \underbrace{|\chi_1(\lambda)e^{i\lambda t}|^2}_{=\chi_1(\lambda)} dF(\lambda) = \underbrace{(F(\lambda_2) - F(\lambda_1))}_{:=\Delta F} + \underbrace{(F((- \lambda_1) -) - F((- \lambda_2) -))}_{=(\Delta F)'} = \Delta F + (\Delta F)' \\ \mathbf{E}x_t^2(x_t^2)^* &= \int_{-\pi}^{\pi} \underbrace{|\chi_2(\lambda)e^{i\lambda t}|^2}_{=\chi_2(\lambda)} dF(\lambda) \\ \gamma(0) = \mathbf{E}x_t x_t' &= \int_{-\pi}^{\pi} (\chi_1(\lambda) + \chi_2(\lambda)) dF(\lambda) = \mathbf{E}x_t^1(x_t^1)^* + \mathbf{E}x_t^2(x_t^2)^* \end{aligned}$$

*Interpretation der Auto-Spektren  $F_{kk}$  bzw.  $f_{kk}$ :*

Der Quotient

$$0 \leq \frac{2\Delta F_{kk}}{\gamma_{kk}(0)} \leq 1$$

ist der relative Anteil der Varianz (=Energie) der k-ten Komponente  $x_{kt}$ , der von den Schwingungen mit Frequenzen im Intervall  $(\lambda_1, \lambda_2]$  erklärt wird. Frequenz-Intervalle, in denen die (Auto-) spektrale Verteilung  $F_{kk}$  stark wächst (also das Inkrement  $\Delta F_{kk}$  groß ist) sind also "wichtig" für den Prozess. Existiert die spektrale Dichte, so können wir dieses Inkrement auch schreiben als

$$\Delta F = \int_{\lambda_1}^{\lambda_2} f(v) dv \approx f(\lambda_1)(\lambda_2 - \lambda_1)$$

Die Approximation  $\Delta F \approx f(\lambda_1)(\lambda_2 - \lambda_1)$  ist natürlich umso besser, je kleiner  $(\lambda_2 - \lambda_1)$  ist. "Peaks" in der Auto- spektralen Dichte  $f_{kk}$  zeigen somit "wichtige Frequenzen" (bzw. Frequenzbereiche) an.

*Interpretation der Kreuz-Spektren  $F_{kl}$  bzw.  $f_{kl}$ :*

$\Delta F_{kl}$  ist die Kovarianz zwischen  $x_{kt}^1$  und  $x_{lt}^1$ , d.h. ein Maß für die lineare Abhängigkeit zwischen  $x_{kt}$  und  $x_{lt}$  im Frequenzband  $(\lambda_1, \lambda_2]$ . Für kleine Intervalle gilt wieder  $\Delta F_{kl} \approx f_{kl}(\lambda_1)(\lambda_2 - \lambda_1)$  und wir können daher die Kreuz-spektrale Dichte  $f_{kl}(\lambda)$  als Maß für die lineare Abhängigkeit der beiden Komponenten  $x_{kt}$  und  $x_{lt}$  in der "Nähe" der Frequenz  $\lambda$  interpretieren. Ein Skalen-unabhängiges Maß für diese Abhängigkeit ist die sogenannte Kohärenz

$$\begin{aligned} C_{kl}^2 : [-\pi, \pi] &\longrightarrow [0, 1] \\ \lambda &\longmapsto \frac{|f_{kl}(\lambda)|^2}{f_{kk}(\lambda)f_{ll}(\lambda)} \end{aligned}$$

die als Frequenz-abhängige quadrierte Korrelation zwischen  $x_{kt}$  und  $x_{lt}$  interpretiert werden kann.

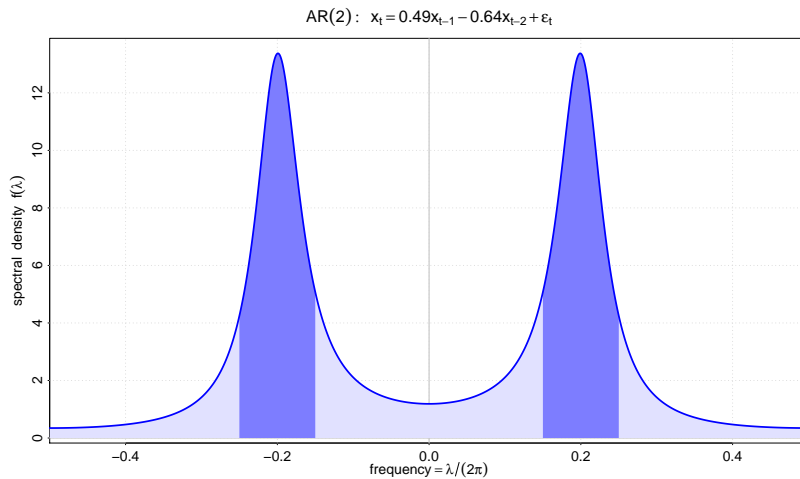


Abbildung 4.4: Spektrale Dichte eines AR(2) Prozesses.

### Beispiele

**HARMONISCHE PROZESSE:** Der Spektralprozess eines reell-wertigen, stationären, harmonischen Prozesses  $x_t = \sum_k z_k e^{i\lambda_k t}$  (siehe Abschnitt 4.2) ist

$$z(\lambda) = \sum_{\{k|\lambda_k \leq \lambda\}} z_k$$

und die spektrale Verteilungsfunktion ist

$$F(\lambda) = \mathbf{E}|z(\lambda)|^2 = \sum_{\{k|\lambda_k \leq \lambda\}} \mathbf{E}|z_k|^2$$

Die Verteilungsfunktion  $F(\cdot)$  ist eine monoton nicht fallende, rechts-stetige Treppenfunktion (mit  $F(-\pi) = 0$  und  $F(\pi) = \mathbf{E}|z(\pi)|^2 = \mathbf{E}x_t^2 < \infty$ ) und definiert daher ein *diskretes* Maß auf dem Intervall  $[-\pi, \pi]$ . Es existiert also keine spektrale Dichte. Die Sprungstellen markieren die Frequenzen  $\lambda_k$  und die Sprunghöhen die entsprechenden Amplituden, genauer gesagt den Erwartungswert der quadrierten Absolutbeträge der (zufälligen) Amplituden.

**WHITE NOISE:** Sei  $(\epsilon_t) \sim \text{WN}(\Sigma)$  ein weißes Rauschen. Die spektrale Dichte von  $(\epsilon_t)$

$$f_\epsilon(\lambda) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \gamma_\epsilon(k) e^{-i\lambda k} = \frac{1}{2\pi} \Sigma \quad 4.29$$

ist konstant (unabhängig von der Frequenz). D.h. alle Frequenzen (Frequenzbänder) sind für den Prozess gleich wichtig. In Analogie zu weißem Licht, das eine gleichmäßige Überlagerung von allen Farben (=Frequenzen) ist, nennt man diese Prozesse "weißes Rauschen" bzw. "white noise"

Ist umgekehrt die spektrale Dichte eines Prozesses  $(x_t)$  konstant ( $f(\lambda) = f_0$ ), dann folgt

$$\gamma(k) = \int_{-\pi}^{\pi} f(\lambda) e^{i\lambda k} d\lambda = f_0 \int_{-\pi}^{\pi} e^{i\lambda k} d\lambda = 0 \quad \text{für } k \neq 0$$

D.h. ein Prozess ist dann und nur dann ein white noise Prozess, wenn die spektrale Dichte konstant ist.

MA(q) PROZESSE: Sei  $(x_t = b_0\epsilon_t + \dots + b_q\epsilon_{t-q})$  ein MA(q) Prozess, wobei  $(\epsilon_t) \sim \text{WN}(\Sigma)$ . Die spektrale Dichte von  $(x_t)$  ist

$$\begin{aligned} f(\lambda) &= \frac{1}{2\pi} \sum_{k=-q}^q \gamma(k) e^{-i\lambda k} = \frac{1}{2\pi} \sum_{k=-q}^q \sum_{j=0}^q b_{j+k} e^{-i\lambda(j+k)} \Sigma b'_j e^{i\lambda j} \\ &\stackrel{(j+k) \rightarrow k}{=} \frac{1}{2\pi} \sum_{k=0}^q \sum_{j=0}^q b_k e^{-i\lambda k} \Sigma b'_j e^{i\lambda j} \\ &= \left( \sum_{j=0}^q b_j e^{-ij\lambda} \right) \left( \frac{1}{2\pi} \Sigma \right) \left( \sum_{j=0}^q b_j e^{-ij\lambda} \right)^* \end{aligned} \quad \boxed{4.30}$$

Hier haben wir, wie in Gleichung (2.4), die Koeffizienten  $b_j$  der Einfachheit halber mit Nullen fortgesetzt, d.h. wir setzen  $b_j = 0$  für  $j < 0$  und  $j > q$ .

MA( $\infty$ ) PROZESSE: Für MA( $\infty$ ) Prozesse  $x_t = \sum_{j=-\infty}^{\infty} b_j \epsilon_{t-j}$  mit absolut summierbaren Koeffizienten ( $\sum_j \|b_j\| < \infty$ ) ist die Autokovarianzfunktion auch absolut summierbar und daher besitzt  $(x_t)$  eine spektrale Dichte. Ganz analog zum MA(q) Fall erhält man

$$f(\lambda) = b(e^{-i\lambda}) f_\epsilon(\lambda) b(e^{-i\lambda})^* \quad \boxed{4.31}$$

wobei  $f_\epsilon(\lambda) = \frac{1}{2\pi} \Sigma$  die spektrale Dichte des white noise Prozesses  $(\epsilon_t)$  ist und

$$b(e^{-i\lambda}) = \sum_{j=-\infty}^{\infty} b_j e^{-i\lambda j}$$

### Schätzung des Spektrums

Die Schätzung der spektralen Verteilungsfunktion bzw. der spektralen Dichte wird in dieser Vorlesung nicht behandelt. Für ein detaillierte Darstellung dieses Schätzproblems, siehe z.B. BROCKWELL & DAVIS<sup>4</sup> (chapter 10). Einige grundlegende Bemerkungen sollen aber doch gemacht werden.

<sup>4</sup> P.J. Brockwell and R.A. Davis. *Time Series: Theory and Methods*. Springer series in statistics. Springer-Verlag, New York, 2nd edition, 1991

1. Vergleicht man die Gleichungen (4.24) und (4.11), so ist klar, dass das Periodogramm  $I(\lambda)$  (genauer gesagt  $\frac{1}{2\pi} I(\lambda)$ ) ein Schätzer für die spektrale Dichte  $f(\lambda)$  ist. Allerdings ist das Periodogramm weder erwartungstreu noch konsistent. (Es ist nur asymptotisch erwartungstreu.) Daher muss man diese "naive" Schätzung noch geeignet adaptieren, um einen konsistenten Schätzer zu bekommen.
2. Die Schätzung der spektralen Verteilungsfunktion  $F(\lambda)$  ist etwas einfacher, da  $\int_{-\pi}^{\lambda} \frac{1}{2\pi} I(v) dv$  ein konsistenter Schätzer für  $F(\lambda) = \int_{-\pi}^{\lambda} f(v) dv$  ist.



3. Als Alternative kann man auch parametrische Schätzer für die spektrale Dichte verwenden. Das heißt, man schätzt zunächst ein parametrisches Modell (z.B. ein AR, bzw. ein ARMA Modell) für den zugrunde liegenden Prozess und schätzt dann die spektrale Dichte mit Hilfe dieses geschätzten Modells.
4. Besonders wichtig ist die Schätzung von  $(2\pi)f(0) = \sum_k e^{i0k}\gamma(k) = \sum_k \gamma(k)$ , da  $(2\pi)f(0)$  gleich der asymptotischen Varianz des Mittelwertschätzers  $\bar{x}_T$  ist. Siehe Theorem 2.13. Wie schon in Punkt 1 gesagt ist der "naive" Schätzer

$$I(0) = \sum_{k=-T+1}^T \hat{\gamma}(k)$$

aber weder erwartungstreu noch konsistent.



# 5

## Lineare, dynamische Filter

Ein linearer, dynamischer Filter ist eine Abbildung der Form

$$(x_t) \mapsto A(x_t) := (y_t = \sum_{j=-\infty}^{\infty} a_j x_{t-j}), \text{ wobei } a_j \in \mathbb{R}^{m \times n}$$

Der Prozess  $(x_t | t \in \mathbb{Z})$  ist der "input" und der transformierte Prozess  $(y_t | t \in \mathbb{Z})$  ist der "output" des Filters.

Die wichtigsten Anwendungen von Filtern sind

- Filter werden verwendet, um stationäre Prozesse zu konstruieren (siehe z.B. MA Prozesse) bzw. zu analysieren.
- Filter werden als Modelle für physikalische, ökonomische, ... (zeitinvariante, lineare, stabile, dynamische) Systeme verwendet.
- Filter sind in der Zeitreihenanalyse (und Signalverarbeitung) wichtige Instrumente, die z.B. zur Schätzung von bestimmten Komponenten einer Zeitreihe verwendet werden (z.B. Trend-Schätzung bzw. Trend-Elimination).

### Backshift-Operator

Ein besonders einfacher Filter ist der *Backshift Operator* B, der definiert ist durch

$$B(x_t | t \in \mathbb{Z}) := (y_t = x_{t-1} | t \in \mathbb{Z})$$

Die  $k$ -te Potenz des Backshift Operators (für  $k \in \mathbb{Z}$ ) ist definiert durch

$$B^k(x_t | t \in \mathbb{Z}) = (y_t = x_{t-k} | t \in \mathbb{Z})$$

Damit können wir Filter auch als (formale) Potenzreihen im Backshift Operator schreiben

$$A = a(B) = \sum_{j=-\infty}^{\infty} a_j B^j$$

$$A(x_t) = \left( \sum_{j=-\infty}^{\infty} a_j B^j \right) (x_t) = \sum_{j=-\infty}^{\infty} \{a_j B^j(x_t)\} = \sum_{j=-\infty}^{\infty} \{a_j(x_{t-j})\} = \sum_{j=-\infty}^{\infty} \{(a_j x_{t-j})\} = \left( \sum_{j=-\infty}^{\infty} a_j x_{t-j} \right)$$

Filter,  $l_1$  Filter, backshift Operator B, Polynome in B (FIR Filter), z-Transformation, zusammengesetzte Filter, inverse Filter, Inverse von FIR Filtern, Analyse im Frequenzbereich, Transfer-Funktion, Amplitudengang (gain), spektrale Verteilung eines gefilterten Prozesses, Datenaufbereitung mit Filtern, Schätzung/Elimination von Trend- bzw. Saisonkomponenten mit Filtern

Einen Filter mit Koeffizienten  $(a_j | j \in \mathbb{Z})$  nennt man

- *dynamisch*, wenn  $a_j \neq 0$  für mindestens ein  $j \neq 0$
- *statisch*, wenn  $a_j = 0$  für alle  $j \neq 0$
- *kausal*, wenn  $a_j = 0$  für alle  $j < 0$
- *anti-kausal*, wenn  $a_j = 0$  für alle  $j > 0$

$B^{-1}$  ist die zu B inverse Abbildung

$$B^{-1}(B(x_t)) = (x_t)$$

$B^0$  ist die Identität

$$B^0(x_t) = (x_t)$$

Für  $k \in \mathbb{N}$  ist  $B^k$  die  $k$ -fache Hintereinanderausführung von B und  $B^{-k}$  die  $k$ -fache Hintereinanderausführung von  $B^{-1}$ . Also z.B.

$$B^2(x_t) = B(B(x_t))$$

Allgemein gilt

$$B^m(B^k(x_t)) = B^{m+k}(x_t)$$

Filter (insbesondere der Backshift Operator) operieren auf Prozessen. Trotzdem verwendet man oft eine etwas schlampige Notation und schreibt z.B. einfach

$$(a_0 + a_1 B + a_2 B^2)x_t = a_0 x_t + a_1 x_{t-1} + a_2 x_{t-2}$$

### 5.1 $l_1$ Filter

Man muss sich natürlich überlegen, für welche “inputs” bzw. welche Filter (d.h. welche Koeffizienten  $(a_j \in \mathbb{R}^{m \times n} \mid j \in \mathbb{Z})$ ) der “output” überhaupt definiert ist. Für endliche Filter, d.h. Filter der Form  $(x_t) \mapsto (y_t = \sum_{j=-K_1}^{K_2} a_j x_{t-j})$  ist der output  $(y_t)$  immer wohldefiniert, aber für Filter mit unendlich vielen Koeffizienten benötigt man Annahmen an die Filterkoeffizienten und den “input” Prozess.

**Definition 5.1.** Einen Filter  $a(B) = \sum_{j=-\infty}^{\infty} a_j B^j$  mit absolut summierbaren Koeffizienten,  $\sum_{j=-\infty}^{\infty} \|a_j\| < \infty$ , nennt man  $l_1$  Filter.

Die Wahl der Matrix Norm ist nicht wesentlich. Im Folgenden werden wir, wenn nichts anderes gesagt wird, die Spektralnorm<sup>1</sup> (2-Norm) verwenden.  $l_1$  Filter sind wichtig auf Grund des folgenden Theorems.

**Theorem 5.2.** Sei ein  $l_1$  Filter  $a(B) = \sum_j a_j B^j$  und ein “input” Prozess  $(x_t \mid t \in \mathbb{Z})$  gegeben.

1. Wenn  $(x_t)$  beschränkt ist (d.h.  $\sup_t \mathbf{E} x_t' x_t \leq c^2 < \infty$ ), dann konvergiert die Folge der Partialsummen  $\sum_{j=-N}^N a_j x_{t-j}$  im quadratischen Mittel und fast sicher. Der “output” Prozess  $(y_t = \sum_{j=-\infty}^{\infty} a_j x_{t-j})$  ist daher wohldefiniert (und beschränkt).
2. Ist  $(x_t)$  ein schwach stationärer Prozess mit Autokovarianzfunktion  $\gamma_x(x)$ , dann gilt
  - (a) der “gestapelte” Prozess  $(x_t', y_t')'$  ist stationär
  - (b) der Erwartungswert von  $y_t$  ist

$$\mathbf{E} y_t = \left( \sum_{j=-\infty}^{\infty} a_j \right) \mathbf{E} x_t \quad \boxed{5.1}$$

- (c) die Kreuz-Kovarianzfunktion zwischen  $(y_t)$  und  $(x_t)$  ist

$$\gamma_{yx}(k) = \mathbf{Cov}(y_{t+k}, x_t) = \sum_{j=-\infty}^{\infty} a_j \gamma_x(k-j) \quad \boxed{5.2}$$

- (d) und die Autokovarianzfunktion von  $(y_t)$  ist

$$\gamma_y(k) = \mathbf{Cov}(y_{t+k}, y_t) = \sum_{j,l=-\infty}^{\infty} a_j \gamma_x(k+l-j) a_l' = \sum_{m=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} a_{k+l-m} \gamma_x(m) a_l' \quad \boxed{5.3}$$

Für den Beweis dieses Theorems fassen wir noch einige nützliche Resultate für Reihen und Doppelreihen im  $\mathbb{L}_2$  zusammen.

<sup>1</sup> Die Spektralnorm einer Matrix  $A \in \mathbb{R}^{m \times n}$  ist  $\|A\| = \max_i \sigma_i(A) = \max_i \sqrt{\lambda_i(A'A)}$ , wobei  $\sigma_i(\cdot)$  die Singulärwerte und  $\lambda_i(\cdot)$  die Eigenwerte einer Matrix bezeichnen. Es gilt  $\|A\| = \|A'\|$ ,  $A'A \leq \|A\|^2 I_n$  und  $\|AB\| \leq \|A\| \|B\|$ .

$l_1$  Filter sind BIBO (bounded input - bounded output) stabil, da für beschränkte inputs  $(x_t)$  gilt

$$\sup_t \mathbf{E} y_t' y_t \leq \rho \sup_t \mathbf{E} x_t' x_t$$

wobei  $\rho < \infty$  nur vom Filter, aber nicht vom input abhängt.

**Lemma 5.3.**

1. Sei  $(u_k \in \mathbb{L}_2 \mid k \in \mathbb{Z})$  eine (zweiseitige) Folge von Zufallsvariablen, die absolut summierbar ist, d.h.  $\sum_{k=-\infty}^{\infty} \|u_k\| < \infty$ . Die Reihe  $\sum_{k=-\infty}^{\infty} u_k$  konvergiert (im  $\mathbb{L}_2$  bzw. im quadratischem Mittel). Sind  $(u_k)$  und  $(v_k)$  zwei solche Folgen,  $u_* = \sum_{k=-\infty}^{\infty} u_k$  und  $v_* = \sum_{k=-\infty}^{\infty} v_k$ , dann gilt

$$\mathbf{E}u_*v_* = \sum_{k,l=-\infty}^{\infty} \mathbf{E}u_kv_l$$

Die Doppelfolge  $(\mathbf{E}u_kv_l \mid k, l \in \mathbb{Z})$  ist absolut summierbar, die Summations-Reihenfolge spielt also keine Rolle. Eine analoge Aussage gilt für  $\mathbf{Cov}(u_*, v_*)$ .

2. Sei  $(u_{k,l} \in \mathbb{L}_2 \mid k, l \in \mathbb{Z})$  eine absolut summierbare Doppelfolge, d.h.  $\sum_{k,l=-\infty}^{\infty} \|u_{k,l}\| < \infty$ . Die entsprechende Doppelreihe konvergiert (im  $\mathbb{L}_2$  bzw. im quadratischem Mittel) und die Summationsreihen-Folge ist beliebig, d.h.

$$\sum_{k=-\infty}^{\infty} \left( \sum_{l=-\infty}^{\infty} u_{kl} \right) = \sum_{l=-\infty}^{\infty} \left( \sum_{k=-\infty}^{\infty} u_{kl} \right) = \sum_{k=-\infty}^{\infty} \left( \sum_{l=-\infty}^{\infty} u_{l,k-l} \right)$$

Diese Aussagen gelten auch für Folgen bzw. Doppelfolgen von Zufallsvektoren, wenn man  $\sqrt{\mathbf{E}u'u}$  als "Norm" für Zufallsvektoren verwendet.

*Beweis von Theorem 5.2.* Für Vektoren  $\alpha \in \mathbb{R}^n$  und Matrizen  $a \in \mathbb{R}^{m \times n}$  gilt  $\alpha' a' a \alpha \leq \alpha' \alpha \|a\|^2$  und daher folgt für  $u_j = a_j x_{t-j}$ , dass  $\sqrt{\mathbf{E}u'_j u_j} \leq \sqrt{\mathbf{E}x'_{t-j} x_{t-j} \|a_j\|^2} \leq c \|a_j\|$ . Die Folge  $(u_j = a_j x_{t-j})$  ist also im Sinne von Lemma 5.3 absolut summierbar und wir sehen, dass  $y_t = \sum_{j=-\infty}^{\infty} a_j x_{t-j} = \text{l.i.m}_{N \rightarrow \infty} \sum_{j=-N}^N a_j x_{t-j}$  existiert. Für den Beweis, dass diese Reihe auch fast sicher konvergiert, siehe z.B. BROCKWELL & DAVIS (Proposition 3.1.2)<sup>2</sup>.

Ist der input  $(x_t)$  stationär, so folgt weiters (mit dem obigem Lemma)

$$\mathbf{E}y_t = \mathbf{E} \text{l.i.m}_{N \rightarrow \infty} \sum_{j=-N}^N a_j x_{t-j} = \lim_{N \rightarrow \infty} \mathbf{E} \sum_{j=-N}^N a_j x_{t-j} = \left( \sum_{j=-\infty}^{\infty} a_j \right) \mathbf{E}x_0$$

$$\mathbf{Cov}(y_{t+k}, y_t) = \sum_{j=-\infty}^{\infty} \mathbf{Cov}(a_j x_{t+k-j}, x_t) = \sum_{j=-\infty}^{\infty} a_j \gamma_x(k-j)$$

$$\mathbf{Cov}(y_{t+k}, y_t) = \sum_{j,l=-\infty}^{\infty} \mathbf{Cov}(a_j x_{t+k-j}, a_l x_{t-l}) = \sum_{j,l=-\infty}^{\infty} a_j \gamma_x(k+l-j) a_l' = \sum_{m=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} a_{k+l-m} \gamma_x(m) a_l'$$

Der Erwartungswert und die Kovarianzen des gestapelten Prozesses sind unabhängig von  $t$  und daher ist  $(x'_t, y'_t)'$  stationär.  $\square$

**Definition 5.4.** Die  $z$ -Transformation eines Filters  $a(B) = \sum_j a_j B^j$  ist die Potenzreihe  $a(z) = \sum_{j=-\infty}^{\infty} a_j z^j$ , wobei  $z \in \mathbb{C}$ .

Für einen  $l_1$  Filter konvergiert die Potenzreihe für  $|z| = 1$  gleichmäßig gegen eine stetige Funktion. Es besteht eine eins-zu-eins Beziehung zwischen den Filter-Koeffizienten und der  $z$ -Transformation eines  $l_1$  Filters. Kennt man die  $z$ -Transformation des Filters für alle  $|z| = 1$ , dann kann man die Koeffizienten bestimmen, indem man diese Funktion in eine (i.A. zweiseitige) Potenzreihe entwickelt.

$$\sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} \underbrace{|\mathbf{E}u_kv_l|}_{\leq \|u_k\| \|v_l\|} \leq \left( \sum_{k=-\infty}^{\infty} \|u_k\| \right) \left( \sum_{k=-\infty}^{\infty} \|v_k\| \right) < \infty$$

Daher gilt

$$\begin{aligned} \sum_{k=-\infty}^{\infty} \left( \sum_{l=-\infty}^{\infty} \mathbf{E}u_kv_l \right) &= \sum_{l=-\infty}^{\infty} \left( \sum_{k=-\infty}^{\infty} \mathbf{E}u_kv_l \right) \\ &= \sum_{k=-\infty}^{\infty} \left( \sum_{l=-\infty}^{\infty} \mathbf{E}u_lv_{k-l} \right) \end{aligned}$$

<sup>2</sup> P.J. Brockwell and R.A. Davis. *Time Series: Theory and Methods*. Springer series in statistics. Springer-Verlag, New York, 2nd edition, 1991

**Definition 5.5.** Seien  $a(B) = \sum_j a_j B^j$  und  $b(B) = \sum_j b_j B^j$  zwei (kompatible)  $l_1$  Filter und  $(c_k = \sum_j b_j a_{k-j} \mid k \in \mathbb{Z})$  die Faltung der Koeffizienten. Dann nennt man  $c(B) = \sum_j c_j B^j$  den Produktfilter von  $a(B)$  und  $b(B)$ . Wir schreiben  $c(B) = b(B)a(B)$ .

**Theorem 5.6.** [Hintereinanderausführung von  $l_1$  Filtern] Seien  $a(B) = \sum_j a_j B^j$  und  $b(B) = \sum_j b_j B^j$  zwei (kompatible)  $l_1$  Filter,  $c(B) = b(B)a(B)$  der Produktfilter und  $(x_t)$  ein beschränkter Prozess.

1. Der Produktfilter  $c(B)$  ist ein  $l_1$  Filter
2. Die  $z$ -Transformation von  $c(B)$  ist gleich dem Produkt der  $z$ -Transformationen, d.h.  $c(z) = b(z)a(z)$
3. Sei  $(y_t) = a(B)(x_t)$  und  $(z_t) = b(B)(y_t) = b(B)(a(B)(x_t))$  dann gilt  $(z_t) = c(B)(x_t)$ . Das heißt die Hintereinanderausführung von  $a(B)$  und  $b(B)$  ist selbst wieder ein  $l_1$  Filter und dieser Filter ist der oben definierte Produktfilter von  $a(B)$  und  $b(B)$ .

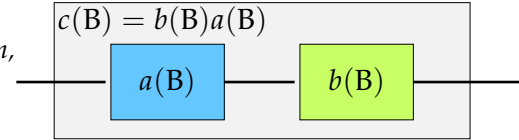


Abbildung 5.1: Block Schema für die Hintereinanderausführung von zwei Filtern

*Beweis.*

1. Der Produktfilter  $c(B)$  hat absolut summierbare Koeffizienten, da

$$\sum_k \|c_k\| = \sum_k \left\| \sum_i b_i a_{k-i} \right\| \leq \sum_k \sum_i \underbrace{\|b_i a_{k-i}\|}_{\leq \|b_i\| \|a_{k-i}\|} \leq \sum_{i,j} \|b_i\| \|a_j\| = \left( \sum_i \|b_i\| \right) \left( \sum_j \|a_j\| \right) < \infty$$

2. Die  $z$ -Transformation von  $c(B)$  ist

$$c(z) = \sum_k c_k z^k = \sum_k \sum_i b_i a_{k-i} z^k = \sum_i \sum_j b_i z^i a_j z^j = \left( \sum_i b_i z^i \right) \left( \sum_j a_j z^j \right) = b(z)a(z)$$

3. Sei  $y_t = a(B)x_t$  und  $z_t = b(B)y_t = b(B)(a(B)x_t)$  dann folgt

$$z_t = \sum_i b_i y_{t-i} = \sum_i b_i \left( \sum_j a_j x_{t-i-j} \right) = \sum_i \sum_j b_i a_j x_{t-i-j} \stackrel{(*)}{=} \sum_k \underbrace{\left( \sum_i b_i a_{k-i} \right)}_{=: c_k} x_{t-k} = \sum_k c_k x_{t-k}$$

Die Doppelfolge  $(u_{ij} = b_i a_j x_{t-i-j} \mid i, j \in \mathbb{Z})$  ist absolut summierbar, da

$$\mathbb{E} u'_{ij} u_{ij} = \mathbb{E} x'_{t-i-j} a'_j b'_i b_i a_j x_{t-i-j} \leq \mathbb{E} x'_{t-i-j} x_{t-i-j} \|b_i a_j\|^2 \leq \sup_t \mathbb{E} x'_t x_t \|b_i\|^2 \|a_j\|^2$$

und  $\sum_i \sum_j \|b_i\| \|a_j\| = \sum_i \|b_i\| \sum_j \|a_j\| < \infty$ . Daher können wir, nach Lemma 5.3, im Schritt  $(*)$  die Summations Reihenfolge ändern.

□

## 5.2 Inverse Filter

In diesem Abschnitt behandeln wir die Frage, welche  $l_1$  Filter invertierbar sind. D.h. für welche Filter kann man den input wieder rekonstruieren, indem man den output durch einen weiteren  $l_1$  filter schickt. Sei  $a(B)$  ein  $(n \times n)$  dimensionaler, quadratischer  $l_1$  Filter. Wir suchen einen  $(n \times n)$  dimensionalen  $l_1$  Filter  $b(B)$ , sodass  $b(B)a(B) = B^0 = a(B)b(B)$ . Für die zugehörigen z-Transformationen muss also  $b(z)a(z) = I_n = a(z)b(z)$  gelten. Daher muss  $a(z)$ , für alle  $|z| = 1$ , regulär sein und daher

$$\det(a(z)) \neq 0, \forall |z| = 1$$

Weiters sehen wir, dass  $b(z) = a^{-1}(z)$ . Wenn man also  $a^{-1}(z)$  in eine Potenzreihe (mit absolut summierbaren Koeffizienten) entwickelt, dann hat man damit auch die Koeffizienten des inversen Filters  $b(B) = a^{-1}(B)$  bestimmt.

**Satz 5.7.** Ein quadratischer, polynomialer Filter  $a(B) = (I_n - a_1 B \cdots - a_p B^p)$  ist dann und nur dann invertierbar, wenn die z-Transformation  $a(z) = (I_n - a_1 z \cdots - a_p z^p)$  für alle Punkte des komplexen Einheitskreises regulär ist. Das heißt, es muss  $\det(a(z)) \neq 0$  für alle  $|z| = 1$  gelten. Wenn  $\det(a(z)) \neq 0$  für alle  $|z| \leq 1$  gilt, dann ist der inverse Filter kausal.

*Beweis.* Die Notwendigkeit der Bedingung  $\det(a(z)) \neq 0$  für alle  $|z| = 1$  ist nach den obigen Erläuterungen klar. Um zu zeigen, dass diese Bedingung auch hinreichend ist, werden wir zeigen, wie der inverse Filter bestimmt werden kann.

FALL A:  $a(B) = 1 - a_1 B, a_1 \in \mathbb{R}$ :

1.  $|a_1| = 1$ : In diesem Fall existiert kein inverser Filter.
2.  $|a_1| < 1$ : Die Inverse der z-Transformation ist

$$(1 - a_1 z)^{-1} = \sum_{j \geq 0} a_1^j z^j$$

Diese Potenzreihe konvergiert für alle  $|z| < |a_1|^{-1}$  und daher insbesondere für  $|z| = 1$ . Die Koeffizienten sind absolut summierbar, da  $\sum_{j \geq 0} |a_1^j| = (1 - |a_1|)^{-1} < \infty$ . Der inverse Filter ist daher

$$(1 - a_1 B)^{-1} = \sum_{j \geq 0} a_1^j B^j$$

3.  $|a_1| > 1$ : Die Inverse der z-Transformation ist

$$(1 - a_1 z)^{-1} = \frac{-1}{a_1 z} \frac{1}{1 - (a_1 z)^{-1}} = -(a_1 z)^{-1} \sum_{j \geq 0} (a_1 z)^{-j} = - \sum_{j < 0} a_1^j z^j$$

Die obige Potenzreihe konvergiert für alle  $|z| > |a_1|^{-1}$  und daher insbesondere für  $|z| = 1$ . Die Koeffizienten sind absolut summierbar, da  $\sum_{j < 0} |a_1^{-j}| = |a_1|^{-1} (1 - |a_1^{-1}|)^{-1} < \infty$ . Der inverse Filter ist also

$$(1 - a_1 B)^{-1} = - \sum_{j < 0} a_1^j B^j$$

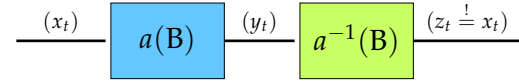


Abbildung 5.2: Block Schema für den inversen Filter

Wir verlangen immer, dass der inverse Filter wieder ein  $l_1$  Filter ist!

Beachte:  $a(z) = 1 - a_1 z = 0$  dann und nur dann, wenn  $z = a_1^{-1}$ . Daher folgt

$$a(z) \neq 0 \forall |z| = 1 \iff |a_1| \neq 1$$

und

$$a(z) \neq 0 \forall |z| \leq 1 \iff |a_1| < 1$$

FALL B:  $a(B) = 1 - a_1 B - \dots - a_p B^p$  mit  $a_j \in \mathbb{R}$  und  $a_p \neq 0$ :

Die zugehörige  $z$ -Transformation ist ein Polynom, das wir faktorisieren können als  $a(z) = (1 - z_1^{-1}z) \cdots (1 - z_p^{-1}z)$ , wobei die  $z_k$ 's die Nullstellen von  $a(z)$  bezeichnen. (Es gilt  $z_k \neq 0$ , da  $a(0) = 1 \neq 0$ .) Wenn keine Nullstelle Betrag 1 hat, dann können wir jeden Faktor, wie oben beschrieben, invertieren und in eine Potenzreihe entwickeln:

$$(1 - z_k^{-1}z)^{-1} = \begin{cases} \sum_{j \geq 0} z_k^{-j} z^j & \text{for } |z_k| > 1 \\ -\sum_{j < 0} z_k^{-j} z^j & \text{for } |z_k| < 1 \end{cases}$$

Das Produkt dieser Potenzreihen gibt eine entsprechende Potenzreihenentwicklung für  $a^{-1}(z)$ . Die Koeffizienten dieser Potenzreihe sind absolut summierbar und die Koeffizienten sind die Koeffizienten des gesuchten inversen Filters. Dieser inverse Filter ist dann und nur dann kausal, wenn alle Nullstellen  $z_k$  außerhalb des Einheitskreises liegen, also dann und nur dann, wenn  $a(z) \neq 0$  für alle  $|z| \leq 1$  gilt.

Die Nullstellen  $z_k$  können natürlich komplex sein. Man kann sich aber leicht überlegen, dass das Produkt der Potenzreihen für  $(1 - z_k^{-1}z)^{-1}$  eine Potenzreihe mit reellen Koeffizienten ergibt. Das Polynom  $a(z)$  hat reelle Koeffizienten und daher treten komplexe Nullstellen immer in zueinander komplex konjugierten Paaren auf.

FALL C:  $a(B) = I_n - a_1 B - \dots - a_p B^p$  wobei  $a_j \in \mathbb{R}^{n \times n}$  und  $a_p \neq 0$ .

Die  $z$ -Transformation kann in diesem Fall als Matrizenwertige Funktion  $a(\cdot) : \mathbb{C} \rightarrow \mathbb{C}^{n \times n}$ ,  $z \mapsto a(z)$  oder als eine Matrix mit polynomialen Elementen interpretiert werden:  $a(\cdot) = (a_{ij}(\cdot))_{i,j=1,\dots,n}$  wobei  $a_{ij}(z)$  skalare Polynome in  $z$  sind. Polynome sind spezielle rationale Funktionen und daher kann man  $a(z)$  auch als rationale Matrix, also eine Matrix deren Elemente rationale Funktionen sind, interpretieren. Mit rationalen Matrizen kann man im wesentlichen so rechnen wie mit "gewöhnlichen", reellen Matrizen. Insbesondere kann man die Inverse (wenn sie existiert) mit folgender Formel berechnen:

$$a^{-1}(z) = [\det(a(z))]^{-1} \text{adj}(a(z))$$

$d(z) = \det(a(z))$  ist ein skalares Polynom (mit  $d(0) = 1$ ) dessen Inverse wir nach dem obigen Schema in eine Potenzreihe entwickeln können. Unter der Bedingung  $d(z) = \det(a(z)) \neq 0$  für alle  $|z| = 1$  erhalten wir  $d^{-1}(z) = \sum_j d_j z^j$  und  $\sum_j |d_j| < \infty$ . Die adjungierte Matrix  $\text{adj}(a(z))$  ist eine polynomiale Matrix, also eine Matrix der Form  $\text{adj}(a(z)) = \sum_{j=0}^{p(n-1)} e_j z^j$ . Die Multiplikation der Potenzreihe  $d^{-1}(z) = \sum_j d_j z^j$  mit der polynomialen Matrix  $\text{adj}(a(z)) = \sum_{j=0}^{p(n-1)} e_j z^j$  ergibt schließlich die gesuchte Potenzreihen-Entwicklung von  $a^{-1}(z)$

$$a^{-1}(z) = \sum_j d_j z^j \left( \sum_{j=0}^{p(n-1)} e_j z^j \right) = \sum_j b_j z^j$$

Der inverse Filter ist damit

$$a^{-1}(B) = \sum_j b_j B^j$$

Die Koeffizienten  $(b_j)$  sind absolut summierbar und wir erhalten einen kausalen Filter, wenn  $\det(a(z))$  keine Nullstelle im Einheitskreis hat.  $\square$



Für den Fall eines kausalen, inversen Filters kann man die Koeffizienten auch rekursiv durch einen "Koeffizienten-Vergleich" bestimmen: Aus der Identität

$$(I_n - a_1 z - a_2 z^2 - \dots - a_p z^p)(b_0 + b_1 z + b_2 z^2 + \dots) = I_n$$

folgt

$$\begin{aligned} I_n b_0 &= I_n \implies b_0 = I_n \\ -a_1 b_0 + I_n b_1 &= 0 \implies b_1 = a_1 \\ -a_2 b_0 - a_1 b_1 + I_n b_2 &= 0 \implies b_2 = a_2 + a_1^2 \\ -a_3 b_0 - a_2 b_1 - a_1 b_2 + I_n b_3 &= 0 \implies b_3 = a_3 + a_2 a_1 + a_1 a_2 + a_1^3 \\ &\vdots \end{aligned} \quad \boxed{5.4}$$

Insbesondere gilt

$$b_0 = I_n \quad \boxed{5.5}$$

**Beispiel 5.8.** Ein einfaches Beispiel für die Berechnung der Inversen einer Polynommatrix:

$$a(z) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} z = \begin{pmatrix} 1 + a_{11}z & a_{12}z \\ a_{21}z & 1 + a_{22}z \end{pmatrix}$$

Die Inverse von  $a(z)$  ist

$$\begin{aligned} a^{-1}(z) &= \frac{1}{\det(a(z))} \text{adj}(a(z)) \\ &= \frac{1}{(1 + a_{11}z)(1 + a_{22}z) - (a_{12}z)(a_{21}z)} \begin{pmatrix} 1 + a_{22}z & -a_{12}z \\ -a_{21}z & 1 + a_{11}z \end{pmatrix} \\ &= \frac{1}{1 + (a_{11} + a_{22})z + (a_{11}a_{22} - a_{12}a_{21})z^2} \begin{pmatrix} 1 + a_{22}z & -a_{12}z \\ -a_{21}z & 1 + a_{11}z \end{pmatrix} \\ &= \begin{pmatrix} \frac{1+a_{22}z}{d(z)} & \frac{-a_{12}z}{d(z)} \\ \frac{-a_{21}z}{d(z)} & \frac{1+a_{11}z}{d(z)} \end{pmatrix} \end{aligned}$$

wobei  $d(z) = \det(a(z)) = 1 + (a_{11} + a_{22})z + (a_{11}a_{22} - a_{12}a_{21})z^2$ .

**Beispiel 5.9.** Die Inverse einer Polynom-Matrix  $a(z)$  ist i.A. eine rationale Matrix. Nur für den Fall  $\det a(z) = d_0 \neq 0$  für alle  $z \in \mathbb{C}$  erhalten wir

$$a^{-1}(z) = \frac{1}{d_0} \text{adj}(a(z))$$

also eine polynomiale Inverse. Matrizen mit dieser Eigenschaft ( $\det(a(z)) = d_0 \neq 0$ ) nennt man *unimodular*. Die Matrix

$$a(z) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} \alpha & \alpha\beta \\ -\alpha\beta^{-1} & -\alpha \end{pmatrix} z = \begin{pmatrix} 1 + \alpha z & \alpha\beta z \\ -\alpha\beta^{-1}z & 1 - \alpha z \end{pmatrix}$$

zum Beispiel ist unimodular für alle  $\alpha, \beta \in \mathbb{R}, \beta \neq 0$ . Die Inverse ist

$$a^{-1}(z) = \begin{pmatrix} 1 - \alpha z & -\alpha\beta z \\ \alpha\beta^{-1}z & 1 + \alpha z \end{pmatrix}$$

**Beispiel 5.10.** Der Differenzen Filter  $\Delta = (1 - B)$  ist nicht invertierbar, da die z-Transformation  $(1 - z)$  für  $z = 1$  gleich null ist. Man kann auch folgendermaßen argumentieren. Sei  $(x_t)$  ein stationärer Prozess mit  $\mathbf{E}x_t = \mu \neq 0$ . Der gefilterte Prozess  $y_t = \Delta x_t = x_t - x_{t-1}$  hat Erwartungswert  $\mathbf{E}y_t = 0$ . Durch die "Differenzen" Bildung ist also die Information über das Mittel von  $(x_t)$  verloren gegangen und es gibt keine Möglichkeit den Erwartungswert von  $(x_t)$  nur mit Hilfe des gefilterten Prozesses ( $y_t = x_t - x_{t-1}$ ) zu bestimmen.

### 5.3 Analyse von Filtern im Frequenzbereich

Sei  $a(B) = \sum_j a_j B^j$  wie üblich ein  $l_1$  Filter mit z-Transformation  $a(z) = \sum_j a_j z^j$ . Weiters sei  $(x_t)$  ein stationärer Prozess mit  $\mathbf{E}x_t = 0$  und Spektraldarstellung

$$x_t = \int_{-\pi}^{\pi} e^{i\lambda t} dz_x(\lambda)$$

Den gefilterten Prozess  $(y_t) = a(B)(x_t)$  können wir schreiben als

$$\begin{aligned} y_t &= \lim_{N \rightarrow \infty} \sum_{j=-N}^N a_j x_{t-j} \\ &= \lim_{N \rightarrow \infty} \sum_{j=-N}^N a_j \left[ \int_{-\pi}^{\pi} e^{i\lambda(t-j)} dz_x(\lambda) \right] \\ &= \lim_{N \rightarrow \infty} \int_{-\pi}^{\pi} e^{i\lambda t} \left[ \sum_{j=-N}^N a_j e^{-i\lambda j} \right] dz_x(\lambda) \\ &= \int_{-\pi}^{\pi} e^{i\lambda t} a(e^{-i\lambda}) dz_x(\lambda) \end{aligned}$$

Vergleichen wir nun diese Darstellung mit der Spektraldarstellung von  $(y_t)$

$$y_t = \int_{-\pi}^{\pi} e^{i\lambda t} dz_y(\lambda)$$

wobei  $z_y()$  den Spektralprozess von  $(y_t)$  bezeichnet. Da die Spektraldarstellung eindeutig ist, folgt

$$dz_y(\lambda) = a(e^{-i\lambda}) dz_x(\lambda)$$

5.6

bzw. genauer

$$z_y(\lambda) = \int_{-\pi}^{\lambda} dz_y(v) = \int_{-\pi}^{\lambda} a(e^{-i\lambda}) dz_x(v)$$

**Definition 5.11.** Die Funktion  $\lambda \in [-\pi, \pi] \mapsto a(e^{-i\lambda}) = \sum_j a_j e^{-i\lambda j}$  nennt man die *Transferfunktion* des Filters  $a(B) = \sum_j a_j B^j$ .

Die Transferfunktion ist die Zeitdiskrete Fourier Transformation der Filter Koeffizienten, siehe (4.1). Wie schon in Abschnitt 5.1 bzw. Abschnitt 4.1 diskutiert, konvergiert die Potenzreihe  $\sum_j a_j e^{-i\lambda j}$  für alle  $\lambda \in [-\pi, \pi]$  gleichmäßig gegen eine stetige Funktion. Nehmen wir nun an, dass  $(x_t)$  eine spektrale Dichte  $f_x()$  besitzt. Die spektrale Dichte kann intuitiv (aber nicht ganz exakt) als Erwartungswert

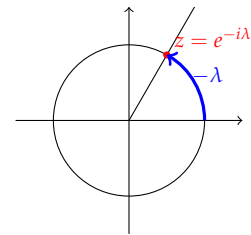


Abbildung 5.3: Zusammenhang zwischen der z-Transformation und der Transferfunktion eines Filters.

der quadrierten Inkremente des entsprechenden Spektralprozesses interpretiert werden. Also für  $f_x()$  z.B.

$$f_x(\lambda)d\lambda = \mathbf{E} [dz_x(\lambda)dz_x(\lambda)^*]$$

Aufgrund dieser Intuition und mit der Gleichung (5.6) können wir schließen, dass auch  $(y_t)$  eine spektrale Dichte besitzt und dass die spektrale Dichte  $f_y()$  und die Kreuz spektrale Dichte  $f_{yx}()$  zwischen  $(y_t)$  und  $(x_t)$  folgendermaßen bestimmt werden können:

$$f_y(\lambda)d\lambda = \mathbf{E} [dz_y(\lambda)dz_y(\lambda)^*] = a(e^{-i\lambda})\mathbf{E} [dz_x(\lambda)dz_x(\lambda)^*] a^*(e^{-i\lambda}) = a(e^{-i\lambda})f_x(\lambda)a^*(e^{-i\lambda})d\lambda$$

$$f_{yx}(\lambda)d\lambda = \mathbf{E} [dz_y(\lambda)dz_x(\lambda)^*] = a(e^{-i\lambda})\mathbf{E} [dz_x(\lambda)dz_x(\lambda)^*] = a(e^{-i\lambda})f_x(\lambda)d\lambda$$

**Theorem 5.12.** Sei  $(x_t)$  ein stationärer Prozess mit Erwartungswert  $\mathbf{E}x_t = 0$  und einer spektralen Dichte  $f_x$ . Weiters sei  $a(B)$  ein  $l_1$  Filter mit Transferfunktion  $a(e^{-i\lambda}) = \sum_j a_j e^{-i\lambda j}$ . Der gefilterte Prozess  $(y_t) = a(B)(x_t)$  besitzt eine spektrale Dichte  $f_y$  und die spektrale Dichte  $f_y()$  und die Kreuz spektrale Dichte  $f_{yx}()$  zwischen  $(y_t)$  und  $(x_t)$  sind gegeben durch

$$f_y(\lambda) = a(e^{-i\lambda})f_x(\lambda)a^*(e^{-i\lambda}) \quad \boxed{5.7}$$

$$f_{yx}(\lambda) = a(e^{-i\lambda})f_x(\lambda) \quad \boxed{5.8}$$

*Beweis.* Wir betrachten den gestapelten Prozess  $(x'_t, y'_t)'$  und dessen Spektraldarstellung

$$\begin{pmatrix} x_t \\ y_t \end{pmatrix} = \int_{-\pi}^{\pi} e^{i\lambda t} \begin{pmatrix} I_n \\ a(e^{-i\lambda}) \end{pmatrix} dz_x(\lambda)$$

Aufgrund der Eigenschaften des stochastischen Integrals erhalten wir folgende Darstellung für die Autokovarianzfunktion des gestapelten Prozesses

$$\begin{pmatrix} \gamma_x(k) & \gamma_{xy}(k) \\ \gamma_{yx}(k) & \gamma_y(k) \end{pmatrix} = \int_{-\pi}^{\pi} e^{i\lambda k} \begin{pmatrix} I_n \\ a(e^{-i\lambda}) \end{pmatrix} f_x(\lambda) \begin{pmatrix} I_n \\ a(e^{-i\lambda}) \end{pmatrix}^* d\lambda$$

Daher ist

$$\begin{pmatrix} I_n \\ a(e^{-i\lambda}) \end{pmatrix} f_x(\lambda) \begin{pmatrix} I_n \\ a(e^{-i\lambda}) \end{pmatrix}^*$$

die spektrale Dichte des gestapelten Prozesses.  $\square$

Mit diesem Theorem kann z.B. die spektrale Dichte eine  $MA(\infty)$  Prozesses  $x_t = \sum_j b_j \epsilon_{t-j}$  mit absolut summierbaren Koeffizienten unmittelbar abgeleitet werden:

$$f_x(\lambda) = b(e^{-i\lambda})f_\epsilon(\lambda)b^*(e^{-i\lambda}) \quad \boxed{5.9}$$

wobei  $b(e^{-i\lambda}) = \sum_j e^{-i\lambda j} b_j$  die Transferfunktion des Filters  $b(B) = \sum_j b_j B^j$  und  $f_\epsilon(\lambda) = \frac{1}{2\pi} \Sigma$  die spektrale Dichte des white noise Prozesses  $(\epsilon_t) \sim WN(\Sigma)$  ist. Vergleiche auch Formel (4.31).

### Interpretation der Transferfunktion

Für die Interpretation der Transferfunktion beschränken wir uns auf den skalaren Fall. Wie in Abschnitt Section 4.4 diskutiert, ist  $f_x(\lambda)\Delta\lambda$  ein Maß für den Beitrag der Schwingungen im Frequenzband<sup>3</sup>  $[\lambda, \lambda + \Delta\lambda]$  zum Prozess. Analog ist auch  $f_y(\lambda)\Delta\lambda$  zu interpretieren. Nach der Formel (5.7) gilt

$$f_y(\lambda)\Delta\lambda = |a(e^{-i\lambda})|^2 f_x(\lambda)\Delta\lambda$$

Daher werden diese Schwingungen

- für  $|a(e^{-i\lambda})| < 1$  abgeschwächt und
- für  $|a(e^{-i\lambda})| > 1$  verstärkt.

Mit Hilfe des Filters kann man also gewisse Schwingungen verstärken bzw. abschwächen. Das erklärt auch den Namen "Filter". Die Funktion  $\lambda \mapsto |a(e^{-i\lambda})|$  nennt man *gain* (oder auch Amplitudengang) des Filters.

#### Beispiel 5.13.

**BACKSHIFT OPERATOR B:** Die Transferfunktion des Backshift Operators ist  $e^{-i\lambda}$  und der gain  $|e^{-i\lambda}| = 1$  ist konstant gleich eins. Der Backshift Operator verändert also die "Amplituden" der Schwingungen nicht. (Allerdings kommt es zu einer Phasenverschiebung.)

**FILTER MIT KONSTANTEM GAIN:** Hier wollen wir zeigen, dass die Filter  $(1 - a_1^{-1}B)^{-1}(1 - a_1B)$  und  $(1 + a_1a_2^{-1}B - a_2^{-1}B^2)(1 - a_1B - a_2B^2)$  konstanten gain haben. Für  $a \in \mathbb{C}$ ,  $a \neq 0$  gilt

$$\frac{1 - az}{1 - \bar{a}^{-1}z} \frac{1 - \bar{a}z^{-1}}{1 - a^{-1}z^{-1}} = \frac{1 - az}{1 - \bar{a}^{-1}z} \frac{(-\bar{a}z^{-1})(1 - \bar{a}^{-1}z)}{(-a^{-1}z^{-1})(1 - az)} = \frac{\bar{a}}{a^{-1}} = |a|^2$$

Für  $z = e^{-i\lambda}$  (und damit  $\bar{z} = z^{-1}$ ) folgt nun

$$\left| \frac{1 - ae^{-i\lambda}}{1 - \bar{a}^{-1}e^{-i\lambda}} \right| = |a|$$

Für den Filter  $(1 - a_1^{-1}B)^{-1}(1 - a_1B)$  ist die Behauptung damit schon gezeigt. Für den zweiten Filter können wir folgendermaßen argumentieren. Sind  $z_1, z_2$  die Nullstellen von  $1 - a_1z - a_2z^2$ , dann sind  $z_1^{-1}, z_2^{-1}$  die Nullstellen von  $(1 + a_1a_2^{-1}z - a_2^{-1}z^2)$ . Faktorisieren wir die beiden Polynome entsprechend, so erhalten wir

$$\left| \frac{(1 - z_1^{-1}z)(1 - z_2^{-1}z)}{(1 - z_1z)(1 - z_2z)} \right| = |z_1^{-1}||z_2^{-1}| = |a_2|$$

Hier ist nur zu beachten, dass im Fall von komplexen Nullstellen,  $z_2 = \bar{z}_1$ .

Ist der gain des Filters konstant, dann ist die spektrale Dichte des outputs bis auf einen Faktor gleich wie die spektrale Dichte des inputs. Daher ist insbesondere der output ein white noise Prozess, wenn der input Prozess ein white noise Prozess ist.

<sup>3</sup> Diese Approximation gilt nur für kleine Frequenzbänder, d.h. für kleine  $\Delta\lambda$ .

**DIFFERENZEN FILTER**  $\Delta = (1 - B)$ : Die Transferfunktion ist  $(1 - e^{-i\lambda}) = (1 - \cos(\lambda) + i \sin(\lambda))$  und der gain ist  $|1 - e^{-i\lambda}| = \sqrt{2 - 2\cos(\lambda)}$ . Langsame Schwingungen ( $\lambda \approx 0$ ) werden also abgeschwächt, während Schwingungen mit hohen Frequenz  $\lambda \approx \pi$  verstärkt werden.

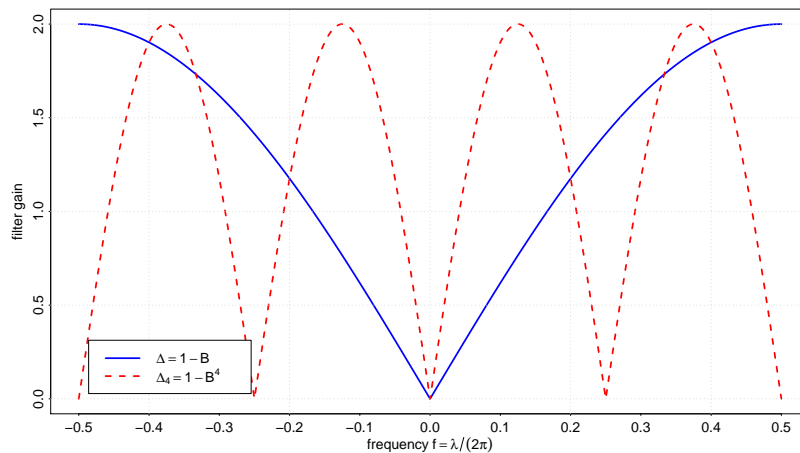


Abbildung 5.4: Differenzen-Filter

**Beispiel 5.14. FILTER ZUR TREND- BZW. SAISON- BEREINIGUNG:** Filter spielen in der Zeitreihenanalyse eine wichtige Rolle. Insbesondere kann man Filter zur Trend- bzw. Saisonbereinigung verwenden. Die Transferfunktion ist dabei ein wichtiges Instrument für die Analyse dieser Filter. Siehe Abbildung 5.5.

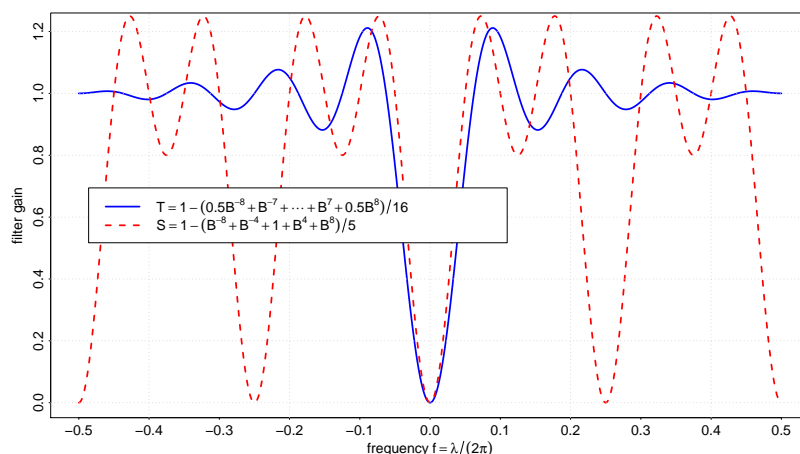


Abbildung 5.5: Filter zur Trend- bzw. Saisonbereinigung

**Beispiel 5.15. TIEFPASS, HOCHPASS UND BANDPASS FILTER:** In der Signalverarbeitung benötigt man oft Filter, die gewisse Frequenzbereiche möglichst komplett löschen und andere dagegen möglichst ungestört durchlassen. Von besonderer Bedeutung sind dabei die folgenden (idealisierten) Filter

- Tiefpass:  $a(e^{-i\lambda}) = \chi_{[-\lambda_0, \lambda_0]}(\lambda)$
- Hochpass:  $a(e^{-i\lambda}) = 1 - \chi_{[-\lambda_0, \lambda_0]}(\lambda)$
- Bandpass:  $a(e^{-i\lambda}) = \chi_{[-\lambda_2, -\lambda_1]}(\lambda) + \chi_{[\lambda_1, \lambda_2]}(\lambda)$

Die Transferfunktionen dieser idealisierten Filter sind nicht stetig, daher können sie mit einem  $l_1$  Filter nicht realisiert werden. Allerdings können diese idealisierten Transferfunktion mit  $l_1$  Filter beliebig genau approximiert werden. Siehe Abbildung 5.6.

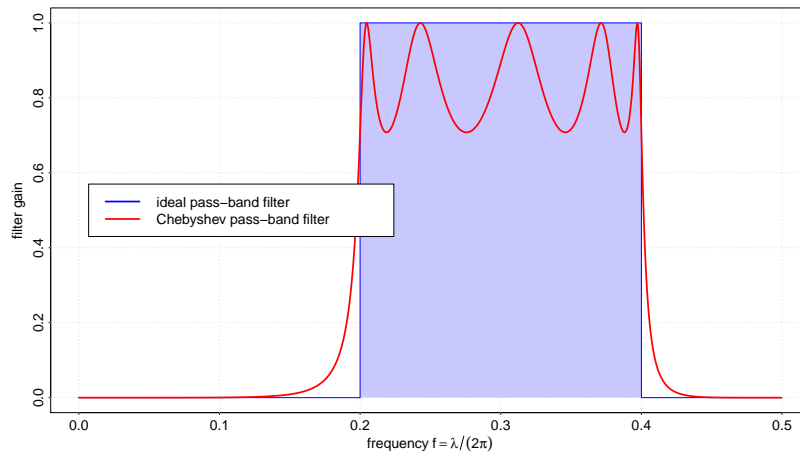


Abbildung 5.6: Bandpass Filter

# 6

## Autoregressive Prozesse

Ein AR (autoregressives) System ist ein Gleichungssystem der Form

$$x_t = a_1 x_{t-1} + \dots + a_p x_{t-p} + \epsilon_t, \quad \forall t \in \mathbb{Z} \quad \boxed{6.1}$$

wobei  $a_j \in \mathbb{R}^{n \times n}$ ,  $a_p \neq 0$  und  $(\epsilon_t) \sim \text{WN}(\Sigma)$ . Eine stationäre Lösung von (6.1), d.h. ein stationärer Prozess  $(x_t)$  der diese Gleichung(en) für alle  $t \in \mathbb{Z}$  erfüllt, ist ein sogenannter AR(p) Prozess (autoregressiver Prozess der Ordnung p).

AR Prozesse haben viele nützliche Eigenschaften und sind daher sehr "beliebt" in der Zeitreihenanalyse. Insbesondere ist die Schätzung und die Prognose besonders einfach.

Hier in diesem Abschnitt sollen folgende wichtige Fragen behandelt werden:

- Für gegebenes AR System:
  - Existiert eine stationäre Lösung? Wenn ja, ist diese eindeutig?
  - Wie kann man die (eindeutige) stationäre Lösung konstruieren?
  - Welche Eigenschaften hat diese (eindeutige) stationäre Lösung? (ACF, Spektrum, ...)
- Für einen gegebenen AR Prozess bzw. eine Zeitreihe:
  - Kann man das zugrundeliegende AR System (eindeutig) bestimmen? (Die Antwort ist natürlich essentiell für das folgende Schätzproblem.)
  - Wie kann man das AR System (d.h. die Parameter  $a_j$  und  $\Sigma$ ) aus einer gegebenen Zeitreihe schätzen?

Das AR System ist äquivalent zu folgender "Filter-Gleichung"

$$a(B)(x_t) = (I_n - a_1 B - \dots - a_p B^p)(x_t) = (\epsilon_t) \quad \boxed{6.2}$$

Die Antworten auf die obigen Fragen hängen essentiell von der z-Transformation  $a(z) = I_n - a_1 z - \dots - a_p z^p$  des AR-Filters  $a(B) = I_n - a_1 B - \dots - a_p B^p$  ab. Die z-Transformation ist eine Polynom-Matrix vom Grad p und wird oft AR-Polynom genannt. Wir werden

AR System, Stabilitätsbedingung, ACF und Spektrum, Partielle Autokorrelationsfunktion, Schätzung, Yule-Walker Gleichungen, Schätzung der Ordnung p, Informationskriterien, Prognose

Ein Gleichungssystem (6.1) nennt man auch *Differenzen Gleichung*.

Wir sind nur an *stationären* Lösungen von (6.1) interessiert.

Einen multivariaten ( $n > 1$ ) AR Prozess, nennt man auch VAR Prozess (vector autoregressive).

zeigen, dass eine eindeutige stationäre Lösung dann und nur dann existiert, wenn

$$\det(a(z)) \neq 0 \quad \forall |z| = 1 \quad \boxed{6.3}$$

Diese eindeutige stationäre Lösung hat eine (i.A. zweiseitige)  $MA(\infty)$  Darstellung. Wenn die stärkere Stabilitätsbedingung

$$\det(a(z)) \neq 0 \quad \forall |z| \leq 1 \quad \boxed{6.4}$$

erfüllt ist, dann hat die Lösung eine *kausale*  $MA(\infty)$  Darstellung.

### 6.1 Der $AR(1)$ Fall

Wir betrachten zunächst ein  $AR(1)$  System, d.h. ein System der Form

$$x_t = ax_{t-1} + \epsilon_t \quad \forall t \in \mathbb{Z} \quad \boxed{6.5}$$

wobei  $a \in \mathbb{R}^{n \times n}$  und  $(\epsilon_t) \sim WN(\Sigma)$  ein  $n$ -dimensionales weißes Rauschen ist. Das obige Gleichungssystem ist äquivalent zu

$$a(B)(x_t) = (I_n - aB)(x_t) = (\epsilon_t) \quad \boxed{6.6}$$

Die  $z$ -Transformation des Filters  $a(B) = I_n - aB$  ist die Polynom-Matrix  $a(z) = I_n - az$ . Nach Satz 5.7, ist der Filter  $a(B)$  dann und nur dann invertierbar, wenn  $\det(a(z))$  keine Nullstelle am Einheitskreis hat. Wenn  $\det(I_n - az) = 0$  gilt, dann ist  $\lambda = z^{-1}$  ein Eigenwert von  $a$ . Ist umgekehrt  $\lambda \neq 0$  ein Eigenwert von  $a$ , dann ist  $z = \lambda^{-1}$  eine Nullstelle von  $\det(a(z))$ . Daher ist der Filter  $a(B)$  genau dann invertierbar, wenn  $a$  keinen Eigenwert mit Betrag eins hat.

Wegen  $a(0) = I_n$ , kann  $z = 0$  keine Nullstelle von  $a(z) = I_n - az$  sein.

Ist der Filter  $a(B)$  invertierbar, dann ist  $(x_t) = a^{-1}(B)(\epsilon_t)$  die einzige, stationäre Lösung des  $AR$  Systems: Es gilt  $a(B)(x_t) = a(B)(a^{-1}(B)(\epsilon_t)) = (\epsilon_t)$  und daher ist  $(x_t) = a^{-1}(B)(\epsilon_t)$  eine Lösung. Sei nun  $(\tilde{x}_t)$  eine beliebige, stationäre Lösung. Wendet man auf beiden Seite der Gleichung  $a(B)(\tilde{x}_t) = (\epsilon_t)$  den inversen Filter  $a^{-1}(B)$  an, so folgt mit  $(\tilde{x}_t) = a^{-1}(B)(a(B)(\tilde{x}_t)) = a^{-1}(B)(\epsilon_t) = (x_t)$  auch die Eindeutigkeit der stationären Lösung.

FALL A: Der Spektralradius von  $a$  ist kleiner als eins, also  $\rho(a) < 1$ . In diesem Fall ist die Stabilitätsbedingung (6.4) erfüllt und  $(a(B))$  hat eine kausale Inverse. Die Koeffizienten können aus der Rekursion (5.4) einfach bestimmt werden,  $b_0 = I_n$ ,  $b_1 = a$ ,  $b_2 = a^2, \dots, b_k = a^k$ . Das heißt

Der Spektralradius  $\rho(a)$  einer quadratischen Matrix  $a$  ist das Maximum der Moduli der Eigenwerte von  $a$ .

$$a^{-1}(B) = \sum_{j \geq 0} a^j B^j$$

Die Koeffizienten  $(b_j = a^j \mid j \in \mathbb{N}_0)$  sind absolut summierbar, da für jedes  $0 < \rho(a) < \rho < 1$  eine Konstante  $c < \infty$  existiert, sodass  $\|a^j\| < c\rho^j$ . Der kausale  $MA(\infty)$  Prozess

$$x_t = \sum_{j \geq 0} a^j \epsilon_{t-j}$$



ist also die einzige, stationäre Lösung des AR Systems. Diese Lösung ist auch regulär und die  $\epsilon_t$ 's sind die Innovation von  $(x_t)$ . Das folgt wegen  $b_0 = I_n$  und  $\mathbb{H}_\epsilon(t) = \mathbb{H}_x(t)$  aus dem Korollar 3.9.

Verwendet man

$$\mathbb{E}x_t \epsilon'_{t-k} = \begin{cases} a^k \Sigma & \text{für } k \geq 0 \\ 0 & \text{für } k < 0 \end{cases}$$

so ist die Autokovarianzfunktion von  $(x_t)$  sehr einfach zu berechnen:

$$\mathbb{E}x_t = 0 \quad \boxed{6.7}$$

$$\begin{aligned} \gamma(0) &= \mathbb{E}x_t x'_t = \mathbb{E}(ax_{t-1} + \epsilon_t)(ax_{t-1} + \epsilon_t)' \\ &= a\gamma(0)a' + \Sigma = \sum_{j \geq 0} a^j \Sigma (a^j)' \end{aligned} \quad \boxed{6.8}$$

$$\begin{aligned} \gamma(k) &= \mathbb{E}x_{t+k} x'_t = \mathbb{E}(ax_{t+k-1} + \epsilon_{t+k})x'_t \\ &= a\gamma(k-1) = a^k \gamma(0) \quad \text{für } k > 0 \\ \gamma(k) &= \gamma(-k)' = \gamma(0)(a^{-k})' \quad \text{für } k < 0 \end{aligned} \quad \boxed{6.9}$$

FALL B: Der Spektralradius von  $a^{-1}$  ist kleiner als eins:  $\rho(a^{-1}) < 1$ . Hier liegen alle Nullstellen von  $\det(a(z))$  innerhalb des Einheitskreises. Die Inverse von  $(I_n - aB)$  ist antikausal und kann z.B. folgendermaßen bestimmt werden:

$$a^{-1}(B) = (I_n - aB)^{-1} = (-a^{-1}B^{-1})(I_n - a^{-1}B^{-1})^{-1} = (-aB)^{-1} \sum_{j \geq 0} (a^{-1})^j B^{-j} = - \sum_{j < 0} a^j B^j$$

Der antikausale MA( $\infty$ ) Prozess

$$x_t = - \sum_{j > 0} a^{-j} \epsilon_{t+j}$$

ist die eindeutige stationäre Lösung. Erwartungswert und Autokovarianzfunktion von  $(x_t)$  sind

$$\begin{aligned} \mathbb{E}x_t &= 0 \\ \gamma(0) &= \mathbb{E}x_{t-1} x'_{t-1} = \mathbb{E}(a^{-1}x_t - a^{-1}\epsilon_t)(a^{-1}x_t - a^{-1}\epsilon_t)' \\ &= a^{-1}\gamma(0)(a^{-1})' + a^{-1}\Sigma(a^{-1})' = \sum_{j > 0} a^{-j} \Sigma (a^{-j})' \\ \gamma(k) &= \mathbb{E}x_{t+k} x'_t = \mathbb{E}(a^{-1}x_{t+k+1} - a^{-1}\epsilon_{t+k+1})x'_t \\ &= a^{-1}\gamma(k+1) = a^k \gamma(0) \quad \text{für } k < 0 \\ \gamma(k) &= \gamma(-k)' = \gamma(0)(a^{-k})' \quad \text{für } k > 0 \end{aligned}$$

Beachte:

$$x_{t-1} = a^{-1}x_t - a^{-1}\epsilon_t$$

und

$$\mathbb{E}x_t \epsilon'_{t+k} = \begin{cases} a^{-k} \Sigma & \text{für } k > 0 \\ 0 & \text{für } k \leq 0 \end{cases}$$

Wir betrachten nun den Prozess  $(\tilde{\epsilon}_t = x_t - \tilde{a}x_{t-1})$  wobei  $\tilde{a} = \gamma(0)(a^{-1})'\gamma(0)^{-1}$ .

Dieser Prozess ist ein white noise Prozess mit Varianz  $\tilde{\Sigma} = \mathbb{E}\tilde{\epsilon}_t \tilde{\epsilon}'_t = \gamma(0) - \gamma(1)\gamma(0)^{-1}\gamma(1)'$ , da

$$\begin{aligned} \mathbb{E}\tilde{\epsilon}_t &= \mathbb{E}x_t - \tilde{a}\mathbb{E}x_{t-1} = 0 \\ \mathbb{E}\tilde{\epsilon}_{t+k} \tilde{\epsilon}'_t &= \mathbb{E}x_{t+k} x'_t - \tilde{a}\mathbb{E}x_{t+k-1} x'_t - \mathbb{E}x_{t+k} x'_{t-1} \tilde{a}' + \tilde{a}\mathbb{E}x_{t+k-1} x'_{t-1} \tilde{a}' \\ &= \gamma(k) - \tilde{a}\gamma(k-1) - \gamma(k+1)\tilde{a}' + \tilde{a}\gamma(k)\tilde{a}' \\ &= \begin{cases} 0 & \text{für } k > 0 \text{ und daher auch für } k < 0 \\ \gamma(0) - \tilde{a}\gamma(0)\tilde{a}' = \gamma(0) - \gamma(1)\gamma(0)^{-1}\gamma(1)' & \text{für } k = 0 \end{cases} \end{aligned}$$

Der Koeffizient  $\tilde{a}$  ist der Koeffizient der 1-Schrittprognose von  $x_{t+1}$  aus einem Wert  $(x_t)$ . D.h.  $\tilde{a} = \gamma(1)\gamma(0)^{-1}$  und  $\tilde{\epsilon}_{t+1} = x_{t+1} - \tilde{a}x_t$  ist der entsprechende Prognosefehler.

Es gilt

$$\tilde{a}\gamma(k) = \gamma(k+1) \quad \text{für } k \geq 0$$

$$\tilde{a}\gamma(-1) = \tilde{a}\gamma(0)\gamma(0)^{-1}a^{-1}\gamma(0) = \tilde{a}\gamma(0)\tilde{a}'$$

Wir haben somit gezeigt, dass der AR(1) Prozess  $(x_t)$  auch durch das folgende AR(1) System

$$x_t = \tilde{a}x_{t-1} + \tilde{\epsilon}_t$$

dargestellt werden kann. Für dieses System gilt  $\rho(\tilde{a}) < 1$  bzw.  $\det(I - \tilde{a}z) \neq 0$  für alle  $|z| \leq 1$ . Daher (siehe Fall A) ist der Prozess  $(x_t)$  auch regulär und die  $\tilde{\epsilon}_t$ 's sind die Innovationen von  $(x_t)$ .

Man kann diese Folgerungen auch für den Fall  $\det(\gamma(0)) = 0$  beweisen. Allerdings ist dann das "stabile" System  $x_t = \tilde{a}x_{t-1} + \tilde{\epsilon}_t$ , d.h. die Matrix  $\tilde{a}$ , nicht mehr eindeutig.

FALL C: Alle Eigenwerte von  $a$  (bzw. alle Nullstellen von  $\det(a(z))$ ) haben Betrag eins ( $\rho(a) = \rho(a^{-1}) = 1$ ). Der Einfachheit halber beschränken wir uns auf den Fall, dass  $a$  diagonalisierbar ist. Es existiert also eine Matrix  $T \in \mathbb{C}^{n \times n}$ , sodass  $T^{-1}aT = \text{diag}(\theta_1, \dots, \theta_n)$  wobei die  $\theta_k$ 's ( $|\theta_k| = 1$ ) die Eigenwerte von  $a$  sind. Wir transformieren die Zufallsvektoren und das AR(1) System (6.5) entsprechend, verwenden aber dieselbe Notation für die transformierten Größen:  $x_t \rightarrow T^{-1}x_t$ ,  $\epsilon_t \rightarrow T^{-1}\epsilon_t$  und  $a \rightarrow T^{-1}aT$ . Durch Vorwärtslösen der Gleichung (6.5) erhält man

$$x_t = a^t x_0 + \sum_{j=0}^{t-1} a^j \epsilon_{t-j} = \text{diag}(\theta_1^t, \dots, \theta_n^t) x_0 + \sum_{j=0}^{t-1} a^j \epsilon_{t-j}$$

Sei  $(x_t)$  nun eine stationäre Lösung. Für die  $k$ -te Komponente  $x_{kt} = \theta_k^t x_{k0} + \sum_{j=0}^{t-1} \theta_k^j \epsilon_{k,t-j}$  folgt somit (mit der Dreiecksungleichung)

$$2\|x_{k0}\| = \|x_{kt}\| + \|\theta_k^t x_{k0}\| \geq \|x_{kt} - \theta_k^t x_{k0}\| = \left\| \sum_{j=0}^{t-1} \theta_k^j \epsilon_{k,t-j} \right\| = \sqrt{t} \|\epsilon_{k0}\|$$

Diese Ungleichung kann nur gelten, wenn  $\epsilon_{kt} = 0$  f.s.

Es gibt also nur dann stationäre Lösungen, wenn  $\epsilon_t = 0$  (f.s.) bzw.  $\Sigma = 0$  gilt. Diese stationären Lösungen sind von der Form  $(x_t = a^t x_0)$ , wobei  $x_0$  ein quadratisch integrierbarer Zufallsvektor ist. Die stationären Lösungen sind harmonische Prozesse und die einzige reguläre Lösung ist  $(x_t = 0)$ . Natürlich erfüllt diese reguläre Lösung die Gleichungen

$$x_t = \tilde{a}x_{t-1} + \epsilon_t$$

für jede beliebige Matrix  $\tilde{a}$ . Insbesondere können wir also  $\rho(\tilde{a}) < 1$  bzw.  $\det(I - \tilde{a}z) \neq 0$  für alle  $|z| \leq 1$  verlangen.

FALL D: Der allgemeine Fall. Hier können sowohl Eigenwerte  $\lambda_i$  mit  $|\lambda_i| < 1$ ,  $|\lambda_i| > 1$  bzw.  $|\lambda_i| = 1$  auftreten. Durch eine Transformation der Zufallsvektoren und des AR(1) Systems ( $x_t \rightarrow T^{-1}x_t$ ,  $\epsilon_t \rightarrow T^{-1}\epsilon_t$  und  $a \rightarrow T^{-1}aT$ ) können wir die Matrix  $a$  in eine Block-Diagonale Form bringen,  $a = \text{diag}(a_{11}, a_{22}, a_{33})$ , wobei die drei Diagonal-Blöcke jeweils den Eigenwerten mit Betrag kleiner, größer oder gleich eins entsprechen. Entsprechend partitioniert man die Prozesse  $(x_t)$  und

Die transformierten Zufallsvariablen sind komplex-wertig. Daher sind ihre Komponenten Elemente von  $\mathbb{L}_2^{\mathbb{C}}$ .

Die Eigenwerte sind von der Form  $\theta_k = e^{i\lambda_k}$  und daher sind die Komponenten von

$$x_t = a^t x_0 = T^{-1} \text{diag}(e^{i\lambda_1 t}, \dots, e^{i\lambda_n t}) T x_0$$

Linearkombinationen der harmonischen Funktionen  $e^{i\lambda_k t}$ . Der Startwert  $x_0$  muss allerdings noch gewisse Zusatzbedingungen erfüllen, damit  $(x_t = a^t x_0)$  ein reell-wertiger, stationärer Prozess ist. Siehe Satz 4.4.

Natürlich müssen nicht immer alle drei Blöcke auftreten.

$(\epsilon_t)$  und diskutiert dann diese Teilprozesse getrennt wie oben im Fall A, B oder C. Die wesentlichen Resultate finden sich im Theorem Theorem 6.2.

## 6.2 Der AR(p) Fall

Der AR(p) Fall kann mit folgendem Trick auf den AR(1) Fall zurückgeführt werden. Wir betrachten den gestapelten Prozess  $x_t^p = (x_t', \dots, x_{t+1-p}')'$ . Man sieht leicht, dass  $(x_t)$  dann und nur dann eine Lösung von (6.1) ist, wenn der gestapelte Prozess eine Lösung des AR(1) Systems

$$x_t^p = Ax_{t-1}^p + B\epsilon_t \quad \boxed{6.10}$$

ist, wobei

$$B = \begin{pmatrix} I_n \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \in \mathbb{R}^{pn \times n}, \quad A = \begin{pmatrix} a_1 & a_2 & \cdots & a_{p-1} & a_p \\ I_n & 0 & \cdots & 0 & 0 \\ 0 & I_n & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & I_n & 0 \end{pmatrix} \in \mathbb{R}^{np \times np} \quad \boxed{6.11}$$

Die Matrix  $A$  nennt man auch *Begleitmatrix* (*companion matrix*) der Polynom Matrix  $a(z)$ . Es ist auch klar, dass  $(x_t)$  dann und nur dann stationär (regulär) ist, wenn  $(x_t^p)$  stationär (bzw. regulär ist). Die Eigenwerte der Matrix  $A$  hängen folgendermaßen mit den Nullstellen von  $\det(a(z))$  zusammen:

**Lemma 6.1.** Die folgenden drei Aussagen sind äquivalent<sup>1</sup>

1.  $\det(I - a_1 z - \cdots - a_p z^p) = 0$
2.  $\det(I - Az) = 0$
3.  $(1/z)$  ist ein Eigenwert von  $A$  (ungleich Null).

*Beweis.*

(2)  $\iff$  (3): Aus  $\det(I - Az) = 0$  folgt  $z \neq 0$ . Daher ist  $\det(I - Az) = 0$  äquivalent zu  $\det(\frac{1}{z}I - A) = 0$ .

(1)  $\iff$  (3): Sei  $\lambda$  ein Eigenwert von  $A$  und  $c = (c_1, \dots, c_p) \in \mathbb{R}^{np \times 1}$ ,  $c \neq 0$  ein zugehöriger Links-Eigenvektor. Das heißt  $cA = \lambda c$  und daher

$$\begin{aligned} c_2 &= \lambda c_1 - c_1 a_1 = c_1 (\lambda I_n - a_1) \\ c_3 &= \lambda c_2 - c_1 a_2 = c_1 (\lambda^2 I_n - \lambda a_1 - a_2) \\ &\vdots \\ c_p &= \lambda c_{p-1} - c_1 a_{p-1} = c_1 (\lambda^{p-1} I_n - \lambda^{p-2} a_1 - \cdots - a_{p-1}) \\ 0 = \lambda c_p - c_1 a_p &= c_1 (\lambda^p I_n - \lambda^{p-1} a_1 - \cdots - a_p) \end{aligned}$$

<sup>1</sup> Daher ist die Stabilitätsbedingung für das AR(p) System (6.1) äquivalent zu der Stabilitätsbedingung für das gestapelte AR(1) System (6.10).

Es muss  $c_1 \neq 0$  gelten, da sonst  $c = 0$  wäre. Für  $\lambda \neq 0$  und  $z = 1/\lambda$  folgt daher  $c_1(I - a_1z - \dots - a_pz^p) = 0$ .

Gilt umgekehrt  $c_1(I - a_1z - \dots - a_pz^p) = 0$  für ein  $z \in \mathbb{C}$  und  $c_1 \in \mathbb{R}^n$ ,  $c_1 \neq 0$ , dann setzen wir  $\lambda = 1/z$  und

$$c = c_1(I_n, (\lambda I_n - a_1), \dots, (\lambda^{p-1}I_n - \lambda^{p-2}a_1 - \dots - a_{p-1})).$$

Dieser Vektor  $c$  ist ein Links-Eigenvektor von  $A$  zum Eigenwert  $\lambda$ .

□

Damit kann man nun im Prinzip den AR(p) Fall analog zum AR(1) diskutieren. Die wesentlichen Resultate sind im folgendem Theorem zusammengefasst.

**Theorem 6.2.** Sei ein AR System (6.1) gegeben und  $a(B) = (I - a_1B - \dots - a_pB^p)$  und  $a(z) = (I - a_1z - \dots - a_pz^p)$ .

1. Das AR System hat dann und nur dann unendlich viele Lösungen, wenn das AR Polynom  $a(z)$  nicht unimodular ist, d.h. wenn  $\det(a(z))$  nicht konstant ist.
2. Das AR System besitzt dann und nur dann eine eindeutige stationäre Lösung, wenn

$$\det(a(z)) \neq 0 \quad \forall |z| = 1$$

6.12

Diese (eindeutige) stationäre Lösung ist regulär und von der Form

$$x_t = a^{-1}(B)\epsilon_t = \sum_{j=-\infty}^{\infty} b_j \epsilon_{t-j}$$

3. Wenn die Stabilitätsbedingung

$$\det(a(z)) \neq 0 \quad \forall |z| \leq 1$$

6.13

erfüllt ist, dann ist die (eindeutige) stationäre Lösung ein kausaler MA( $\infty$ ) Prozess

$$x_t = a^{-1}(B)\epsilon_t = \sum_{j=0}^{\infty} b_j \epsilon_{t-j}$$

Der Prozess  $(x_t)$  ist regulär und  $(\epsilon_t)$  sind die Innovationen von  $(x_t)$ .

4. Ist die Bedingung (6.12) nicht erfüllt, dann gibt es nur dann stationäre Lösungen, wenn  $\Sigma = \mathbf{E}\epsilon_t \epsilon_t'$  singulär ist. Falls es stationäre Lösungen gibt, dann gibt es immer unendlich viele stationäre Lösungen. Allerdings gibt es unter diesen stationären Lösungen nur eine reguläre Lösung.
5. Ist  $(x_t)$  ein regulärer AR Prozess, dann existiert immer ein zugehöriges AR System, das die Stabilitätsbedingung erfüllt. Dieses AR System ist eindeutig, falls  $\Gamma_p = \mathbf{Var}(x_t^p)$  positiv definit ist. Die Bedingung  $\Gamma_p > 0$  ist insbesondere erfüllt, wenn die Varianz der Innovationen von  $(x)$  positiv definit ist.

*Beweis.* Wir werden dieses Theorem nicht in allen Details beweisen, sondern manchmal nur auf die entsprechende Diskussion für den AR(1) Fall verweisen.

PUNKT 1: Ist die z-Transformation  $a(z)$  unimodular<sup>2</sup>, siehe Beispiel 5.9, dann ist der inverse Filter  $a^{-1}(z)$  polynomial. Durch Anwenden des inversen Filter auf beiden Seiten der Gleichung (6.2) erhalten wir

$$(x_t) = a^{-1}(B)a(B)(x_t) = a^{-1}(B)(\epsilon_t)$$

Es existiert also nur eine Lösung und diese Lösung ist ein kausaler MA(q) Prozess. Auch die Umkehrung lässt sich relativ leicht zeigen. Insbesondere im Fall  $\det(a_p) \neq 0$  sieht man, dass man für jede Wahl von "Startwerten"  $x_{-p+1}, \dots, x_0$  eine Lösung konstruieren kann, indem das AR System rekursiv vorwärts

$$x_t = a_1 x_{t-1} + \dots + a_p x_{t-p} + \epsilon_t \text{ für } t \geq 1$$

und rückwärts

$$x_{t-p} = a_p^{-1}(x_t - a_1 x_{t-1} - \dots - a_{p-1} x_{t-p+1} - \epsilon_t) \text{ für } t \leq 0$$

löst.

PUNKT 2: In Satz 5.7 wurde gezeigt, dass  $a(B)$  dann und nur dann invertierbar ist, wenn (6.12) erfüllt ist. Nehmen wir nun an, dass der inverse Filter  $a^{-1}(B) = \sum_{j=-\infty}^{\infty} b_j B^j$  existiert. Dann folgt

$$a(B)(a^{-1}(B)(\epsilon_t)) = (\epsilon_t)$$

Das heißt  $(x_t) = a^{-1}(B)(\epsilon_t) = (\sum_j b_j \epsilon_{t-j})$  ist eine stationäre Lösung. Sei nun  $(\tilde{x}_t)$  eine beliebige stationäre Lösung, dann folgt mit

$$(\tilde{x}_t) = a^{-1}(B)(a(B)(\tilde{x}_t)) = a^{-1}(B)(\epsilon_t) = (x_t)$$

die Eindeutigkeit der stationären Lösung. Wie im Fall  $p = 1$ , kann man nun zeigen, dass  $(x_t)$  eine äquivalente AR Darstellung besitzt, die die Stabilitätsbedingung erfüllt. Daher ist  $(x_t)$  nach Punkt 3) regulär.

PUNKT 3: Der inverse Filter  $a^{-1}(B)$  ist, nach Satz 5.7 dann und nur dann kausal, wenn die Stabilitätsbedingung erfüllt ist. Aus  $x_t = \sum_{j \geq 0} b_j \epsilon_{t-j}$  und  $\epsilon_t = x_t - a_1 x_{t-1} - \dots - a_p x_{t-p}$  folgt  $\mathbb{H}_\epsilon(t) = \mathbb{H}_x(t)$ . Zusammen mit  $b_0 = I_n$  (siehe Gleichung (5.5)) zeigt nun das Korollar 3.9, dass  $(x_t)$  regulär ist und dass  $(\epsilon_t)$  die Innovationen sind.

PUNKTE 4 UND 5: Siehe die Diskussion zum AR(1) Fall. Die Frage der Eindeutigkeit des "stabilen" Systems wird in Abschnitt 6.7 noch genauer behandelt werden.  $\square$

Im Folgenden betrachten wir nur *reguläre* AR Prozesse und setzen daher auch o.E.d.A immer die Stabilitätsbedingung voraus.

<sup>2</sup> In diesem Fall ist natürlich auch die Stabilitätsbedingung erfüllt.

### 6.3 Prognose

Die Prognose von AR Prozessen ist besonders einfach. Aufgrund der Stabilitätsbedingung ist  $(x_t)$  ein *kausaler* MA( $\infty$ ) Prozess und es gilt (siehe Beweis von Punkt 3 des Theorems 6.2)

$$\begin{aligned}\mathbb{H}_x(t) &= \mathbb{H}_{\epsilon(t)} \\ \epsilon_s &\perp \mathbb{H}_x(t) = \mathbb{H}_{\epsilon(t)} \text{ für alle } s > t\end{aligned}$$

Für die Ein-Schrittprognose aus  $k \geq p$  vergangenen Werten verwenden wir die Linearität der Projektion (Prognose). Sei  $\mathbb{M} = \text{span}\{x_t, \dots, x_{t+1-k}\} \subseteq \mathbb{H}_x(t)$ :

$$\begin{aligned}\hat{x}_{t,1,k} = \mathbf{P}_{\mathbb{M}}(x_{t+1}) &= a_1 \underbrace{\mathbf{P}_{\mathbb{M}}x_t}_{x_t} + \dots + a_p \underbrace{\mathbf{P}_{\mathbb{M}}x_{t+1-p}}_{x_{t+1-p}} + \underbrace{\mathbf{P}_{\mathbb{M}}\epsilon_{t+1}}_{=0} \\ &= a_1 x_t + \dots + a_p x_{t+1-p} \\ \hat{u}_{t,1,k} = (x_{t+1} - \hat{x}_{t,1,k}) &= \epsilon_{t+1}\end{aligned}$$

Die Koeffizienten der optimalen Ein-Schrittprognose sind also genau die Koeffizienten der (stabilen) AR Darstellung und die  $\epsilon_t$ 's sind die Ein-Schrittprognosefehler. Für die Prognose werden nur  $p$  Werte benötigt. Das heißt auch, dass die Prognose aus der unendlichen Vergangenheit mit der Prognose aus den letzten  $p$  Werten übereinstimmt<sup>3</sup> und dass die  $\epsilon_t$ 's die Innovationen des Prozesses sind.

Für  $k < p$  ist  $a_1 x_t + a_k x_{t+1-k}$  natürlich nicht mehr die optimale Prognose!

$$\begin{aligned}\hat{x}_{t,1} &= \hat{x}_{t,1,k} = \hat{x}_{t,1,p} \\ \epsilon_t = \hat{u}_{t,1} &= \hat{u}_{t,1,k} = \hat{u}_{t,1,p} \text{ für } k \geq p \\ \Sigma &= \Sigma_1 = \Sigma_{1,k} = \Sigma_{1,p}\end{aligned}$$

<sup>3</sup> Diese Eigenschaft ist ein *Charakteristikum* von AR Prozessen. D.h. ein stationärer Prozess  $(x_t)$  ist dann und nur dann ein AR Prozess, wenn die Prognose aus der unendlichen Vergangenheit nur von endlich vielen Werten abhängt.

Die Mehrschritt-Prognose kann einfach rekursiv bestimmt werden. Für  $h = 2$  und  $h = 3$  erhalten wir z.B.

$$\begin{aligned}\hat{x}_{t,2,k} = \mathbf{P}_{\mathbb{M}}(x_{t+2}) &= a_1 \underbrace{\mathbf{P}_{\mathbb{M}}x_{t+1}}_{\hat{x}_{t,1,k}} + a_2 \underbrace{\mathbf{P}_{\mathbb{M}}x_t}_{x_t} + \dots + a_p \underbrace{\mathbf{P}_{\mathbb{M}}x_{t+2-p}}_{x_{t+2-p}} + \underbrace{\mathbf{P}_{\mathbb{M}}\epsilon_{t+2}}_{=0} \\ &= a_1 \hat{x}_{t,1,k} + a_2 x_t + \dots + a_p x_{t+2-p} \\ \hat{u}_{t,2,k} = (x_{t+2} - \hat{x}_{t,2,k}) &= \epsilon_{t+2} + a_1 \hat{u}_{t,1,k} = \epsilon_{t+2} + a_1 \epsilon_{t+1} \\ \hat{x}_{t,3,k} = \mathbf{P}_{\mathbb{M}}(x_{t+3}) &= a_1 \underbrace{\mathbf{P}_{\mathbb{M}}x_{t+2}}_{\hat{x}_{t,2,k}} + a_2 \underbrace{\mathbf{P}_{\mathbb{M}}x_{t+1}}_{\hat{x}_{t,1,k}} + \dots + a_p \underbrace{\mathbf{P}_{\mathbb{M}}x_{t+3-p}}_{x_{t+3-p}} + \underbrace{\mathbf{P}_{\mathbb{M}}\epsilon_{t+3}}_{=0} \\ &= a_1 \hat{x}_{t,2,k} + a_2 \hat{x}_{t,1,k} + a_3 x_t + \dots + a_p x_{t+3-p} \\ \hat{u}_{t,3,k} = (x_{t+3} - \hat{x}_{t,3,k}) &= \epsilon_{t+3} + a_1 \hat{u}_{t,2,k} + a_2 \hat{u}_{t,1,k} = \epsilon_{t+3} + a_1 \epsilon_{t+2} + (a_1^2 + a_2) \epsilon_{t+1}\end{aligned}$$

Auch für die Mehrschrittprognose genügt es also die letzten  $p$  Werte zu verwenden. Man kann sich auch leicht überzeugen, dass die  $h$ -Schrittprognosefehler gegeben sind durch  $\hat{u}_{t,h} = \hat{u}_{t,h,k} = \sum_{j=0}^{h-1} b_j \epsilon_{t+h-j}$ , siehe auch Korollar 3.9.

Der Vektor  $x_t^p$  enthält alle Informationen, die für die Prognose der *gesamten* Zukunft des Prozesses notwendig sind.

### 6.4 Autokovarianzfunktion und spektrale Dichte

Multipliziert man die Definitionsgleichung  $x_t = a_1 x_{t-1} + \dots + a_p x_{t-p} + \epsilon_t$  des AR Prozesses  $(x_t)$  von rechts mit  $x_{t-j}'$  und bildet auf beiden

Seiten der Gleichung den Erwartungswert, so erhält man die sogenannten *Yule-Walker Gleichungen*:

$$\gamma(0) = a_1\gamma(-1) + \dots + a_p\gamma(-p) + \Sigma \quad \text{für } j = 0 \quad \boxed{6.14}$$

$$\gamma(j) = a_1\gamma(j-1) + \dots + a_p\gamma(j-p) \quad \text{für } j > 0 \quad \boxed{6.15}$$

Hierbei verwendet man  $x_t = \sum_{k \geq 0} b_k \epsilon_{t-k}$  und  $b_0 = I_n$  (nach Gleichung (5.5)) und daher

$$\mathbf{E}\epsilon_t x'_{t-j} = \begin{cases} 0 & \text{für } j > 0 \\ \Sigma b'_0 = \Sigma & \text{für } j = 0 \\ \Sigma b'_j & \text{für } j < 0 \end{cases}$$

Die Yule-Walker Gleichungen repräsentieren den Zusammenhang zwischen der Autokovarianzfunktion des Prozesses und den Parametern  $(a_1, \dots, a_p, \Sigma)$  des zugrunde liegenden AR Systems. Man kann also einerseits, für gegebene Parameter die Autokovarianzfunktion und andererseits für gegebene Autokovarianzfunktion die Parameter bestimmen. Setzt man in die Yule-Walker Gleichungen die geschätzten Autokovarianzen  $\hat{\gamma}(k)$  ein und löst die Gleichungen nach den Parametern, so erhält man die sogenannten Yule Walker Schätzer für die System Parameter. Siehe Abschnitt 6.7.

#### *Autokovarianzfunktion*

Zunächst wollen wir die Autokovarianzfunktion  $\gamma(k)$  des AR Prozesses  $(x_t)$  bestimmen. Dazu betrachten wir, wie in Abschnitt 6.2 den gestapelten Prozess  $(x_t^p)$ , der vom AR(1) System

$$x_t^p = A x_{t-1}^p + B \epsilon_t \quad \boxed{6.16}$$

$$B = \begin{pmatrix} I_n \\ 0 \\ 0 \\ \dots \\ 0 \end{pmatrix}, \quad A = \begin{pmatrix} a_1 & a_2 & \dots & a_{p-1} & a_p \\ I_n & 0 & \dots & 0 & 0 \\ 0 & I_n & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & I_n & 0 \end{pmatrix} \quad \boxed{6.17}$$

erzeugt wird. Die Stabilitätsbedingung (6.13) ist äquivalent zu der Bedingung  $\rho(A) < 1$ . Die Autokovarianzfunktion von  $(x_t^p)$  kann daher mit den Gleichungen (6.8) und (6.9) berechnet werden.

$$\Gamma_p = A \Gamma_p A' + B \Sigma B' \quad \boxed{6.18}$$

$$\Gamma_p(j) = A^j \Gamma_p \quad \text{für } j > 0 \quad \boxed{6.19}$$

Hier bezeichnet  $\Gamma_p(k)$  die Autokovarianz von  $x_t^p$  zum lag  $k$ , d.h.

$$\Gamma_p(j) = \mathbf{E} x_{t+j}^p (x_t^p)' = (\gamma(j+l-k))_{k,l=1,\dots,n}$$

Man kann sich leicht überzeugen, dass die Gleichungen (6.18), (6.19) (algebraisch) äquivalent sind zu den Yule-Walker Gleichungen (6.14)

und (6.15). Da der Spektralradius von  $A$  kleiner als eins ist, hat (6.18) eine *eindeutige* Lösung

$$\Gamma_p = \sum_{k \geq 0} A^k B \Sigma B' (A^k)' \quad (6.20)$$

Die Autokovarianzen  $\gamma(j)$  für  $j \geq p$  können dann rekursiv aus (6.19) bzw. (6.15) bestimmt werden.

### Spektrale Dichte

Die spektrale Dichte eines (regulären) AR Prozess folgt sofort mit Theorem 5.12:

$$f(\lambda) = \frac{1}{2\pi} a^{-1}(e^{-i\lambda}) \Sigma (a^{-1}(e^{-i\lambda}))^* \quad (6.21)$$

Im skalaren Fall (mit  $\epsilon_t \sim \text{WN}(\sigma^2)$ ) erhält man eine etwas einfachere Darstellung

$$f(\lambda) = \frac{\sigma^2}{2\pi |a(e^{-i\lambda})|^2} \quad (6.22)$$

Es ist klar, dass die spektrale Dichte im wesentlichen von der Position der Nullstellen des AR Polynoms  $a(z)$  bestimmt wird. Insbesondere erzeugen Nullstellen nahe am Einheitskreis "peaks" im Spektrum.

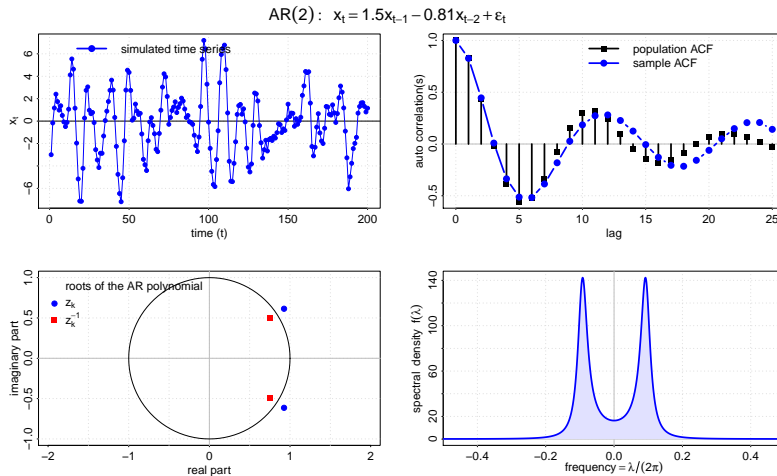


Abbildung 6.1: AR(2) Prozess  $x_t = 1.5x_{t-1} - 0.81x_{t-2} + \epsilon_t$

## 6.5 Granger Kausalität

GRANGER<sup>4</sup> hat 1969 eine Definition für Kausalität präsentiert, die wesentlich auf Prognosen beruht. Er sagt, eine Variable  $x_2$  hat einen kausalen Einfluss auf eine Variable  $x_1$ , wenn die Prognose von  $x_1$  durch Verwendung der Variablen  $x_2$  verbessert wird. Um diese Idee etwas genauer zu definieren, betrachtet man einerseits die Informationsmenge  $\mathbb{I}_t$ , die zum Zeitpunkt  $t$  ohne Informationen über die Variable  $x_2$  zur Verfügung steht und andererseits  $\mathbb{I}_t \cup \mathbb{X}_{2t}$  die Information bis  $t$  inklusive der Informationen über  $x_2$ . Die Prognose für  $x_{1,t+h}$ , gegeben diese Informationsmengen sei  $\hat{x}_{1,t+h|\mathbb{I}_t}$  bzw.



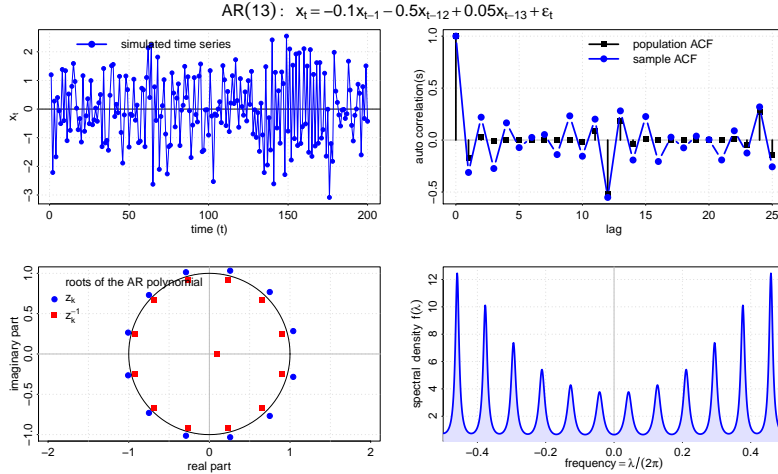


Abbildung 6.2: AR(13) Prozess  $x_t = -0.1x_t - 0.5x_{t-12} + 0.05x_{t-13} + \epsilon_t$

$\hat{x}_{1,t+h} | \mathbb{I}_t \cup \mathbb{X}_{2t}$ . Man sagt  $x_2$  ist *Granger-nicht-kausal* für  $x_1$ , wenn

$$\hat{x}_{1,t+h} | \mathbb{I}_t = \hat{x}_{1,t+h} | \mathbb{I}_t \cup \mathbb{X}_{2t} \quad \text{für alle } h > 0$$

Umgekehrt sagt man,  $x_2$  ist *Granger-kausal* für  $x_1$ , wenn die Hinzunahme der Information über  $x_2$  die Prognose für  $x_1$  (für ein  $h > 0$ ) verbessert.

Um diese Idee operabel zu machen, d.h. um Test-Prozeduren für Granger-Kausalität zu entwickeln, muss man die Begriffe "Information" bzw. Prognose natürlich etwas genauer spezifizieren. Besonders einfach wird diese Diskussion, wenn man sich auf lineare Prognose beschränkt und die relevanten Variablen mit einem (V)AR Prozess modelliert.

Wir betrachten einen  $n$ -dimensionalen, regulären AR Prozess  $(x_t)$ , der folgendermaßen partitioniert wird  $x_t = (x_{1t}, x_{2t}, x'_{3t})'$ . Hier sind  $x_{1t}$  und  $x_{2t}$  die Werte der beiden Variablen  $x_1$  und  $x_2$  (zum Zeitpunkt  $t$ ), deren kausalen Zusammenhang wir untersuchen wollen. Im  $(n-2)$  dimensionalen Vektor  $x'_{3t}$  stecken alle anderen relevanten Variablen.

$$x_t = \begin{pmatrix} x_{1t} \\ x_{2t} \\ x_{3t} \end{pmatrix} = \sum_{j=1}^p \begin{pmatrix} a_{11,j} & a_{12,j} & a_{13,j} \\ a_{21,j} & a_{22,j} & a_{23,j} \\ a_{31,j} & a_{32,j} & a_{33,j} \end{pmatrix} \begin{pmatrix} x_{1,t-j} \\ x_{2,t-j} \\ x_{3,t-j} \end{pmatrix} + \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \\ \epsilon_{3t} \end{pmatrix}$$

Die  $h$ -Schritt Prognose für  $x_{t+h}$  aus der unendlichen Vergangenheit sei

$$\hat{x}_{t,h} = c_1^h x_t + \dots + c_p^h x_{t+1-p}$$

Die Koeffizienten  $c_j^h$  kann man, wie in Abschnitt 6.3 diskutiert, rekursiv berechnen

$$\begin{aligned} (c_1^1, \dots, c_p^1) &= (a_1, \dots, a_p) \\ (c_1^2, \dots, c_p^2) &= (a_2, \dots, a_p, 0) + a_1(a_1, \dots, a_p) \\ &\vdots \\ (c_1^{h+1}, \dots, c_p^{h+1}) &= (c_2^h, \dots, c_p^h, 0) + c_1^h(a_1, \dots, a_p) \end{aligned}$$

Die Wahl der Variablen  $x_3$  hat großen Einfluss auf die Frage der Granger Kausalität, bzw. nicht-Kausalität.

$$\begin{aligned} x_{t+h+1} &= \hat{x}_{t+1,h} + \hat{u}_{t+1,h} \\ &= c_1^h x_{t+1} + c_2^h x_t + \dots + c_p^h x_{t+2-p} + \hat{u}_{t+1,h} \\ &= c_1^h a_1 x_t + \dots + c_1^h a_p x_{t+1-p} + c_1^h \epsilon_{t+1} + \\ &\quad c_2^h x_t + \dots + c_p^h x_{t+2-p} + \hat{u}_{t+1,h} \end{aligned}$$

Die Variable  $x_2$  ist daher dann und nur dann Granger nicht-kausal für  $x_1$ , wenn

$$c_{12,j}^h = 0 \text{ für } h \geq 0, \text{ und } j = 1, \dots, p$$

Das sind komplizierte, nicht-lineare Bedingungen an die AR Koeffizienten  $a_1, \dots, a_p$ , die nicht so einfach zu testen sind. Es ist aber klar, dass

$$a_{12,j} = 0 \text{ für } j = 1, \dots, p$$

6.23

zumindest *notwendige* Bedingungen sind. Für den Fall  $n = 2$ , wenn also keine extra Variablen  $x_3$  verwendet werden, dann sind diese (linearen) Bedingungen an die AR Koeffizienten auch hinreichend. Die Restriktionen (6.23) können z.B. mit einem Wald-Test, siehe Abschnitt 6.7, überprüft werden. Für eine genauere Diskussion siehe z.B. LÜTKEPOHL <sup>5</sup>.

## 6.6 Impulsantwort und Zerlegung der Prognosefehlervarianz

Die  $MA(\infty)$  Darstellung von  $(x_t)$

$$x_t = \sum_{k \geq 0} b_k \epsilon_{t-k}$$

(mit  $b_0 = I_n$ ) erlaubt es den Einfluss der Innovationen  $\epsilon_t$  auf zukünftige Werte des Prozesses zu analysieren. Ändert man  $\epsilon_{jt}$  um eine Einheit (während man alle anderen Größen  $\epsilon_{ls}$  für  $s \neq t$  oder  $l \neq j$  gleich lässt), dann ändert sich  $x_{i,t+k}$  um den Wert  $b_{ij,k}$ . Hier bezeichnet  $b_{ij,k}$  das  $ij$ -te Element von  $b_k$ . Die Folge  $(b_{ij,k} | k \geq 0)$  gibt also die Änderung der zukünftigen  $x_{it}$ 's an, wenn der "input" ( $\epsilon_{it}$ ) mit einem Impuls ( $\delta_t = 1, \delta_s = 0$  für  $s \neq t$ ) gestört wird. Man spricht daher auch von einer *Impulsantwort*. Die Koeffizienten  $b_k$  sind absolut summierbar, d.h.  $b_{ij,k} \rightarrow 0$  für  $k \rightarrow \infty$ . Der Effekt eines Impulses (Shocks) nimmt also im Laufe der Zeit (relativ) schnell ab.

Die Komponenten von  $\epsilon_t$  sind typischerweise miteinander korreliert ( $\Sigma = E\epsilon_t\epsilon_t'$  ist nicht diagonal). Daher ist eine naive/direkte Interpretation der Impulsantwort etwas fragwürdig. Ein Schock in der  $j$ -ten Variable hat einen unmittelbaren Einfluss auf die Zukunft (über die Impulsantwort ( $b_{ij,k}$ )) und einen mittelbaren Einfluss über die anderen Komponenten ( $\epsilon_{lt}$ ).

Um diese Schwierigkeiten (teilweise) zu umgehen, betrachtet man oft eine alternative  $MA(\infty)$  Darstellung

$$x_t = \sum_{k \geq 0} \underbrace{b_k T}_{=\tilde{b}_k} \underbrace{T^{-1} \epsilon_{t-k}}_{=\tilde{\epsilon}_{t-k}}$$

wobei die Transformationsmatrix  $T \in \mathbb{R}^{n \times n}$  so gewählt wird, dass  $TT' = \Sigma$  bzw.  $E\tilde{\epsilon}_t\tilde{\epsilon}_t' = T^{-1}\Sigma(T^{-1})' = I_n$ . Die Komponenten von  $\tilde{\epsilon}_t$  sind also unkorreliert und  $(\tilde{b}_{ij,k} | k \geq 0)$  ist daher besser geeignet den Einfluss von "individuellen" Shocks auf das System zu analysieren. Allerdings ist die Transformationsmatrix  $T$  nicht eindeutig durch die

<sup>5</sup> Helmut Lütkepohl. *Palgrave Handbook of Econometrics, Volume 1 Econometric Theory*, chapter Vector Autoregressive Models, pages 477–510. Palgrave MacMillan, 2007

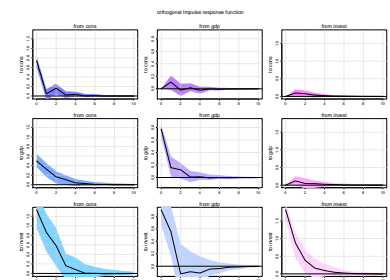


Abbildung 6.3: Impulsantwort

Bedingung  $TT' = \Sigma$  bestimmt, da  $(TO)(TO)' = \Sigma$  für jede orthogonale<sup>6</sup> Matrix  $O$  gilt. Oft wählt man für  $T$  eine untere Dreiecksmatrix, d.h. man bestimmt  $T$  über eine Cholesky Zerlegung von  $\Sigma$ . Dann ist auch  $\tilde{b}_0 = T$  eine untere Dreiecksmatrix. Die  $j$ -te Komponente  $\tilde{\epsilon}_{jt}$  hat daher einen unmittelbaren Einfluss auf die Komponenten  $x_{lt}$  für  $l \geq j$ . Die erste Komponente  $\tilde{\epsilon}_{1t}$  beeinflusst also alle Komponenten von  $x_t$  ohne Zeitverzögerung, während die letzte Komponente  $\tilde{\epsilon}_{nt}$  nur  $x_{nt}$  unmittelbar beeinflusst und alle anderen Komponenten nur mit einer gewissen Zeitverzögerung. Natürlich ist hier wichtig die Variablen richtig anzuordnen.

Die Varianz des  $h$ -Schritt Prognosefehlers ist nach Korollar 3.9 gleich

$$\Sigma_h = \sum_{j=0}^{h-1} b_j \Sigma b_j' = \sum_{j=0}^{h-1} b_j T (b_j T)' = \sum_{j=0}^{h-1} \tilde{b}_j \tilde{b}_j'$$

Die Varianz der  $k$ -ten Komponente des Prognosefehlers ist damit gleich

$$\Sigma_{kk,h} = \sum_{j=0}^{h-1} (\tilde{b}_{k1,j}^2 + \cdots + \tilde{b}_{kn,j}^2) = \sum_{i=1}^n (\tilde{b}_{ki,0}^2 + \cdots + \tilde{b}_{ki,h-1}^2)$$

und der Quotient

$$\frac{\tilde{b}_{ki,0}^2 + \cdots + \tilde{b}_{ki,h-1}^2}{\Sigma_{kk,h}}$$

ist der Anteil der Prognosefehlervarianz, der von der  $i$ -ten Komponente der transformierten Innovationen  $\tilde{\epsilon}_t$  herrührt.

Für eine genauere Diskussion der Impulsantwort und dieser Varianzzerlegung siehe z.B. LÜTKEPOHL<sup>7</sup>.

## 6.7 Schätzung von AR Systemen

### Yule-Walker Schätzer

Wir betrachten nun die Yule-Walker Gleichungen (6.14), (6.15) als ein Gleichungssystem für die AR Parameter  $a_1, \dots, a_p$  und  $\Sigma$ . Typischerweise verwendet man nur die Gleichungen für  $j = 0$  und  $j = 1, \dots, p$ . Wir erhalten also folgendes Gleichungssystem:

$$\gamma(0) = (a_1, \dots, a_p)(\gamma(1), \dots, \gamma(p))' + \Sigma \quad 6.24$$

$$(\gamma(1), \dots, \gamma(p)) = (a_1, \dots, a_p) \Gamma_p \quad 6.25$$

Die Autokovarianzfunktion  $\gamma(\cdot)$  muss dabei nicht unbedingt von einem regulärem AR(p) Prozess kommen. Wir wollen insbesondere auch die folgenden Fälle behandeln:

1. Die Autokovarianzfunktion stammt von einem AR( $p_0$ ) Prozess, aber  $p_0 \neq p$ .

$${}^6 OO' = O'O = I_n$$

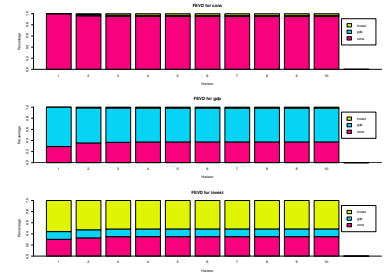


Abbildung 6.4: Zerlegung der Prognosefehlervarianz

<sup>7</sup> Helmut Lütkepohl. *Palgrave Handbook of Econometrics, Volume 1 Econometric Theory*, chapter Vector Autoregressive Models, pages 477–510. Palgrave MacMillan, 2007

2. Der Prozess ist kein AR(p) Prozess. Man versucht den Prozess  $(x_t)$  durch einen AR Prozess (bzw. ein AR System) zu approximieren.
3. Die Autokovarianzfunktion ist eine empirische Autokovarianzfunktion. D.h. wir setzen empirischen Autokovarianzen in die Yule-Walker Gleichungen ein und lösen dann die Gleichungen nach den Parametern auf, um Schätzer für die AR Parameter zu erhalten. Diese Schätzer nennt man *Yule-Walker Schätzer*.

Die Yule-Walker Gleichungen (6.24) und (6.25) sind genau die "Prognosegleichungen" für die Ein-Schrittprognose aus  $k = p$  vergangenen Werten, siehe (3.3) und (3.4). Daher können wir sofort schließen, dass die Yule-Walker Gleichungen *immer lösbar*<sup>8</sup> sind. Die Varianz  $\Sigma = \Sigma_{1,p}$  ist eindeutig bestimmt und die AR Koeffizienten  $a_1, \dots, a_p$  sind dann und nur dann eindeutig aus den YW-Gleichungen bestimmt, wenn  $\Gamma_p$  positiv definit ist.

<sup>8</sup> vorausgesetzt die Folge  $\gamma(k)$  ist eine Autokovarianzfunktion, d.h. positiv semidefinit.

Im Folgenden bezeichnen  $a_1, \dots, a_p, \Sigma$  Lösungen der YW-Gleichungen, bzw. der Prognosegleichungen. D.h. insbesondere  $\Sigma = \Sigma_{1,p}$ .

**Satz 6.3.** *Die Yule-Walker Gleichungen sind immer lösbar. Die Varianz  $\Sigma$  ist eindeutig bestimmt und die AR Koeffizienten  $a_1, \dots, a_p$  sind dann und nur dann eindeutig bestimmt, wenn  $\Gamma_p$  positiv definit ist.*

Für den Fall  $\Gamma_{p+1} > 0$  liefern die Yule-Walker Gleichungen ein stabiles AR(p) System, d.h. es gilt  $\det(I - a_1 z - \dots - a_p z^p) \neq 0$  für alle  $|z| < 1$ .

*Beweis.* Wir müssen nur noch den zweiten Teil des Satzes beweisen. Nach Lemma 6.1 müssen wir dazu zeigen, dass der Spektralradius der Begleitmatrix  $A$  kleiner als eins ist. Sei also  $c = (c_1, c_2, \dots, c_p) \in \mathbb{C}^{1 \times np}$  ein Links-Eigenvektor von  $A$  zum Eigenwert  $\lambda$ , d.h.  $cA = \lambda c$  und  $c \neq 0$ . Aus dem Beweis von Lemma 6.1 folgt auch  $c_1 \neq 0$ . Aus den Yule Walker Gleichungen folgt (siehe (6.18))

$$\Gamma_p = A\Gamma_p A' + B\Sigma B'$$

und daher

$$c\Gamma_p c^* = cA\Gamma_p A c^* + cB\Sigma B' c^* = |\lambda|^2 c\Gamma_p c^* + c_1 \Sigma c_1^*$$

Nach (3.10) ist die Annahme  $\Gamma_{p+1} > 0$  äquivalent zu  $\Gamma_p > 0$  und  $\Sigma = \Sigma_{1,p} > 0$ . Daher folgt  $|\lambda| < 1$  aus  $(1 - |\lambda|^2)c\Gamma_p c^* = c_1 \Sigma c_1^*$ .  $\square$

**Beispiel 6.4.** Die folgenden Beispiele sollen illustrieren, was passieren kann, wenn die Annahme  $\Gamma_{p+1} > 0$  nicht erfüllt ist.

1. Sei  $|a_1| < 1$ ,  $\epsilon_t \sim \text{WN}(\sigma^2)$ ,  $\sigma^2 > 0$  und  $z$  eine quadratisch integrierbare Zufallsvariable, für die  $\mathbb{E}z^2 > 0$  und  $\mathbb{E}\epsilon_t z = 0$  für alle  $t$ . Der Prozess

$$x_t = \begin{pmatrix} (1 - a_1 B)^{-1} \epsilon_t \\ z \end{pmatrix}$$

ist ein nicht regulärer AR(1) Prozess und das zugehörige AR(1) System

$$x_t = \begin{pmatrix} a_1 & 0 \\ 0 & 1 \end{pmatrix} x_{t-1} + \begin{pmatrix} \epsilon_t \\ 0 \end{pmatrix} \quad \boxed{6.26}$$

erfüllt die Stabilitätsbedingung *nicht*. Es gilt  $\Gamma_1 > 0$  und  $\Sigma = \Sigma_{1,1} = \text{diag}(\sigma^2, 0)$  ist singular. Die Yule-Walker Gleichungen haben eine eindeutige Lösung, nämlich die Parameter des wahren Systems (6.26).

2. Das zweite Beispiel ist ganz analog zu oben, nur nehmen wir hier an, dass  $z = 0$ . Der Prozess  $(x_t)$  ist daher regulär. Das zugehörige AR System ist aber nicht mehr eindeutig. Der Prozess  $(x_t)$  ist eine stationäre Lösung von allen AR Systemen der Form

$$x_t = \begin{pmatrix} a_1 & a_{12} \\ 0 & a_2 \end{pmatrix} x_{t-1} + \begin{pmatrix} \epsilon_t \\ 0 \end{pmatrix} \quad \boxed{6.27}$$

wobei  $a_2, a_{12}$  beliebige reelle Zahlen sind. Alle diese System sind auch Lösungen der Yule-Walker Gleichungen, da die Varianz  $\Gamma_1$  singular ist. Das heißt natürlich insbesondere, dass unter den Lösungen der Yule-Walker Gleichungen auch Systeme sind, die die Stabilitätsbedingung nicht erfüllen.

3. Hier betrachten wir den Prozess

$$x_t = \begin{pmatrix} a_1 & 0 \\ 0 & a_2 \end{pmatrix} x_{t-1} + \begin{pmatrix} \epsilon_t \\ \epsilon_t \end{pmatrix} \quad \boxed{6.28}$$

wobei  $|a_1| < 1$ ,  $|a_2| < 1$ ,  $a_1 \neq a_2$  und  $(\epsilon_t) \sim \text{WN}(\sigma^2)$ . Man kann leicht zeigen, dass die Varianz  $\Gamma_1$  positiv definit ist. Die Yule-Walker Gleichungen haben also eine eindeutige Lösung, nämlich genau die obigen wahren Parameter. Die Varianz der Innovationen  $\Sigma = \sigma^2(1, 1)'(1, 1)$  ist hier zwar singular (und daher auch die Töplitz Matrix  $\Gamma_2$ ), aber das stellt kein wirkliches Problem dar.

4. Wenn man im obigen Beispiel  $a_1 = a_2$  statt  $a_1 \neq a_2$  verlangt, dann folgt  $\det(\Gamma_1) = 0$ . Die Yule Walker Gleichungen sind nicht mehr eindeutig lösbar und unter den Lösungen gibt es auch nicht stabile Systeme. Dieser Fall ist ganz analog zum zweiten Beispiel, da

$$\begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix} x_t = \begin{pmatrix} x_{1t} \\ 0 \end{pmatrix} = \begin{pmatrix} (1 - a_1 B)^{-1} \epsilon_t \\ 0 \end{pmatrix}$$

Im skalaren Fall gibt der folgende Satz eine einfache hinreichende Bedingung für  $\Gamma_k > 0$  für alle  $k \geq 0$ .

**Satz 6.5.** Für den skalaren Fall folgt aus  $\gamma(0) > 0$  und  $\lim_{k \rightarrow \infty} \gamma(k) = 0$ , dass die Töplitz Matrizen  $\Gamma_k$  (für alle  $k \geq 1$ ) positiv definit sind.

*Beweis.* Wir führen einen Widerspruchsbeweis und nehmen an, dass  $\Gamma_k > 0$  und  $\det(\Sigma_{k+1})$  für ein  $k > 0$  gilt. Daher folgt  $\Gamma_k > 0$  und  $\Sigma_{1,k} = 0$  und damit  $x_{t+1} = a_1 x_t + \dots + a_k x_{t+1-k}$  für die entsprechenden "Prognose" Koeffizienten  $a_1, \dots, a_k$ . Wir betrachten den gestapelten Prozess  $(x_t^k)$  und die entsprechende Begleitmatrix  $A$ , siehe 6.17. Es gilt  $x_{t+1}^k = A x_t^k$  und daher  $x_{t+m}^k = A^m x_t^k$  für alle  $m \geq 0$ . Den gesuchten Widerspruch liefern nun

$$\Gamma_k = \text{Var}(x_{t+m}^k) = \text{Var}(A^m x_t^k) = A^m \Gamma_k (A')^m = (A^m \Gamma_k) \Gamma_k^{-1} (A^m \Gamma_k)'$$

Dieser Satz gilt für beliebige skalare Autokovarianzfunktionen, insbesondere also auch für empirische Autokovarianzfunktion  $\hat{\gamma}(k)$ . Da  $\hat{\gamma}_T(k) = 0$  für  $k \geq T$  folgt also, dass die Töplitz Matrizen  $\hat{\Gamma}_{T,k} = (\hat{\gamma}_T(j-i))_{i,j=1,\dots,k}$  immer positiv definit sind (solange  $\hat{\gamma}_T(0) > 0$ ). Die entsprechenden Yule-Walker Schätzer sind daher eindeutig und erfüllen die Stabilitätsbedingung.

und

$$A^m \Gamma_k = \mathbf{Cov}(A^m x_t^k, x_t^k) = \mathbf{Cov}(x_{t+m}^k, x_t^k) \xrightarrow{m \rightarrow \infty} 0$$

□

Im multivariaten Fall ist die Sache komplizierter, wie schon die obigen Beispiele zeigen. Es gilt aber folgender Satz:

**Satz 6.6.** Sei  $\gamma(\cdot)$  die Autokovarianzfunktion eines regulären Prozesses. Wenn die Varianz der Innovationen positiv definit ist, dann gilt  $\Gamma_k > 0$  für alle  $k \geq 0$ .

Das gilt natürlich auch für reguläre AR Prozesse.

*Beweis.* Sei  $\Sigma_0$  die Varianz der Innovationen. Aus  $\Sigma_{1,k-1} \geq \Sigma_0$  und der Beziehung (3.11) folgt die Behauptung unmittelbar. □

Gilt  $\hat{\gamma}(k) \rightarrow \gamma(k)$  f.s., dann gilt  $(\det(\Gamma_{T,k}) > 0 \text{ für } T \geq T_0)$  f.s.

### OLS Schätzer

Es gibt eine Reihe von alternativen Schätzern für AR Systeme. Die meisten dieser Schätzer sind asymptotisch äquivalent zu den Yule-Walker Schätzern.

Im Folgenden wollen wir auch AR Prozesse mit Mittelwert ungleich Null betrachten. Genauer gesagt, betrachten wir Prozesse der Form  $x_t = \mu + \tilde{x}_t$ , wobei  $\tilde{x}_t$  ein regulärer AR( $p$ ) Prozess ist. Den Mittelwert kann man auch gleich in das AR System integrieren, indem man folgendes inhomogene Modell

$$x_t = c_0 + a_1 x_{t-1} + \dots + a_p x_{t-p} + \epsilon_t \quad \boxed{6.29}$$

betrachtet. Die eindeutige, stationäre Lösung ist (wegen der Stabilitätsbedingung  $\det(I - a_1 z - \dots - a_p z^p) \neq 0, \forall |z| \leq 1$ ) gleich

$$x_t = (I - a_1 B - \dots - a_p B^p)^{-1} (c_0 + \epsilon_t) = \mu + (I - a_1 B - \dots - a_p B^p)^{-1} \epsilon_t = \mu + \tilde{x}_t$$

wobei

$$\mu = (I - a_1 - \dots - a_p)^{-1} c_0$$

Das inhomogene AR-System (6.29) ist ein multivariates Regressionsmodell

$$x_t = \beta y_t + \epsilon_t \quad \boxed{6.30}$$

Achtung: Die übliche Notation für Regressionsmodelle mit einer skalaren, abhängigen Variablen ist

$$y_t = x_t \beta + u_t$$

mit

$$\beta = (c_0, a_1, \dots, a_p) \text{ und } y_t = (1, x'_{t-1}, \dots, x'_{t-p})'$$

Hier sind die Rollen von  $y_t$  und  $x_t$  vertauscht und die abhängige Variable  $x_t$  ist  $n$ -dimensional.

Um die Notation etwas zu vereinfachen, nehmen wir im Folgenden an, dass wir eine Stichprobe  $x_{-p+1}, \dots, x_0, x_1, \dots, x_T$  zur Verfügung haben. Das obige Regressions-Modell (6.30) kann einfach mit OLS geschätzt werden

$$\hat{\beta} = \left( \frac{1}{T} \sum_{t=1}^T x_t y_t' \right) \left( \frac{1}{T} \sum_{t=1}^T y_t y_t' \right)^{-1} \quad \boxed{6.31}$$

Mit Hilfe des Frisch-Waugh Theorems kann man zeigen, dass der OLS Schätzer für die AR Koeffizienten  $(a_1, \dots, a_p)$  gegeben ist durch

$$(\hat{a}_1, \dots, \hat{a}_p) = (\tilde{\gamma}(0, 1), \dots, \tilde{\gamma}(0, p)) \tilde{\Gamma}_p^{-1}$$

wobei

$$\tilde{\Gamma}_p = [\tilde{\gamma}(i, j)]_{i,j=1,\dots,p}, \quad \tilde{\gamma}(i, j) = \frac{1}{T} \sum_{t=1}^T (x_{t-i} - \bar{x}^{(i)})(x_{t-j} - \bar{x}^{(j)})' \quad \text{und} \quad \bar{x}^{(i)} = \frac{1}{T} \sum_{t=1}^T x_{t-i}$$

Der OLS Schätzer für  $c_0$  folgt aus der Bedingung, dass das empirische Mittel der (OLS) Fehler  $\hat{\epsilon}_t = x_t - \hat{\beta}y_t$  gleich Null sein muss.

$$\hat{c}_0 = \bar{x}^{(0)} - \hat{a}_1 \bar{x}^{(1)} - \dots - \hat{a}_p \bar{x}^{(p)}$$

Für große Stichproben ( $T \rightarrow \infty$ ) gilt  $\bar{x}^{(i)} \approx \bar{x} = \frac{1}{T+p} \sum_{t=-p+1}^T x_t$  und  $\tilde{\gamma}(i, j) \approx \hat{\gamma}(j-i)$ . Daher sind die OLS Schätzer  $(\hat{a}_1, \dots, \hat{a}_p)$  asymptotisch äquivalent zu den Yule-Walker Schätzern. Für den Erwartungswert  $\mu$  liefert die OLS Schätzung

$$\hat{\mu} = (I_n - \hat{a}_1 - \dots - \hat{a}_p)^{-1} \hat{c}_0 \approx \bar{x}$$

Dieser (OLS) Schätzer  $\hat{\mu}$  ist also asymptotisch äquivalent zum empirischen Mittel  $\bar{x}$ .

### Maximum Likelihood Schätzer

Wenn die Innovationen  $(\epsilon_t)$  normal verteilt sind, also  $\epsilon_t \sim N(0, \Sigma)$ , dann sind auch die  $x_t$ 's multivariat normal verteilt. Es gilt

$$x_T^{T+p} \sim N(\iota_{T+p} \otimes \mu, \Gamma_{T+p})$$

und die negative Log-Likelihood Funktion ist daher (bis auf eine Konstante) gleich

$$l(\theta | x_{-p+1}, \dots, x_0, x_1, \dots, x_T) = \frac{1}{2} \ln \det \Gamma_T + \frac{1}{2} (x_T^{T+p} - \iota_{T+p} \otimes \mu)' \Gamma_{T+p}^{-1} (x_T^{T+p} - \iota_{T+p} \otimes \mu)$$

Im Parametervektor  $\theta$  sind alle Parameter des Modells zusammengefasst ( $\theta \triangleq (c_0, a_1, \dots, a_p, \Sigma)$ ). Diese Likelihood Funktion lässt sich erheblich vereinfachen, wenn man die Innovationen

$$\begin{aligned} \epsilon_t &= x_t - a_1 x_{t-1} - \dots - a_p x_{t-p} - c_0 \\ &= (x_t - \mu) - a_1 (x_{t-1} - \mu) - \dots - a_p (x_{t-p} - \mu) \end{aligned}$$

verwendet:

$$e := \begin{pmatrix} \epsilon_T \\ \vdots \\ \epsilon_1 \\ x_0^p - \iota_p \otimes \mu \end{pmatrix} = A(x_T^{T+p} - \iota_{T+p} \otimes \mu)$$

Die Matrix  $A$  ist eine obere Dreiecksmatrix, deren Diagonalelemente gleich eins sind. Die erste Block-Zeile von  $A$  ist z.B. gleich  $(I_n, -a_1, \dots, -a_p, 0, \dots, 0)$ . Die Varianz-Kovarianzmatrix von  $e$  ist gleich

$$\Xi := \mathbb{E} e e' = \text{diag}(\Sigma, \dots, \Sigma, \Gamma_p) = A \Gamma_{T+p} A'$$

$\otimes$  steht für das Kroneckerprodukt und  $\iota_m = (1, \dots, 1)' \in \mathbb{R}^{m \times 1}$ .

Später werden wir auch den vec bzw. den vech Operator benötigen. Für Matrizen  $A = (a_{ij}) \in \mathbb{R}^{m \times n}$  definiert man:

$$\text{vec}(A) = (a_{11}, \dots, a_{m1}, a_{12}, \dots, a_{mn}) \in \mathbb{R}^{mn \times 1}$$

und für symmetrische Matrizen  $A = A' \in \mathbb{R}^{n \times n}$

$$\text{vech}(A) = (a_{11}, \dots, a_{n1}, a_{22}, \dots, a_{n2}, a_{33}, \dots, a_{nn})$$

Siehe auch Anhang 6.8.

Daher gilt  $\Gamma_{T+p}^{-1} = A' \Xi^{-1} A$  und

$$\ln \det \Gamma_{T+p} = \ln \det \Xi = T \ln \det \Sigma + \ln \det(\Gamma_p).$$

Damit können wir die (negative) Log Likelihood Funktion auch schreiben als

$$\begin{aligned} l(\theta | x_{-p+1}, \dots, x_0, x_1, \dots, x_T) &= \frac{T}{2} \ln \det \Sigma + \frac{1}{2} \sum_{t=1}^T \epsilon_t' \Sigma^{-1} \epsilon_t + \\ &\quad + l(\theta | x_{-p+1}, \dots, x_0) \\ l(\theta | x_{-p+1}, \dots, x_0) &= \frac{1}{2} \ln \det \Gamma_p + \frac{1}{2} (x_0^p - \iota_p \otimes \mu)' \Gamma_p^{-1} (x_0^p - \iota_p \otimes \mu) \end{aligned}$$

Für große Stichproben kann man den Beitrag der ersten  $p$  Beobachtungen zur Log Likelihood vernachlässigen und nur den Term<sup>9</sup>

$$\tilde{l}(\theta | x_1, \dots, x_T) := \frac{T}{2} \ln \det \Sigma + \frac{1}{2} \sum_{t=1}^T \epsilon_t' \Sigma^{-1} \epsilon_t$$

betrachten. Dieser Ausdruck wird minimal, wenn man für  $c_0, a_1, \dots, a_p$  die OLS Schätzer einsetzt und für die Varianz  $\Sigma$  den Schätzer

$$\hat{\Sigma} = \frac{1}{T} \sum_{t=1}^T \hat{\epsilon}_t \hat{\epsilon}_t'$$

einsetzt, wobei  $\hat{\epsilon}_t = x_t - \hat{a}_1 x_{t-1} - \dots - \hat{a}_p x_{t-p} - \hat{c}_0$  die entsprechenden OLS Residuen bezeichnet. Das heißt die ML Schätzer sind asymptotisch äquivalent zu den OLS Schätzern (und damit auch zu den Yule-Walker Schätzern). Das Minimum der (bedingten) Log Likelihood Funktion  $\tilde{l}$  ist gleich

$$\tilde{l}(\hat{\theta} | x_1, \dots, x_T) = \frac{T}{2} \ln \det \hat{\Sigma} + \frac{Tn}{2} \quad \boxed{6.32}$$

### Asymptotik der Schätzer

Die oben diskutierten Schätzer sind i.A. nicht erwartungstreu. Aber es ist klar, dass sie konsistent sind, wenn die empirischen Momente  $(\bar{x}, \hat{\gamma}(k))$  konsistente Schätzer für die Populations Momente  $(\mu$  und  $\gamma(k))$  sind.

Das folgende Resultat (siehe HAMILTON<sup>10</sup> (Proposition 11.1 und 11.2)) behandelt die OLS Schätzer. Da die ML- und die Yule-Walker Schätzer asymptotisch äquivalent zu den OLS Schätzern sind, genügt es die OLS Schätzer zu betrachten.

**Theorem 6.7.** *Der Prozess  $(x_t)$  ist die eindeutige stationäre Lösung von*

$$x_t = c_0 + a_1 x_{t-1} + \dots + a_p x_{t-p} + \epsilon_t$$

*Das AR System erfüllt die Stabilitätsbedingung  $(\det(a(z))) \neq 0 \forall |z| \leq 1$  und die Innovationen  $(\epsilon_t)$  sind unabhängig, identisch verteilt mit  $\mathbb{E} \epsilon_t = 0$ ,  $\mathbb{E} \epsilon_t \epsilon_t' = \Sigma > 0$  und  $\mathbb{E} \epsilon_{it} \epsilon_{jt} \epsilon_{lt} \epsilon_{mt} < \infty$  für alle  $i, j, l, m$  und  $t \in \mathbb{Z}$ . Dann gilt*

Diese Zerlegung der Log Likelihood entspricht der Faktorisierung der Dichte als Produkt von bedingten Dichten:

$$\begin{aligned} f(x_T, \dots, x_1, x_0, \dots, x_{-p+1}) &= f(x_T | x_0^p) f(x_0^p) = \\ &= f(x_T | x_{T-1}^{T+p-1}) \dots f(x_1 | x_0^p) f(x_0^p) \end{aligned}$$

<sup>9</sup> Das ist die Log Likelihood bedingt auf die ersten  $p$  Beobachtungen  $x_{-p+1}, \dots, x_0$ .

Für endliche Stichproben gibt es aber sehr wohl Unterschiede. So ist z.B. für die OLS Schätzer, bzw. die ML Schätzer nicht garantiert, dass das geschätzte System die Stabilitätsbedingung erfüllt.

<sup>10</sup> J.D. Hamilton. *Time Series Analysis*. Princeton, NJ: Princeton University Press, 1994

Die Bedingung  $\Sigma > 0$  garantiert  $\Gamma_p > 0$  und damit auch  $\Omega > 0$ .



$$1. \quad \hat{\Omega} := \frac{1}{T} \sum_{t=1}^T y_t y_t' \xrightarrow{p} \mathbf{E} y_t y_t' =: \Omega$$

$$2. \quad \hat{\beta} \xrightarrow{p} \beta \text{ und } \hat{\Sigma} \xrightarrow{p} \Sigma$$

3.

$$\sqrt{T} \begin{pmatrix} \text{vec}(\hat{\beta}' - \beta') \\ \text{vech}(\hat{\Sigma} - \Sigma) \end{pmatrix} \xrightarrow{d} N \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \Sigma \otimes \Omega^{-1} & 0 \\ 0 & 2D_n^{\dagger}(\Sigma \otimes \Sigma)(D_n^{\dagger})' \end{bmatrix} \right)$$

Hier bezeichnet  $D_n$  die eindeutig definierte Matrix für die  $\text{vec}(A) = D_n \text{vech}(A)$  für alle  $A = A' \in \mathbb{R}^{n \times n}$  gilt und  $D_n^{\dagger} = (D_n' D_n)^{-1} D_n$ .

*Beweis.* Wir werden dieses Theorem nicht vollständig beweisen, sondern nur den Beweis für die asymptotische Verteilung von  $\hat{\beta}$  skizzieren. Mit der Stabilitätsbedingung und den Annahmen an die Innovation folgt, dass der Prozess  $(x_t)$  strikt stationär und ergodisch ist. Damit ist auch der Prozess  $((1, x_t', \dots, x_{t-p}')' \mid t \in \mathbb{Z})$  ergodisch und es folgt mit Theorem 2.17

$$\begin{aligned} \frac{1}{T} \sum_{t=1}^T y_t y_t' &\xrightarrow{p} \mathbf{E} y_t y_t' = \Omega \\ \frac{1}{T} \sum_{t=1}^T x_t y_t' &\xrightarrow{p} \mathbf{E} x_t y_t' \end{aligned}$$

Damit ist Punkt 1) bewiesen und auch der zweite Punkt folgt unmittelbar.

Für den OLS Schätzer  $\hat{\beta}$  gilt

$$\sqrt{T}(\hat{\beta} - \beta) = \left( \frac{1}{\sqrt{T}} \sum_{t=1}^T \epsilon_t y_t' \right) \left( \frac{1}{T} \sum_{t=1}^T y_t y_t' \right)^{-1}$$

bzw.

$$\sqrt{T} \text{vec}(\hat{\beta}' - \beta') = (I \otimes \hat{\Omega}^{-1}) \left( \frac{1}{\sqrt{T}} \sum_{t=1}^T \text{vec}(y_t \epsilon_t') \right)$$

Der Prozess  $(z_t = \text{vec}(y_t \epsilon_t'))$  ist strikt stationär und ergodisch. Weiters ist dieser Prozess eine Martingal-Differenz, da

$$\mathbf{E}(y_t \epsilon_t' \mid x_{t-1}, \dots, x_{-p+1}) = y_t \underbrace{\mathbf{E}(\epsilon_t' \mid x_{t-1}, \dots, x_{-p+1})}_{=0} = 0 \text{ für } t \geq 1$$

und daher

$$\mathbf{E}(z_t \mid z_{t-1}, \dots, z_1) = \mathbf{E}(\mathbf{E}(z_t \mid x_{t-1}, \dots, x_{-p+1}) \mid z_{t-1}, \dots, z_1) = 0 \text{ für } t > 1$$

Die Varianz-Kovarianz Matrix von  $z_t$  ist gleich

$$\mathbf{E} z_t z_t' = \Sigma \otimes \Omega$$

da  $\epsilon_t$  und  $y_t$  unabhängig sind und daher  $\mathbf{E}(y_t \epsilon_{kt})(y_t \epsilon_{lt})' = \mathbf{E} \epsilon_{kt} \epsilon_{lt} \mathbf{E} y_t y_t' = \Sigma_{kl} \Omega$ . Nach Theorem 2.19 folgt also

$$\frac{1}{\sqrt{T}} \sum_{t=1}^T z_t \xrightarrow{d} N(0, \Sigma \otimes \Omega)$$

Schließlich erhalten wir mit  $(I \otimes \hat{\Omega}^{-1}) \rightarrow (I \otimes \Omega^{-1})$  und  $(I \otimes \Omega^{-1})(\Sigma \otimes \Omega)(I \otimes \Omega^{-1})' = (\Sigma \otimes \Omega^{-1})$  das gewünschte Resultat für  $\hat{\beta}$ .  $\square$

Die Varianz  $\Omega$  und ihre Inverse  $\Omega^{-1}$  ist gleich

$$\Omega = \begin{pmatrix} 1 & (\boldsymbol{\iota}_p \otimes \boldsymbol{\mu})' \\ \boldsymbol{\iota}_p \otimes \boldsymbol{\mu} & \Gamma_p + (\boldsymbol{\iota}_p \boldsymbol{\iota}_p') \otimes (\boldsymbol{\mu} \boldsymbol{\mu}') \end{pmatrix}, \quad \Omega^{-1} = \begin{pmatrix} 1 + (\boldsymbol{\iota}_p \otimes \boldsymbol{\mu})' \Gamma_p^{-1} (\boldsymbol{\iota}_p \otimes \boldsymbol{\mu}) & -(\boldsymbol{\iota}_p \otimes \boldsymbol{\mu})' \Gamma_p^{-1} \\ -(\boldsymbol{\iota}_p \otimes \boldsymbol{\mu}) \Gamma_p^{-1} & \Gamma_p^{-1} \end{pmatrix}$$

Für die AR Parameter  $\boldsymbol{a} = (a_1, \dots, a_p)$  und deren Schätzer  $\hat{\boldsymbol{a}} = (\hat{a}_1, \dots, \hat{a}_p)$  erhält man also folgendes CLT:

$$\sqrt{T} \text{vec}(\hat{\boldsymbol{a}}' - \boldsymbol{a}') \xrightarrow{d} N(0, \Sigma \otimes \Gamma_p^{-1})$$

Mit den geschätzten Koeffizienten  $\hat{c}_0, \hat{a}_1, \dots, \hat{a}_p$  erhält man auch eine Schätzung für den Erwartungswert  $\boldsymbol{\mu} = \mathbf{E} \boldsymbol{x}_t$

$$\hat{\boldsymbol{\mu}} = (I_n - \hat{a}_1 - \dots - \hat{a}_p)^{-1} \hat{c}_0$$

Mit der Delta Methode können wir die asymptotische Verteilung dieses Schätzers bestimmen. Für  $\boldsymbol{\mu} = (I_n - a_1 - \dots - a_p)^{-1} c_0$  gilt

$$\frac{\partial \boldsymbol{\mu}}{\partial \text{vec}(\boldsymbol{\beta}')} = (I - a_1 - \dots - a_p)^{-1} [I_n \otimes (1, (\boldsymbol{\iota}_p \otimes \boldsymbol{\mu})')]$$

Mit den Formeln für die asymptotische Varianz von  $\text{vec}(\hat{\boldsymbol{\beta}}')$  folgt nun

$$\sqrt{T}(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}) \xrightarrow{d} N(0, M)$$

$$\begin{aligned} M &= \frac{\partial \boldsymbol{\mu}}{\partial \text{vec}(\boldsymbol{\beta}')} (\Sigma \otimes \Omega^{-1}) \left( \frac{\partial \boldsymbol{\mu}}{\partial \text{vec}(\boldsymbol{\beta}')} \right)' \\ &= (I - a_1 - \dots - a_p)^{-1} \Sigma ((I - a_1 - \dots - a_p)^{-1})' \\ &= \left( \sum_{j \geq 0} b_j \right) \Sigma \left( \sum_{j \geq 0} b_j \right)' \end{aligned}$$

Wir haben schon bei der Diskussion der OLS Schätzung angemerkt, dass der Schätzer  $\hat{\boldsymbol{\mu}}$  asymptotisch äquivalent zum Stichprobenmittel  $\bar{\boldsymbol{x}}$  ist. Das wird durch die obigen Formeln bestätigt. Die asymptotische Verteilung von  $\hat{\boldsymbol{\mu}}$  ist genau gleich der asymptotischen Verteilung von  $\bar{\boldsymbol{x}}$ , siehe Theorem 2.13.

Die asymptotische Varianz  $(\Sigma \otimes \Omega^{-1})$  von  $\text{vec}(\hat{\boldsymbol{\beta}}')$  kann konsistent mit  $(\hat{\Sigma} \otimes \hat{\Omega}^{-1})$  geschätzt werden. Diese Beobachtung ist wichtig für das Testen von Hypothesen und für die Konstruktion von Konfidenzintervallen.

Die  $k$ -te Zeile von  $\boldsymbol{\beta}$  bzw. von  $\hat{\boldsymbol{\beta}}$  (die im Folgenden mit  $\beta_k$  bzw.  $\hat{\beta}_k$  bezeichnet werden) entsprechen der Regression der  $k$ -ten Komponente  $x_{kt}$  auf  $y_t$ . Da

$$\sqrt{T}(\hat{\beta}_k - \beta_k) \xrightarrow{d} N(0, \Sigma_{kk} \Omega^{-1})$$

kann man die Verteilung  $\hat{\beta}_k$  für große  $T$  durch folgende Normalverteilung approximieren

$$\hat{\beta}_k \approx N \left( \beta_k, \hat{\sigma}_k^2 \left( \sum_{t=p+1}^T y_t y_t' \right)^{-1} \right)$$

$$\frac{\partial \boldsymbol{\mu}}{\partial c_0} = (I_n - a_1 - \dots - a_p)^{-1} \boldsymbol{u}_i$$

$$\frac{\partial \boldsymbol{\mu}}{\partial a_{ij,k}} = (I_n - a_1 - \dots - a_p)^{-1} \boldsymbol{u}_i \boldsymbol{\mu}_j$$

$\boldsymbol{u}_i$  ist der  $i$ -te Einheitsvektor im  $\mathbb{R}^n$

$$\boldsymbol{u}_i = (0, \dots, 0, 1, 0, \dots, 0)' \in \mathbb{R}^{n \times 1}$$

$$(I - a_1 z - \dots - a_p z^p)^{-1} = \sum_{j \geq 0} b_j z^j$$

und daher speziell für  $z = 1$

$$(I - a_1 - \dots - a_p)^{-1} = \sum_{j \geq 0} b_j$$

Alternativ kann man die asymptotische Varianz auch mit der "Yule-Walker" Prozedur schätzen. In der Formel für  $\Omega$  (bzw.  $\Omega^{-1}$ ) ersetzt man  $\boldsymbol{\mu}$  durch  $\bar{\boldsymbol{x}}$  und  $\Gamma_p$  durch  $\hat{\Gamma}_p = (\hat{\gamma}(j-i))_{i,j=1,\dots,p}$ . Die Varianz  $\Sigma$  schätzt man mit

$$\hat{\gamma}(0) - (\hat{\gamma}(1), \dots, \hat{\gamma}(p)) \hat{\Gamma}_p^{-1} (\hat{\gamma}(1), \dots, \hat{\gamma}(p))'$$

Hier ist  $\hat{\sigma}_k^2 = T^{-1} \sum_{t=1}^T \hat{\epsilon}_{kt}^2$  der (ML) Schätzer für die Varianz von  $\epsilon_{kt}$ . Diese Approximation entspricht aber genau der asymptotischen Verteilung von OLS Schätzern in einem klassischen Modell. Daher kann man z.B. wie "üblich"  $t$ -Tests durchführen.

Diese asymptotischen Resultate gelten auch, wenn einige der Koeffizienten  $a_j$  gleich Null sind. Insbesondere gilt das Theorem 6.7 auch für den Fall, dass man die Ordnung  $p$  zu groß gewählt hat. In diesem Fall ist der Prozess  $(x_t)$  ein  $AR(p_0)$  Prozess mit  $p_0 < p$  ist und die letzten  $(p - p_0)$  AR Koeffizienten sind gleich Null ( $a_{p_0+1} = \dots = a_p = 0 \in \mathbb{R}^{n \times n}$ ). Die Schätzer sind aber weiterhin konsistent und asymptotisch Normalverteilt. Allerdings sind die Schätzer nicht mehr effizient.

### Testen von linearen Parameter Restriktionen

Lineare Hypothesen der Form

$$R\text{vec}(\beta') = r \in \mathbb{R}^s$$

kann man mit dem Wald Test überprüfen

$$W = T(R\text{vec}(\hat{\beta}') - r)' \left[ R(\hat{\Sigma} \otimes \hat{\Omega}^{-1})R' \right]^{-1} (R\text{vec}(\hat{\beta}') - r) \xrightarrow{d} \chi_s^2$$

Unter der Nullhypothese konvergiert diese Teststatistik in Verteilung gegen eine (zentrale) Chi-Quadrat Verteilung mit  $s$  Freiheitsgraden.

**Beispiel 6.8.** Granger Kausalität in einem bivariaten Modell: Es genügt hier folgende Regression

$$x_{1t} = c_{10} + a_{11,1}x_{1,t-1} + \dots + a_{11,p}x_{1,t-p} + a_{12,1}x_{2,t-1} + \dots + a_{12,p}x_{2,t-p} + \epsilon_{1t}$$

zu betrachten und die Hypothese

$$a_{12,1} = \dots = a_{12,p} = 0$$

zu testen. Das kann z.B. mit einem "klassischem" F-Test gemacht werden.

### Ordnungsschätzung

Für eine vollständige Schätzung eines (V)AR Modells benötigt man natürlich auch eine Schätzung der "wahren" Ordnung, die wir im Folgenden mit  $p_0$  bezeichnen. Typischerweise startet man mit einer Vermutung  $p_{\max} \geq p_0$ , d.h. man nimmt an, dass zumindest eine obere Schranke für die wahre Ordnung zur Verfügung steht.

**SEQUENTIELLE TEST-PROZEDUR:** Hier testet man der Reihe nach die Hypothesen  $H_0 : a_{p_{\max}} = 0$ ,  $H_0 : a_{p_{\max}-1} = 0$ , ... und wählt dann die Ordnung  $p$  für die die Nullhypothese  $a_p = 0$  zum ersten Mal verworfen wird. Diese Hypothesen können z.B. mit dem Wald Test getestet werden.

INFORMATIONSKRITERIEN: Informationskriterien sind von der Form

$$I(p) = \ln \det \hat{\Sigma}_p + c_T \varphi(p) \quad (6.33)$$

wobei  $\hat{\Sigma}_p = \frac{1}{T} \sum_{t=1}^p \hat{\epsilon}_t \hat{\epsilon}_t'$  die Schätzung für die Innovations-Varianz für ein Modell der Ordnung  $p$  bezeichnet.  $c_T$  ist eine Funktion der Stichprobengröße  $T$  und  $\varphi(p)$  ist Maß für die Komplexität des AR( $p$ ) Modells. Für die unten beschriebenen Informationskriterien wählt man  $\varphi(p) = pn^2$ . Das heißt die Komplexität des Modells wird einfach mit der Zahl<sup>11</sup> der zu schätzenden Parameter gleichgesetzt. Der erste Term von  $I(p)$  ist ein Maß für die Anpassungsgüte des Modells. (Wie gut passt das Modell für die gegebene Stichprobe?) Dieser Term ist im wesentlichen gleich dem minimierten Wert der negativen Log Likelihood Funktion (siehe (6.32)). Der "fit" eines Modells wird umso besser, je komplexer es ist, d.h. der erste Term wird mit wachsendem  $p$  abnehmen. Der zweite Term ist ein "Strafterm", der zu komplexe Modelle bestraft. Als Schätzer für die Modell Ordnung  $p_0$  wählt man nun den Wert von  $p$ , der das Informationskriterium minimiert

<sup>11</sup> Die Zahl der Parameter ist eigentlich  $n + pn^2 + n(n+1)/2$ . Für die Wahl von  $p$  ist aber nur der Teil  $pn^2$  relevant.

$$\hat{p} = \arg \min_{p=0, \dots, p_{\max}} I(p)$$

Man hofft auf diese Weise ein möglichst einfaches Modell zu finden, das aber trotzdem einen guten "fit" liefert. Die folgenden Informationskriterien werden oft verwendet:

Akaike's Informationskriterium:

$$AIC(p) = \ln \det \hat{\Sigma}_p + \frac{2}{T} (pn^2) \quad (6.34)$$

Hannan-Quinn Kriterium:

$$HQ(p) = \ln \det \hat{\Sigma}_p + \frac{2 \ln \ln T}{T} (pn^2) \quad (6.35)$$

Schwarz-Kriterium (Rissanen Kriterium, Bayes Informationskriterium):

$$BIC(p) = \ln \det \hat{\Sigma}_p + \frac{\ln T}{T} (pn^2) \quad (6.36)$$

Ab  $T \geq 16$  gilt  $BIC(p) > HQ(p) > AIC(p)$  und daher  $\hat{p}(BIC) \leq \hat{p}(HQ) \leq \hat{p}(AIC)$  für die entsprechenden Ordnungsschätzer. Man kann unter recht allgemeinen Voraussetzungen zeigen, dass  $BIC$  und  $HQ$  konsistente Schätzungen für die Ordnung  $p_0$  liefern, wenn der Prozess ein AR Prozess ist.  $AIC$  hingegen tendiert dazu die Ordnung zu überschätzen, hat aber dafür Vorteile, wenn der Prozess kein AR Prozess ist (bzw. wenn  $(x_t)$  ein  $AR(\infty)$  Prozess ist).

#### PARTIELLE AUTOKORRELATIONSFUNKTION (PACF)

Für einen skalaren, stationäre Prozess  $(x_t)$ , mit Erwartungswert  $Ex_t = 0$ , definiert man die *partielle Autokorrelationsfunktion* durch folgende Prozedur. Sei  $M(t : s) = \text{span}\{x_t, \dots, x_s\} \subseteq \mathbb{L}_2$ , dann setzt

man

$$\begin{aligned}\delta(0) &= 1 \\ \delta(1) &= \rho(1) = \mathbf{Corr}(x_1, x_0) \\ \delta(k) &= \mathbf{Corr}(x_k - \mathbf{P}_{\mathbb{M}(1:k-1)}x_k, x_0 - \mathbf{P}_{\mathbb{M}(1:k-1)}x_0) \quad \text{für } k \geq 2 \\ \delta(-k) &= \delta(k) \quad \text{für } k < 0\end{aligned}$$

Die partielle Korrelation zum lag  $k$  versucht also den "direkten" Zusammenhang zwischen  $x_k$  und  $x_0$  zu quantifizieren, indem zuerst der Einfluss der dazwischen liegenden Werte  $x_1, \dots, x_{k-1}$  eliminiert wurde. Die partielle Autokorrelation zum lag  $k$  kann alternativ auch durch die Koeffizienten der optimalen Ein-Schritt Prognose berechnet werden. Sei

$$\hat{x}_{k-1,1,k-1} = c_1^{(k)}x_{k-1} + \dots + c_k^{(k)}x_0 = \mathbf{P}_{\mathbb{M}(0:k-1)}x_k \quad \boxed{6.37}$$

die optimale Ein-Schrittprognose für  $x_k$  aus den Werten  $x_{k-1}, \dots, x_0$ . Dann gilt (für  $k \geq 1$ )

$$c_k^{(k)} = \delta(k) \quad \boxed{6.38}$$

Um diese Behauptung zu zeigen, definieren wir  $u_k = x_k - \mathbf{P}_{\mathbb{M}(1:k-1)}x_k$  und  $u_0 = x_0 - \mathbf{P}_{\mathbb{M}(1:k-1)}x_0$ . Der Unterraum  $\mathbb{M}(0:k-1)$  ist die Summe der beiden orthogonalen Räume  $\mathbb{M}(1:k-1)$  und  $\text{span}(u_0)$ . Wendet man auf beiden Seiten von (6.37) die Projektion  $\mathbf{P}_{\mathbb{M}(1:k-1)^\perp}$  auf den Orthogonalraum von  $\mathbb{M}(1:k-1)$  an, so erhält man

$$\begin{aligned}c_k^{(k)}u_0 &= \mathbf{P}_{\mathbb{M}(1:k-1)^\perp}(c_1^{(k)}x_{k-1} + \dots + c_k^{(k)}x_0) = \\ &= \mathbf{P}_{\mathbb{M}(1:k-1)^\perp}(\underbrace{\mathbf{P}_{\mathbb{M}(1:k-1)} + \mathbf{P}_{\text{span}(u_0)}}_{=\mathbf{P}_{\mathbb{M}(0:k-1)}})(u_k + \underbrace{\mathbf{P}_{\mathbb{M}(1:k-1)}x_k}_{=x_k}) = \mathbf{P}_{\text{span}(u_0)}u_k\end{aligned}$$

Das heißt

$$c_k^{(k)} = \frac{\mathbf{E}u_k u_0}{\mathbf{E}u_0^2} = \frac{\mathbf{E}u_k u_0}{\sqrt{\mathbf{E}u_0^2} \sqrt{\mathbf{E}u_k^2}} = \mathbf{Corr}(u_k, u_0)$$

Hier haben wir noch  $\mathbf{E}u_0^2 = \mathbf{E}u_k^2$  benutzt. Diese Identität folgt z.B. aus der Symmetrie der Autokovarianzfunktion  $\gamma(k) = \gamma(-k)$ .

Für einen  $\text{AR}(p)$  Prozess sind die Koeffizienten der Einschritt Prognose ab dem lag  $p$  gleich Null. Ein Prozess  $(x_t)$  ist daher dann und nur dann ein  $\text{AR}(p)$  Prozess, wenn  $\delta(p) \neq 0$  und  $\delta(k) = 0 \quad \forall k > p$  gilt. Diese Charakterisierung von AR Prozessen ist analog zur Charakterisierung von MA Prozessen mit Hilfe der Autokovarianzfunktion.

Die Gleichung (6.38) zeigt auch, wie die PACF geschätzt werden kann. Man schätzt einfach die Koeffizienten der optimalen Ein-Schrittprognose (mit den geschätzten Autokovarianzen  $\hat{\gamma}(k)$ ) und setzt dann  $\hat{\delta}(k) = \hat{c}_k^{(k)}$ . Alternativ kann man auch ein  $\text{AR}(k)$  Modell schätzen und setzt  $\hat{\delta}(k) = \hat{a}_k$ .

Mit Hilfe von Theorem 6.7 können wir die asymptotische Verteilung der geschätzten PACF Koeffizienten angeben. Sei  $(x_t - \mathbf{E}x_t)$  ein

AR( $p$ ) Prozess, der die Bedingungen von Theorem 6.7 erfüllt. Wir nehmen auch an, dass  $a_p \neq 0$  gilt. Das heißt,  $p$  ist die wahre (minimale) Ordnung von  $(x_t - \mathbb{E}x_t)$ . Für  $k > p$  folgt nun

$$\sqrt{T}\hat{\delta}(k) \xrightarrow{d} N(0, \sigma^2(\Gamma_k^{-1})_{kk})$$

wobei  $\sigma^2 = \mathbb{E}\epsilon_t^2 > 0$  die Varianz der Innovationen ist. Man kann leicht zeigen, dass das  $k$ -te Diagonal-Element von  $\Gamma_k^{-1}$  (für  $k > p$ ) gleich  $\sigma^2$  ist. Das heißt

$$\sqrt{T}\hat{\delta}(k) \xrightarrow{d} N(0, 1)$$

Diese asymptotische Aussage gilt insbesondere auch für den Fall  $p = 0$ , d.h. für IID Prozesse  $(x_t = \mu + \epsilon_t)$ . Vergleiche auch die asymptotische Verteilung der Autokorrelationsfunktion (2.10).

### Modellvalidierung

Nachdem man ein (V)AR Modell (inklusive der Ordnung) geschätzt hat, sollte man überprüfen, ob das Modell eine adäquate Beschreibung des zugrundeliegende Prozesses ist. Mit anderen Worten, man sollte also überprüfen, ob der Prozess  $(x_t)$  wirklich ein AR Prozess ist. Fast alle Validierungs Prozeduren verwenden dazu die geschätzten Innovationen  $\hat{\epsilon}_t$ . Die wahren Innovationen sind ein white noise Prozess und daher sollten die geschätzten Innovation sich auch ähnlich wie ein weißes Rauschen verhalten. Sie sollten also keine signifikanten seriellen Korrelationen zeigen.

DGP: data generating process

PORTMANTEAU TEST: Um die Nullhypothese  $\mathbb{E}\epsilon_t\epsilon'_{t-k} = 0$  zu testen verwendet man oft folgende Test Statistik

Im skalaren Fall erhält man die Box-Pierce Test Statistik (2.11).

$$Q = T \sum_{k=1}^h \text{tr} \left( \hat{\gamma}_\epsilon(k)' \hat{\gamma}_\epsilon^{-1}(0) \hat{\gamma}_\epsilon(k) \hat{\gamma}_\epsilon^{-1}(0) \right)$$

wobei

$$\hat{\gamma}_\epsilon(k) = \frac{1}{T} \sum_{t=k+1}^T \hat{\epsilon}_t \hat{\epsilon}'_{t-k}$$

die empirische Autokovarianzfunktion der geschätzten Innovation  $\hat{\epsilon}_t$  bezeichnet. Unter der Annahme, dass der wahre Prozess ein AR( $p$ ) Prozess ist (und dass ein AR( $p$ ) Modell geschätzt wurde) konvergiert die Test Statistik für  $T \rightarrow \infty$  gegen eine Chi-Quadrat Verteilung mit  $n^2(h-p)$  Freiheitsgraden:

$$Q \xrightarrow{d} \chi_{n^2(h-p)}^2$$

Die Zahl der Freiheitsgrade muss korrigiert werden, da die  $\hat{\epsilon}_t$ 's die Residuen eines geschätzten AR( $p$ ) Modells sind!

Für "kleine" Stichproben liefert die leicht adaptierte Test Statistik

Vergleiche auch die Ljung-Box Test Statistik (2.12).

$$Q^* = T^2 \sum_{k=1}^h \frac{1}{T-k} \text{tr} \left( \hat{\gamma}_\epsilon(k)' \hat{\gamma}_\epsilon^{-1}(0) \hat{\gamma}_\epsilon(k) \hat{\gamma}_\epsilon^{-1}(0) \right)$$

etwas bessere Resultate.

LM TEST (BREUSCH-GODFREY TEST): Der LM Test auf Autokorrelation der Ordnung  $h$  testet, ob die Koeffizienten  $(d_1, \dots, d_h)$  in dem Modell

$$\epsilon_t = d_1\epsilon_{t-1} + \dots + d_h\epsilon_{t-h} + \eta_t$$

gleich Null sind. Der Prozess  $(\eta_t)$  ist eine white noise Prozess, d.h. falls einige dieser Koeffizienten ungleich Null sind, dann ist der Prozess  $(\epsilon_t)$  selbst ein AR Prozess. Zur Berechnung der Test Statistik betrachtet man folgendes Hilfs-Modell

$$\hat{\epsilon}_t = e_0 + e_1x_{t-1} + \dots + e_px_{t-p} + d_1\hat{\epsilon}_{t-1} + \dots + d_h\hat{\epsilon}_{t-h} + v_t$$

Die  $\hat{\epsilon}_t$ 's sind wieder die geschätzten Residuen (Innovationen) des ursprünglichen AR(p) Modells. Man schätzt nun die Parameter des Hilfsmodell (mit OLS) und erhält eine Schätzung  $\hat{v}_t$  für die Fehler  $v_t$ . Die LM Test Statistik ist dann

$$Q_{LM} = T(n - \text{tr}(\hat{\Sigma}_\epsilon^{-1}\hat{\Sigma}_v))$$

wobei

$$\hat{\Sigma}_\epsilon = \frac{1}{T} \sum_{t=1}^T \hat{\epsilon}_t \hat{\epsilon}_t' \quad \text{und} \quad \hat{\Sigma}_v = \frac{1}{T} \sum_{t=1}^T \hat{v}_t \hat{v}_t'$$

Unter der Nullhypothese hat diese Teststatistik eine asymptotische Chi-Quadratverteilung mit  $(hn^2)$  Freiheitsgraden:

$$Q_{LM} \xrightarrow{d} \chi_{hn^2}^2$$

## 6.8 Anhang

Das Kronecker Produkt von zwei Matrizen  $A \in \mathbb{R}^{m \times n}$  und  $B \in \mathbb{R}^{p \times q}$  ist definiert als

$$A \otimes B = \begin{pmatrix} a_{11}B & a_{12}B & \dots & a_{1n}B \\ a_{21}B & a_{22}B & \dots & a_{2n}B \\ \vdots & \vdots & & \vdots \\ a_{m1}B & a_{m2}B & \dots & a_{mn}B \end{pmatrix} \in \mathbb{R}^{(mp) \times (nq)}$$

Es gelten folgende Rechenregeln

$$\begin{aligned} (A \otimes B)(C \otimes D) &= (AC) \otimes (BD) \\ (A \otimes B)^{-1} &= (A^{-1} \otimes B^{-1}) \\ (A \otimes B)' &= (A' \otimes B') \end{aligned}$$

Hier wird immer stillschweigend angenommen, dass die Matrizen kompatibel sind.

Der vec Operator formt einen Vektor aus einer Matrix, indem er die Spalten der Matrix übereinanderstapelt. Für  $A = (a_{ij}) \in \mathbb{R}^{m \times n}$  setzt man

$$\text{vec}(A) = (a_{11}, \dots, a_{m1}, a_{12}, \dots, a_{m2}, a_{13}, \dots, a_{mn})' \in \mathbb{R}^{mn \times 1}$$

Es gilt

$$\text{vec}(ABC) = (C' \otimes A)\text{vec}(B)$$

Für symmetrische Matrizen verwendet man oft den vech Operator, der nur die Elemente auf und unterhalb der Hauptdiagonalen verwendet. Das heißt, für  $A = A' \in \mathbb{R}^{n \times n}$  definiert man

$$\text{vech}(A) = (a_{11}, a_{21}, \dots, a_{n1}, a_{22}, \dots, a_{n2}, a_{33}, \dots, a_{nn})' \in \mathbb{R}^{(n(n+1)/2) \times 1}$$

Die Matrix  $D_n$  ist die eindeutig definierte Matrix, für die gilt

$$\text{vec}(A) = D_n \text{vech}(A), \quad \forall A = A' \in \mathbb{R}^{n \times n}$$

Im Theorem 6.7 wird folgende Links-Inverse von  $D_n$  benötigt  $D_n^\dagger = (D_n' D_n)^{-1} D_n'$ . Es gilt

$$\text{vech}(A) = D_n^\dagger \text{vec}(A), \quad \forall A = A' \in \mathbb{R}^{n \times n}$$



# 7

## ARMA Prozesse

ARMA System / Prozess, Stabilitätsbedingung, minimum-phase Bedingung, Identifizierbarkeit, ACF, (rationale) spektrale Dichte, ML Schätzung, VDE's

ARMA Prozesse sind eine Verallgemeinerung von AR(p) und MA(q) Prozessen. Ein ARMA System ist eine Differenzen Gleichung der Form

$$x_t = a_1 x_{t-1} + \dots + a_p x_{t-p} + \epsilon_t + b_1 \epsilon_{t-1} + \dots + b_q \epsilon_{t-q} \quad \boxed{7.1}$$

wobei  $(\epsilon_t) \sim \text{WN}(\Sigma)$  ein  $n$ -dimensionales weißes Rauschen ist,  $a_j \in \mathbb{R}^{n \times n}$ ,  $b_j \in \mathbb{R}^{n \times n}$  sind Matrizen und  $(x_t)$  ist ein  $n$ -dimensionaler (stationärer) Prozess. Analog zum AR Fall nennt man eine stationäre Lösung dieser Differenzen Gleichung ARMA Prozess. Wesentliche Fragen sind:

ARMA steht für Auto Regressive Moving Average. Der multivariate Fall ( $n > 1$ ) wird oft auch mit VARMA (Vector ARMA) bezeichnet.

1. Sei ein ARMA System (d.h. die Parameter  $a_j$ ,  $b_j$  und der white noise Prozess  $\epsilon_t$ ) gegeben. Existiert eine stationäre Lösung? Wenn ja, ist diese Lösung eindeutig? Wie kann man diese (eindeutige, stationäre) Lösung konstruieren und welche Eigenschaften hat sie (Mittelwert, ACF, spektrale Verteilung, ...)?
2. Sei umgekehrt ein ARMA Prozess gegeben. Kann man das zugrunde liegende ARMA System (d.h. die Parameter  $a_j$ ,  $b_j$  und  $\Sigma$ ) eindeutig bestimmen? Falls nur eine Zeitreihe gegeben ist, wie kann man das ARMA System schätzen?

Wenn  $a_p \neq 0$  und  $b_q \neq 0$ , dann ist  $(p, q)$  die Ordnung des ARMA Systems (ARMA Prozesses) und die (eindeutige) stationäre Lösung nennt man ARMA( $p, q$ ) Prozess.

### 7.1 Struktur Theorie

Das ARMA System kann auch mit Hilfe von Filtern geschrieben werden:

$$a(B)x_t = b(B)\epsilon_t$$

wobei

$$\begin{aligned} a(B) &= I_n - a_1 B - \dots - a_p B^p \\ b(B) &= I_n + b_1 B + \dots + b_p B^p \end{aligned}$$

Die z-Transformation dieser beiden Filter ist

$$\begin{aligned} a(z) &= I_n - a_1 z - \dots - a_p z^p \\ b(z) &= I_n + b_1 z + \dots + b_p z^p \end{aligned}$$

Die Antwort auf die oben gestellten Fragen hängen wesentlich von den folgenden Bedingungen an das ARMA System ab:

- Stationaritäts-Bedingung:

$$\det(a(z)) \neq 0 \quad \forall |z| \neq 1 \quad \boxed{7.2}$$

- Stabilitäts-Bedingung:

$$\det(a(z)) \neq 0 \quad \forall |z| \leq 1 \quad \boxed{7.3}$$

- minimum phase Bedingung:

$$\det(b(z)) \neq 0 \quad \forall |z| < 1 \quad \boxed{7.4}$$

- strikte minimum phase Bedingung:

$$\det(b(z)) \neq 0 \quad \forall |z| \leq 1 \quad \boxed{7.5}$$

- Koprimitäts Bedingung: Das Paar  $\{a(z), b(z)\}$  ist links koprimit.

**Theorem 7.1.**

1. Das ARMA System hat eine eindeutige, stationäre Lösung, wenn die Stationaritäts-Bedingung (7.2) erfüllt ist. Diese Lösung ist regulär und hat die MA( $\infty$ ) Darstellung

$$x_t = a^{-1}(B)b(B)\epsilon_t = \sum_{j=-\infty}^{\infty} \bar{b}_j \epsilon_{t-j}$$

Wenn die stärkere Stabilitäts-Bedingung erfüllt ist, dann hat  $(x_t)$  eine kausale MA( $\infty$ ) Darstellung

$$x_t = a^{-1}(B)b(B)\epsilon_t = \sum_{j=0}^{\infty} \bar{b}_j \epsilon_{t-j}$$

2. Gilt die Stabilitätsannahme und die minimum phase Annahme, dann ist  $(\epsilon_t)$  der Innovations Prozess von  $(x_t)$ . Falls die etwas stärkere strikte minimum phase Bedingung erfüllt ist, dann besitzt der Prozess  $(x_t)$  eine sogenannte AR( $\infty$ ) Darstellung

$$\epsilon_t = b^{-1}(B)a(B)x_t = \sum_{j \geq 0} \bar{a}_j x_{t-j}$$

3. Für jeden regulären ARMA Prozess existiert ein (zugehöriges) ARMA System, das die Stabilitäts-Bedingung, die minimum phase Bedingung und die Koprimitäts Bedingung erfüllt.
4. Im skalarem Fall ist das ARMA System durch die Stabilitäts-Bedingung, die minimum phase Bedingung und die Koprimitäts Bedingung eindeutig bestimmt. D.h. in diesem Fall sind die ARMA Parameter identifizierbar. Im multivariaten Fall jedoch sind diese Bedingungen nicht ausreichend, um Identifizierbarkeit zu erreichen.

Seien  $a(z)$ ,  $b(z)$  und  $c(z)$  drei Polynom Matrizen (passender Dimension). Wenn  $a(z) = b(z)c(z)$  gilt, dann nennt man  $b(z)$  einen links Teiler von  $a(z)$ . Die Matrix  $a(z)$  ist links prim, wenn alle links Teiler  $b(z)$  von  $a(z)$  unimodular sind (d.h.  $\det(b(z)) = \det(b(0)) \neq 0 \quad \forall z \in \mathbb{C}$ ). Ein Paar von Polynom Matrizen  $\{a(z), b(z)\}$  nennt man links koprimit, wenn die zusammengesetzte Matrix  $[a(z), b(z)]$  links prim ist.

Wir beweisen hier dieses Theorem nicht, sondern verweisen z.B. auf HANNAN, DEISTLER (1988)<sup>1</sup>. Einige Beispiele und Bemerkungen sollen aber die Behauptungen doch motivieren und erläutern. (Punkt 2. wird im Abschnitt 7.3 diskutiert.)

<sup>1</sup> E. J. Hannan and M. Deistler. *The Statistical Theory of Linear Systems*. John Wiley & Sons, New York, 1988

Ist die Stationaritäts-Bedingung erfüllt, so ist der Filter  $a(B)$  invertierbar (siehe Satz 5.7) und es folgt unmittelbar, dass  $x_t = a^{-1}(B)b(B)\epsilon_t$  die eindeutige stationäre Lösung ist. Wenn  $\det(a(z))$  Nullstellen am Einheitskreis hat, dann existiert entweder keine stationäre Lösung (z.B. im skalarem Fall  $x_t = x_{t-1} + \epsilon_t$  mit  $E\epsilon_t^2 > 0$ ) oder es existieren unendlich viele stationäre Lösungen. Das passiert z.B. für das ARMA System

$$(1 - B)x_t = (1 - B)\epsilon_t$$

Für dieses System ist jeder Prozess der Form

$$x_t = z + \epsilon_t$$

(wobei  $Ez'z < \infty$  und  $Ez\epsilon_t' = 0 \forall t \in \mathbb{Z}$ ) eine stationäre Lösung. Da  $z$  (im Wesentlichen) beliebig ist, sehen wir, dass es unendlich viele stationäre Lösungen gibt. Die einzige reguläre Lösung ( $x_t = \epsilon_t$ ) kann durch das (triviale) ARMA System

$$x_t = \epsilon_t$$

dargestellt werden, das natürlich die Stabilitäts-Bedingung, die strikte minimum phase Bedingung und die Koprimitäts Bedingung erfüllt.

Betrachten wir nun ein skalares ARMA System  $a(B)x_t = b(B)\epsilon_t$ , das die Stabilitätsbedingung erfüllt. Wenn  $a(z)$  eine Nullstelle  $z_k \in \mathbb{R}$ ,  $|z_k| < 1$  im Einheitskreis hat, dann ist

$$\tilde{a}(z) = (1 - z_k z)(1 - z_k^{-1} z)^{-1} a(z)$$

ein Polynom vom Grad  $p$  mit einer Nullstelle  $z_k^{-1}$ ,  $|z_k^{-1}| > 1$  (statt der Nullstelle  $z_k$ ). Weiters hat der Filter  $(1 - z_k B)(1 - z_k^{-1} B)^{-1}$  einen konstanten "gain" (siehe Beispiel 5.13) und daher ist  $\tilde{\epsilon}_t = (1 - z_k B)(1 - z_k^{-1} B)^{-1} \epsilon_t$  ein weißes Rauschen. Damit haben wir gezeigt, dass

$$\underbrace{(1 - z_k B)(1 - z_k^{-1} B)^{-1} a(B)}_{\tilde{a}(B)} x_t = b(B) \underbrace{(1 - z_k B)(1 - z_k^{-1} B)^{-1} \epsilon_t}_{\tilde{\epsilon}_t}$$

eine äquivalente ARMA Darstellung von  $(x_t)$  ist. Falls  $a(z)$  ein Paar konjugiert komplexer Nullstellen  $z_k, \bar{z}_k$  mit  $|z_k| = |\bar{z}_k| < 1$  hat, dann können wir den Filter  $(1 - z_k B)(1 - \bar{z}_k B)(1 - z_k^{-1} B)^{-1}(1 - \bar{z}_k^{-1} B)^{-1}$  verwenden, um ein äquivalentes ARMA System zu konstruieren. Wiederholt man nun diese Prozedur für alle "instabilen" Nullstellen  $z_k$ ,  $|z_k| < 1$ , so erhält man schließlich ein ARMA System, das die Stabilitätsbedingung erfüllt.

Derselbe Trick funktioniert auch für Nullstellen von  $b(z)$ . Sei  $z_k$ ,  $|z_k| < 1$  eine (relle) Nullstelle von  $b(z)$ . Dann ist

$$a(B)x_t = \underbrace{b(B)(1 - z_k^{-1} B)^{-1}(1 - z_k B)}_{\tilde{b}(B)} \underbrace{(1 - z_k^{-1} B)(1 - z_k B)^{-1} \epsilon_t}_{\tilde{\epsilon}_t}$$

Beachten Sie, dass hier  $a(z) = b(z)$ !

ein äquivalentes ARMA System, bei dem die "böse" Nullstelle  $z_k, |z_k| < 1$  des MA Polynoms durch eine "brave" Nullstelle  $z_k^{-1}, |z_k^{-1}| > 1$  ersetzt wurde. Wiederholt man diesen Trick für alle Nullstellen des MA Polynoms innerhalb des Einheitskreises, so bekommt man schließlich ein ARMA System, das die minimum phase Bedingung erfüllt.

Sei nun ein ARMA System  $a(B)x_t = b(B)\epsilon_t$  gegeben, das die Stabilitätsbedingung und die minimum phase Bedingung erfüllt. Weiters sei  $c(z)$  eine quadratische Polynom Matrix mit  $c(0) = I_n$  und  $\det(c(z)) \neq 0$  für alle  $|z| \leq 1$ . Das ARMA System

$$\underbrace{c(B)a(B)}_{\tilde{a}(B)} x_t = \underbrace{c(B)b(B)}_{\tilde{b}(B)} \epsilon_t$$

erfüllt auch die Stabilitätsbedingung und die minimum phase Bedingung und beide Systeme haben dieselbe stationäre Lösung

$$\tilde{a}^{-1}(B)\tilde{b}(B)\epsilon_t = a^{-1}(B)\underbrace{c^{-1}(B)c(B)}_{=I_n} b(B)\epsilon_t = a^{-1}(B)b(B)\epsilon_t$$

Diese beiden Systeme nennt man daher *Beobachtungs äquivalent*. Ist  $c(z)$  nicht unimodular, dann sind  $(\tilde{a}(z), \tilde{b}(z))$  nicht links prim. Um eine eindeutige Darstellung zu bekommen, müssen wir zunächst gemeinsame (nicht unimodulare) links Teiler des Paares  $(\tilde{a}(z), \tilde{b}(z))$  kürzen. Die Ordnung  $(\tilde{p}, \tilde{q})$  des ARMA Systems  $\tilde{a}(B)x_t = \tilde{b}(B)\epsilon_t$  ist größer gleich der Ordnung  $(p, q)$  des ARMA Systems  $a(B)x_t = b(B)\epsilon_t$ , da  $\tilde{p} = \deg(\tilde{a}) = \deg(ca) \geq \deg(a) = p$  und  $\tilde{q} = \deg(\tilde{b}) = \deg(cb) \geq \deg(b) = q$ . Mit der Koprimheits-Bedingung stellen wir auch sicher, dass die Ordnung des Systems möglichst klein ist und wir daher keine unnötigen Parameter verwenden.

Im *skalaren* Fall garantieren die Stabilitäts-Bedingung, die minimum phase Bedingung und die Koprimheits-Bedingung zusammen mit der Normierung  $a(0) = a_0 = 1, b(0) = b_0 = 1$ , dass das ARMA System eindeutig bestimmt ist. Im multivariaten Fall sind diese Bedingungen noch nicht ausreichend, um ein eindeutiges ARMA System festzulegen, wie man am folgenden Beispiel sieht:

$$\begin{aligned} c(z)(a(z), b(z)) &= \begin{pmatrix} 1 & c_{12}z \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 + a_{11}z & a_{12}z & | & 1 + b_{11}z & b_{12}z \\ 0 & 1 & | & 0 & 1 \end{pmatrix} = \\ &= \begin{pmatrix} 1 + a_{11}z & (a_{12} + c_{12})z & | & 1 + b_{11}z & (b_{12} + c_{12})z \\ 0 & 1 & | & 0 & 1 \end{pmatrix} = (\tilde{a}(z), \tilde{b}(z)) \end{aligned}$$

Hier erhält man für jedes  $c_{12} \in \mathbb{R}$  ein (äquivalentes) ARMA(1,1) System, das alle Bedingungen erfüllt!

## 7.2 Die spektrale Dichte von (regulären) ARMA Prozessen

Die spektrale Dichte eines ARMA Prozesses  $x_t = a^{-1}(B)b(B)\epsilon_t$  folgt unmittelbar aus Theorem 5.12

$$f(\lambda) = \frac{1}{2\pi} a^{-1}(e^{-i\lambda}) b(e^{-i\lambda}) \Sigma b(e^{-i\lambda})^* (a^{-1}(e^{-i\lambda}))^* \quad \boxed{7.6}$$

Die Matrix  $c(z)$  ist unimodular und  $\deg(\tilde{a}(z)) = \deg(c(z)a(z)) = \deg(a(z)) = 1$  und  $\deg(\tilde{b}(z)) = \deg(c(z)b(z)) = \deg(b(z)) = 1$ .

Der skalare Fall ist einfacher, weil eine  $1 \times 1$  Polynom Matrizen  $c(z)$  dann und nur dann unimodular ist, wenn sie konstant ist (d.h.  $c(z) = c(0) \forall z \in \mathbb{C}$ ).

Die Berechnung der Autokovarianz-Funktion von ARMA Prozessen ist nicht ganz so einfach wie für den AR Fall. Man kann die ACF z.B. mit Hilfe der MA( $\infty$ ) Darstellung bestimmen, siehe (2.5) oder mit Hilfe der spektralen Dichte, siehe (4.23).

Im skalaren Fall (mit  $E\epsilon_t^2 = \sigma^2$ ) vereinfacht sich diese Formel zu

$$f(\lambda) = \frac{\sigma^2}{2\pi} \frac{|b(e^{-i\lambda})|^2}{|a(e^{-i\lambda})|^2} \quad \boxed{7.7}$$

Wertet man die *rationale* (Matrix) Funktion

$$\underline{f}(z) = \frac{1}{2\pi} a^{-1}(z) b(z) \Sigma b(z^{-1})' (a^{-1}(z^{-1}))' \quad \boxed{7.8}$$

an der Stelle  $z = e^{-i\lambda}$  aus, so erhält man die spektrale Dichte  $f(\lambda)$

$$f(\lambda) = \underline{f}(e^{-i\lambda}), \quad \lambda \in [-\pi, \pi]$$

Mit anderen Worten, die Elemente  $f_{ij}(\lambda)$  der spektralen Dichte  $f(\lambda)$  sind rationale Funktionen in  $e^{-i\lambda}$ . Die Funktion  $f_{ij}(\lambda)$  kann also als Quotient von zwei Polynomen in  $e^{-i\lambda}$  dargestellt werden:

$$f_{ij}(\lambda) = \frac{c_{ij,0}(e^{-i\lambda})^0 + c_{ij,1}(e^{-i\lambda})^1 + \dots + c_{ij,r}(e^{-i\lambda})^r}{d_{ij,0}(e^{-i\lambda})^0 + d_{ij,1}(e^{-i\lambda})^1 + \dots + d_{ij,r}(e^{-i\lambda})^r}$$

Spektrale Dichten mit dieser Eigenschaft nennt man *rational*. Man kann zeigen, dass ein (regulärer) Prozess dann und nur dann ein ARMA Prozess ist, wenn er eine rationale spektrale Dichte besitzt.

Im skalaren Fall folgt weiters

$$\underline{f}(z) = \frac{\sigma^2}{2\pi} \frac{b(z)b(z^{-1})}{a(z)a(z^{-1})} = \frac{\sigma^2}{2\pi} \frac{b(z)\tilde{b}(z)}{a(z)\tilde{a}(z)} z^{p-q}$$

wobei  $\tilde{b}(z) = z^q b(z^{-1}) = 1z^q + \dots + b_q z^0$  und  $\tilde{a}(z) = z^p a(z^{-1}) = 1z^p + \dots + a_p z^0$ . Diese rationale Funktion ist (bis auf einen Skalierungsfaktor) durch die Nullstellen und Pole bestimmt, d.h. durch die Nullstellen von  $a(z)$  und  $b(z)$ . Seien nun  $z_1^b, \dots, z_q^b$  und  $z_1^a, \dots, z_p^a$  die Nullstellen von  $b(z)$  bzw. von  $a(z)$ . Dann sind  $z_1^b, (z_1^b)^{-1}, \dots, z_q^b, (z_q^b)^{-1}$  die Nullstellen von  $\underline{f}(z)$  und  $z_1^a, (z_1^a)^{-1}, \dots, z_p^a, (z_p^a)^{-1}$  die Pole von  $\underline{f}(z)$ . Zusätzlich ist  $z = 0$  eine Nullstelle von  $\underline{f}(z)$  im Fall  $p > q$  und  $z = 0$  ist ein Pol wenn  $p < q$  gilt. Daher kann man sich ein ungefähres Bild der spektralen Dichte machen, wenn man die Nullstellen von  $a(z)$  und  $b(z)$  kennt.

Umgekehrt kann man das ARMA System bestimmen, wenn man die spektrale Dichte des Prozesses kennt. Die Pole von  $\underline{f}(z)$  außerhalb des Einheitskreises sind die Nullstellen von  $a(z)$  und damit ist  $a(z)$  eindeutig bestimmt (da  $a(0) = 1$ ). Die Nullstellen von  $\underline{f}(z)$  außerhalb des Einheitskreises sind Nullstellen von  $b(z)$ . Falls  $\underline{f}(z)$  Nullstellen am Einheitskreis hat, dann müssen diese gerade Vielfachheit haben. Diese Nullstellen am Einheitskreis werden auch  $b(z)$  zugeordnet, allerdings nur mit der halben Vielfachheit. Das Polynom  $b(z)$  ist nun durch seine Nullstellen und die Normierung  $b(0) = 1$  eindeutig bestimmt. Die Varianz  $\sigma^2$  erhält man schließlich aus  $\underline{f}(1) = f(0) = \frac{\sigma^2}{2\pi} \frac{b(1)b(1)}{a(1)a(1)}$ . Mit dieser Prozedur bekommen wir automatisch ein (eindeutiges) ARMA System, das die Stabilitäts-Bedingung, die minimum phase Bedingung und die Koprimitätsbedingung erfüllt.

Wir haben also im skalaren Fall gezeigt, dass rationale spektrale Dichten zu ARMA Systemen führen. Eine entsprechende Verallgemeinerung gilt auch für den multivariaten Fall  $n > 1$ .

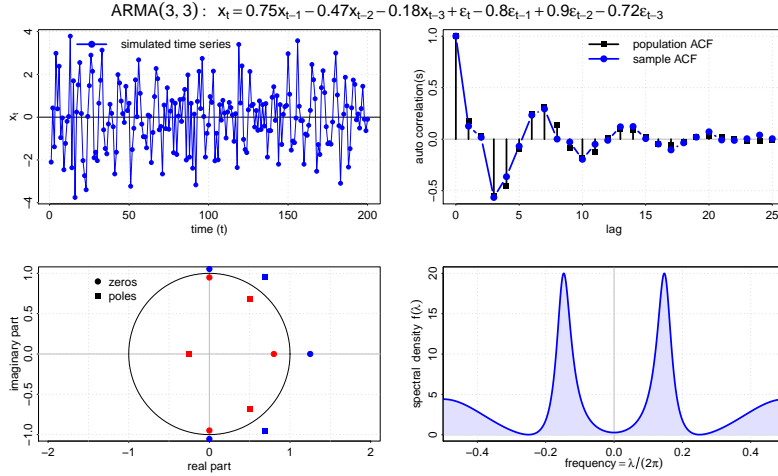


Abbildung 7.1: ARMA(3,3) Prozess  
 $x_t = 0.75x_{t-1} - 0.47x_{t-2} - 0.18x_{t-3} + \epsilon_t - 0.8\epsilon_{t-1} + 0.9\epsilon_{t-2} - 0.72\epsilon_{t-3}$ .

### 7.3 Prognose

Um die Analyse etwas zu vereinfachen, setzen wir hier die etwas stärkere strikte minimum phase Bedingung. Damit folgt eine sogenannte AR( $\infty$ ) Darstellung für  $(x_t)$ :

$$\epsilon_t = b^{-1}(z)a(z)x_t = \sum_{j \geq 0} \bar{a}_j x_{t-j} \quad (7.9)$$

mit  $\bar{a}_0 = I_n$ . Die Stabilitätsbedingung impliziert eine kausale MA( $\infty$ ) Darstellung für  $(x_t)$

$$x_t = a^{-1}(z)b(z)\epsilon_t = \sum_{j \geq 0} \bar{b}_j \epsilon_{t-j} \quad (7.10)$$

wobei  $\bar{b}_0 = I_n$ . Aus (7.9) bzw. (7.10) folgt nun  $\epsilon_s \in \mathbb{H}_x(t)$ ,  $s \leq t$  und  $x_s \in \mathbb{H}_\epsilon(t)$ ,  $s \leq t$  und somit  $\mathbb{H}_x(t) = \mathbb{H}_\epsilon(t)$ . Der ARMA Prozess  $(x_t)$  erfüllt also die Bedingungen von Korollar 3.9 und ist daher regulär. Die  $h$ -Schritt Prognose aus der unendlichen Vergangenheit kann folgendermaßen bestimmt werden:

$$x_{t+h} = \underbrace{\sum_{j=0}^{h-1} \bar{b}_j \epsilon_{t+h-j}}_{\hat{a}_{t,h}} + \underbrace{\sum_{j=h}^{\infty} \bar{b}_j \epsilon_{t+h-j}}_{\hat{x}_{t,h}}$$

Die Prognose  $\hat{x}_{t,h}$  kann auch als Funktion der gegenwärtigen und vergangenen Beobachtungen  $(x_s, s \leq t)$  dargestellt werden, da

$$\hat{x}_{t,h} = \sum_{j=h}^{\infty} \bar{b}_j \epsilon_{t+h-j} = \sum_{j=h}^{\infty} \bar{b}_j \left( \sum_{k=0}^{\infty} \bar{a}_k x_{t+h-j-k} \right) = \sum_{k \geq 0} c_k^h x_{t-k}$$

wobei die Koeffizienten  $c_k^h$  z.B. aus folgender Gleichung bestimmt werden können:

$$\left( \sum_{j=h}^{\infty} \bar{b}_j z^j \right) \left( \sum_{k=0}^{\infty} \bar{a}_k z^k \right) = \sum_{k=0}^{\infty} c_k^h z^k$$

(7.10) ist äquivalent zu

$$x_t = -\bar{a}_1 x_{t-1} - \bar{a}_2 x_{t-2} - \dots + \epsilon_t$$

Die MA( $\infty$ ) Koeffizienten  $(\bar{b}_j)$  sind die Impuls Antwort des ARMA Systems. Diese Koeffizienten haben dieselbe Interpretation wie im AR Fall. Man kann auch eine orthogonalisierte Impuls Antwort und eine Zerlegung der Prognose Fehler Varianz ganz analog zum AR Fall definieren.

Die Einschrittprognose Fehler sind  $\hat{u}_{t,1} = \epsilon_{t+1}$  und daher sind die  $\epsilon_t$ 's die Innovationen des Prozesses ( $x_t$ ).

Folgende Darstellung der Prognose folgt unmittelbar aus der AR-MA Darstellung, der Beziehung  $\mathbb{H}_x(t) = \mathbb{H}_\epsilon(t)$  und  $\hat{x}_{t,h} = \mathbb{P}_{\mathbb{H}_x(t)} x_{t+h}$ :

$$\hat{x}_{t,h} = a_1 \hat{x}_{t,h-1} + a_2 \hat{x}_{t,h-2} + \cdots + a_p \hat{x}_{t,h-p} + b_1 \hat{\epsilon}_{t,h-1} + \cdots + b_q \hat{\epsilon}_{t,h-q}$$

wobei

$$\begin{aligned} \hat{x}_{t,h-k} &= \mathbb{P}_{\mathbb{H}_x(t)} x_{t+h-k} = \begin{cases} \hat{x}_{t,h-k} & h-k > 0 \\ x_{t+h-k} & h-k \leq 0 \end{cases} \\ \hat{\epsilon}_{t,h-k} &= \mathbb{P}_{\mathbb{H}_x(t)} \epsilon_{t+h-k} = \begin{cases} 0 & h-k > 0 \\ \epsilon_{t+h-k} & h-k \leq 0 \end{cases} \end{aligned}$$

## 7.4 Schätzung

Die Schätzung von ARMA Systemen ist erheblich schwieriger als die Schätzung von AR Systemen. Einerseits ist die Struktur des Parameterraums komplizierter (die Bedingungen für Identifizierbarkeit sind viel komplexer) und andererseits gibt es auch keine einfache und trotzdem statistisch effiziente Schätzprozedur (analog zur Yule Walker Schätzung im AR Fall). Diese Schwierigkeiten treten besonders im multivariaten Fall auf.

Die Maximum Likelihood (ML) Schätzung basiert typischerweise auf der Annahme, dass die Fehler und damit auch der beobachtete Prozess normal verteilt sind. Die log Likelihood ist dann gegeben durch

$$l(a_1, \dots, a_p, b_1, \dots, b_q, \Sigma) = -\frac{T}{2} \ln(2\pi) - \frac{1}{2} \ln \det \Gamma_T - \frac{1}{2} (x_T^T)' \Gamma_T^{-1} x_T^T$$

7.11

Die Autokovarianzfunktion (und damit  $\Gamma_T$ ) ist eine Funktion der AR-MA Parameter  $(a_1, \dots, a_p, b_1, \dots, b_q, \Sigma)$ . Der ML Schätzer (d.h. der Maximierer der Likelihood Funktion) muss mit numerischen Optimierungsverfahren berechnet werden, da (außer in einfachen Spezialfällen) keine analytische Lösung existiert. Die Likelihood Funktion kann lokale Maxima haben und der Raum der "zulässigen" Parameter hat keine einfache Struktur. Daher ist die Bestimmung des ML Schätzers keine triviale Aufgabe.

Ein direkte Berechnung der Likelihood mit der oben angegebene Formel wäre numerisch sehr ineffizient. Daher wird die Likelihood als Funktion der Einschritt Prognosefehler (und deren Varianzen) ausgedrückt. Sei  $\hat{u}_t (= \hat{u}_{t-1,1,t-1})$  der Fehler der Prognose für  $x_t$  gegeben die Beobachtungen  $x_1, \dots, x_{t-1}$ . Die entsprechenden Varianzen werden mit  $\Sigma_t = \mathbf{E} \hat{u}_t \hat{u}_t' (= \Sigma_{1,t-1})$  bezeichnet. Der Vektor  $\hat{u}_T = (\hat{u}_T', \dots, \hat{u}_1')'$  kann nun dargestellt werden als

$$\hat{u}_1 = x_1 \text{ und } \Sigma_1 = \gamma(0).$$

$$\hat{u}_T = C x_T^T = \begin{pmatrix} I_n & c_{12} & \cdots & c_{1T} \\ 0 & I_n & \cdots & c_{2T} \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & I_n \end{pmatrix} \begin{pmatrix} x_T \\ x_{T-1} \\ \vdots \\ x_1 \end{pmatrix}$$

wobei  $C$  eine obere Block Dreiecksmatrix mit Einheits Matrizen auf der Hauptdiagonalen ist. Die Varianz-Kovarianz Matrix von  $\hat{u}_T$

$$\Omega_T := \mathbf{E} \hat{u}_T \hat{u}_T' = \begin{pmatrix} \Sigma_T & 0 & \cdots & 0 \\ 0 & \Sigma_{T-1} & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \Sigma_1 \end{pmatrix} = C \Gamma_T C'$$

ist eine Block Diagonal Matrix, da die Einschrift Prognosefehler zu-einander unkorreliert sind. Damit folgt  $\Gamma_T = C^{-1} \Omega_T (C^{-1})'$ ,  $\det \Gamma_T = \det \Omega_T = \prod_{t=1}^T \det \Sigma_t$  und

$$(x_T^T)' \Gamma_T^{-1} x_T^T = (x_T^T)' C' \Omega_T^{-1} C x_T^T = \hat{u}_T' \Omega_T^{-1} \hat{u}_T = \sum_{t=1}^T \hat{u}_t' \Sigma_t^{-1} \hat{u}_t$$

Zusammengefasst folgt nun diese Darstellung der Likelihood Funktion

$$l(a_1, \dots, a_p, b_1, \dots, b_q, \Sigma) = -\frac{T}{2} \ln(2\pi) - \frac{1}{2} \sum_{t=1}^T \ln \det \Sigma_t - \frac{1}{2} \sum_{t=1}^T \hat{u}_t' \Sigma_t^{-1} \hat{u}_t$$

7.12

die mit dem Kalman Filter besonders effizient (see e.g. HAMILTON (1994)<sup>2</sup>) berechnet werden kann. Für große Stichproben kann die Likelihood Funktion auch durch

$$l(a_1, \dots, a_p, b_1, \dots, b_q, \Sigma) \approx -\frac{T}{2} \ln(2\pi) - \frac{1}{2} \sum_{t=1}^T \ln \det \Sigma - \frac{1}{2} \sum_{t=1}^T \hat{\epsilon}_t' \Sigma^{-1} \hat{\epsilon}_t$$

approximiert werden, wobei  $\hat{\epsilon}_t$  Schätzungen der Innovationen  $\epsilon_t$  sind, die durch folgende Rekursion berechnet werden

$$\hat{\epsilon}_t = x_t - a_1 x_{t-1} - \cdots - a_p x_{t-p} - b_1 \hat{\epsilon}_{t-1} - \cdots - b_q \hat{\epsilon}_{t-q}, \quad t = 1, 2, 3, \dots$$

Man kann nun, unter gewissen Regularitätsannahmen, zeigen, dass der ML Schätzer konsistent, asymptotisch normal verteilt und effizient ist. Als Beispiel zitieren wir hier für den skalaren Fall ein Resultat aus BROCKWELL & DAVIS<sup>3</sup> (Theoreme 10.8.1 und 10.8.2). Die Parameter des ARMA Systems werden im Vektor  $\theta = (a_1, \dots, a_p, b_1, \dots, b_q)$  zusammengefasst. Ein Parametervektor  $\theta$  ist "zulässig", wenn die Stabilitäts Bedingung, die strikte minimum phase Bedingung und die Koprimitätsbedingung erfüllt sind. Zusätzlich fordert man  $a_p \neq 0$  und  $b_q \neq 0$ . Die Menge aller zulässigen Parametervektoren wird mit  $\Theta_{p,q} \subset \mathbb{R}^{p+q}$  bezeichnet. Die ML Schätzer  $(\hat{\theta}, \hat{\sigma}^2)$  für  $\theta$  und die Fehler Varianz  $\sigma^2$  sind die Maximierer der (Gauss'schen) Likelihood Funktion über dem Parameter-Raum  $(\Theta_{p,q} \times \mathbb{R}^+)$ .

**Theorem 7.2.** Sei  $(x_t)$  ein ARMA( $p, q$ ) Prozess, der von einem ARMA System  $a(B)x_t = b(B)\epsilon_t$  mit unabhängig identisch verteilten Fehlern  $(\epsilon_t) \sim \text{IID}(\sigma_0^2)$ ,  $\sigma_0^2 > 0$  und Parametern  $\theta_0 \in \Theta_{p,q}$  erzeugt wird. Dann gilt:

$$1. (\hat{\theta}, \hat{\sigma}^2) \longrightarrow (\theta_0, \sigma_0^2) \text{ f.s.}$$

<sup>2</sup> J.D. Hamilton. *Time Series Analysis*. Princeton, NJ: Princeton University Press, 1994

Unbekannte Startwerte  $(x_t, \hat{\epsilon}_t \ t \leq 0)$  werden hier einfach durch Null Vektoren ersetzt.

<sup>3</sup> P.J. Brockwell and R.A. Davis. *Time Series: Theory and Methods*. Springer series in statistics. Springer-Verlag, New York, 2nd edition, 1991

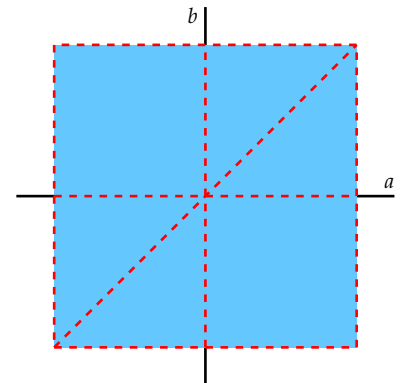


Abbildung 7.2: Die Menge  $\Theta_{1,1} = \{(a, b) \mid |a| < 1, |b| < 1, a \neq 0, b \neq 0, a \neq b\} \subseteq \mathbb{R}^2$ . Die strichlierten Linien sind nicht in  $\Theta_{1,1}$  enthalten.



2. Die ML Schätzer sind asymptotisch normal verteilt, d.h

$$\sqrt{T}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, \sigma_0^2 W^{-1})$$

wobei  $W = \mathbf{E} z_t z_t'$ ,  $z_t = ((u_t^p)', (v_t^q)')'$  und  $(u_t)$ ,  $(v_t)$  sind zwei AR Prozesse, die durch  $a(B)u_t = \epsilon_t$  und  $b(B)v_t = \epsilon_t$  definiert sind.

Die Likelihood Funktion basiert auf der Annahme von normal verteilten Fehlern. Trotzdem gilt obiges Resultat auch für beliebige IID Fehler (quasi ML Schätzung).

### Ordnungsschätzung

Die Ordnung  $(p, q)$  kann wie im AR Fall mit Hilfe von Informationskriterien geschätzt werden. Einige häufig benutzte Informationskriterien für den skalaren Fall sind:

- Akaikes Informations Kriterium:  $AIC(p, q) = -2l(p, q) + 2(p + q)$
- korrigiertes AIC:  $AIC_c(p, q) = AIC(p, q) + \frac{2(p+q+1)(p+q+2)}{T-p-q-2}$
- (Schwarz) Bayes Informations Kriterium:  $BIC(p, q) = -2l(p, q) + \ln(T)(p + q)$

Hier bezeichnet  $l(p, q)$  das Maximum der log Likelihood Funktion eines ARMA Modells mit Ordnung  $(p, q)$ . Der Schätzer für die Ordnung  $(p, q)$  ist das Paar  $(\hat{p}, \hat{q})$  welches das betrachtete Informationskriterium (in einer vorgegebenen Menge von Ordnungen) minimiert. Wenn der wahre Prozess ein ARMA(p,q) Prozess ist, dann liefert BIC eine konsistente Schätzung der Ordnung, während AIC und  $AIC_c$  die Ordnung eher überschätzen.

Für die Modell-Validierung kann man z.B. die Ljung Box Test Statistik verwenden, die im skalaren Fall gegeben ist durch:

$$Q^* = T(T+2) \sum_{k=1}^h \frac{1}{T-k} \hat{\rho}_{\tilde{u}}^2(k)$$

Hier bezeichnet  $\hat{\rho}_{\tilde{u}}(\cdot)$  die Stichproben Autokorrelations-Funktion der normalisierten, Einschnitt Prognosefehler

$$\tilde{u}_t = \frac{\hat{u}_t}{\sigma_t}, \quad \sigma_t^2 = \mathbf{E} \hat{u}_t^2$$

Die normalisierten Prognosefehler sind ein weißes Rauschen (wenn sie mit den wahren Prozess Parametern berechnet werden). Wenn die richtige Modell Ordnung gewählt wurde, dann sind die ML Schätzer, für große Stichproben, nahe bei den wahren Parametern und daher verhalten sich auch die geschätzten, normalisierten Einschnitt Prognose Fehler ähnlich wie ein white noise Prozess. Man kann zeigen, dass die Test Statistik für  $T \rightarrow \infty$  gegen eine Chi-Quadrat Verteilung konvergiert:  $Q^* \xrightarrow{d} \chi_{h-p-q}^2$ . Die normalisierten Einschnitt-Prognose Fehler können auch durch die oben beschriebenen Schätzer  $\hat{\epsilon}_t$  ersetzt werden.

### 7.5 Der SARIMAX Zoo

Viele Modell- bzw. Prozess- Klassen können mit Hilfe von Differenzen Gleichungen analog zum AR bzw. ARMA Fall definiert werden. Einige davon sollen hier exemplarisch aufgezählt werden. Im ARX bzw. ARMAX Fall betrachtet man zusätzlich noch einen Prozess von exogenen Variablen  $(z_t)$ , der stationär und unkorreliert zu den Fehlern (d.h.  $\text{Cov}(z_t, \epsilon_s) = 0 \forall t, s$ ) ist.

$$MA(q): x_t = b(B)\epsilon_t$$

$$AR(p): a(B)x_t = \epsilon_t$$

$$ARMA(p, q): a(B)x_t = b(B)\epsilon_t$$

$$ARX: a(B)x_t = d(B)z_t + \epsilon_t$$

$$ARMAX: a(B)x_t = d(B)z_t + b(B)\epsilon_t$$

$ARIMA(p, d, q): a(B)(1 - B)^d x_t = b(B)\epsilon_t$ . Hier nimmt man an, dass der Prozess  $(x_t)$  selbst nicht stationär ist, aber der Prozess der  $d$ -ten Differenzen  $(1 - B)^d x_t$  soll stationär sein. Genauer gesagt nimmt man an, dass  $(1 - B)^s x_t$  für  $0 \leq s < d$  nicht stationär ist, aber  $(1 - B)^d x_t$  ist stationär<sup>4</sup>. Weiters nimmt man an, dass  $(1 - B)^d x_t$  ein  $ARMA(p, q)$  Prozess ist.

<sup>4</sup> Einen Prozess mit dieser Eigenschaft nennt man *integriert der Ordnung d*. Oft verwendet man die Notation  $(x_t) \sim I(d)$  für solche Prozesse

$$SARIMA(p, d, q) \times (\bar{p}, \bar{d}, \bar{q})_s: a(B)\bar{a}(B^s)(1 - B)^d(1 - B^s)^{\bar{d}} x_t = b(B)\bar{b}(B^s)\epsilon_t.$$

Saisonale ARIMA Systeme können verwendet werden, um Prozesses mit einer starken saisonalen Komponente zu modellieren. Die Periodenlänge der Saison Komponente wird mit  $s$  bezeichnet. Für Quartalsdaten wäre also z.B.  $s = 4$ .

## Mehrgleichungs-Modelle

Interdependente Mehrgleichungs Modelle, SVAR, strukturelle / reduzierte Form, Identifizierbarkeit, exogene / endogene / vorbestimmte Variable, Havelmo Bias, IV Schätzung, 2SLS Schätzung

In der Makroökonomie werden oft so genannte *interdependente Mehrgleichungsmodelle*, d.h. Modelle der Form

$$a_0 y_t = c + a_1 y_{t-1} + \dots + a_p y_{t-p} + d_0 x_t + \dots + d_s x_{t-s} + u_t \quad 8.1$$

betrachtet. Hier bezeichnet  $y_t$  einen  $n$ -dimensionalen Vektor von *endogenen* Variablen,  $x_t$  ist ein  $m$ -dimensionaler Vektor von *exogenen* Variablen und  $u_t$  ist ein  $n$ -dimensionaler Vektor von Störungen. Die Matrizen  $a_i \in \mathbb{R}^{n \times n}$ ,  $d_i \in \mathbb{R}^{n \times m}$ , der Vektor  $c \in \mathbb{R}^n$  und die Varianz  $\Sigma = \mathbf{E} u_t u_t' \in \mathbb{R}^{n \times n}$  sind die Parameter des Modells.

Folgende Annahmen werden gesetzt

1. Die Diagonal Elemente der Matrix  $a_0$  sind gleich eins<sup>1</sup>.
2. Die Prozesse  $(x_t)$  und  $(y_t)$  sind gemeinsam stationär<sup>2</sup>.
3. Die Störungen  $(u_t)$  sind ein white noise Prozess, d.h.  $(u_t) \sim \text{WN}(\Sigma)$ .
4. Der Prozess  $(x_t)$  ist unkorreliert zu den Störungen  $(u_t)$  (zu allen leads und lags), d.h.  $\mathbf{Cov}(x_t, u_s) = \mathbf{E} x_t u_s' = 0$  für alle  $t, s \in \mathbb{Z}$ .

Bei der Analyse dieser Modellklasse treten zwei Schwierigkeiten auf

**IDENTIFIZIERBARKEIT:** Die Parameter des Modells sind i.A. *nicht* identifiziert, d.h. ohne weitere Annahmen existieren mehrere äquivalente Darstellungen für den Prozess  $(y_t)$ . Multipliziert man die Gleichung (8.1) von links mit einer nicht singulären Matrix  $T$ , so erhält man ein äquivalentes Modell

$$\tilde{a}_0 y_t = \tilde{c} + \tilde{a}_1 y_{t-1} + \dots + \tilde{a}_p y_{t-p} + \tilde{d}_0 x_t + \dots + \tilde{d}_s x_{t-s} + \tilde{u}_t \quad 8.2$$

wobei  $\tilde{a}_i = T a_i$  (für  $i = 0, \dots, p$ ),  $\tilde{d}_i = T d_i$  (für  $i = 0, \dots, s$ ),  $\tilde{c} = T c$  und  $\tilde{u}_t = T u_t$ . Die einzige Bedingung an  $T$  ist, dass die Diagonalelemente von  $T a_0$  gleich eins sein müssen. Daher benötigt man Restriktionen an die Parameter, um Identifizierbarkeit zu erreichen. Typischerweise sind das "Null-Restriktionen", d.h. man verlangt, dass bestimmte Elemente der Parameter Matrizen gleich null sind. Diese Restriktionen sollten natürlich nicht willkürlich gewählt werden sondern das a

<sup>1</sup> Das heißt, die  $i$ -te Gleichung beschreibt die  $i$ -te Variable als Funktion der eigenen lags und als Funktion der anderen endogenen und der exogenen Variablen.

<sup>2</sup> Viele Makro-ökonomische Variable sind nicht stationär (integriert) und man formuliert daher ein Mehrgleichungsmodell für die ersten Differenzen bzw. für die Wachstumsraten. Für eine direkte Modellierung der Variablen bieten sich zum Beispiel Kointegrationsmodelle an.

priori Wissen über die Beziehungen der Variablen zueinander widerspiegeln. Das Modell (8.1) zusammen mit den Restriktionen nennt man *strukturelle Form*.

**SCHÄTZUNG:** Eine (naive) Schätzung der Modellparameter mit OLS liefert im Allgemeinen verzerrte Schätzer (Haavelmo Bias). Dieses Problem werden wir in Abschnitt Section 8.3 genauer diskutieren.

Das Mehrgleichungs-Modell kann auch mittels Filtern geschrieben werden

$$a(B)y_t = c + d(B)x_t + u_t$$

wobei

$$\begin{aligned} a(B) &= a_0 - a_1 B - \dots - a_p B^p \\ d(B) &= d_0 + d_1 B + \dots + d_s B^s \end{aligned}$$

Die z-Transformationen dieser Filter sind  $a(z) = a_0 - a_1 z - \dots - a_p z^p$  und  $d(z) = d_0 + d_1 z + \dots + d_s z^s$ . Das dynamische Mehrgleichungsmodell ist im Prinzip ein VARX Modell, allerdings ohne die Normalisierung  $a(0) = a_0 = I_n$ . Es ist auch klar, dass eine eindeutige stationäre Lösung für  $(y_t)$  existiert, wenn  $\det(a(z)) \neq 0$  für  $|z| = 1$  gilt, da in diesem Fall der Filter  $a(B)$  invertierbar ist. Wenn die stärkere Stabilitätsbedingung

$$\det(a(z)) \neq 0 \quad \forall |z| \leq 1 \quad \boxed{8.3}$$

erfüllt ist, dann ist die stationäre Lösung kausal d.h.

$$\begin{aligned} y_t &= a^{-1}(B)c + a^{-1}(B)d(B)x_t + a^{-1}(B)u_t \\ &= \mu + \sum_{j \geq 0} \bar{d}_j x_{t-j} + \sum_{j \geq 0} \bar{b}_j u_{t-j} \end{aligned} \quad \boxed{8.4}$$

Weiters folgt, dass  $(a_0^{-1}u_{t+1})$  die bedingten Einschnittprognosefehler sind, d.h. die Fehler der Prognose für  $y_{t+1}$  gegeben  $y_s, s \leq t$  und  $x_s, s \in \mathbb{Z}$ .

Bei der Konstruktion der Lösung(en) von (8.1) kann im Vergleich zum Standard VARX Modell ( $a_0 = I$ ) noch eine Schwierigkeit auftreten. Wenn die Polynom-Matrix  $a(z)$  für  $z \in \mathbb{C}$  singular ist ( $\det(a(z)) \equiv 0$ ) dann existiert entweder überhaupt keine Lösung (auch keine nicht-stationäre Lösung), oder es existieren – auch für gegebene Startwerte – unendlich viele Lösungen. Das zeigt folgendes einfache Beispiel:

*inkonsistente Gleichungen*

$$\begin{aligned} y_{1t} + 2y_{2t} &= u_{1t} \\ 0.5y_{1t} + y_{2t} &= u_{2t} \quad \text{wobei } u_{2t} \neq 0.5u_{1t} \end{aligned}$$

*unvollständige Gleichungen*

$$\begin{aligned} y_{1t} + 2y_{2t} &= u_{1t} \\ 0.5y_{1t} + y_{2t} &= u_{2t} \quad \text{wobei } u_{2t} = 0.5u_{1t} \end{aligned}$$

Von jetzt an setzen wir immer voraus, dass die Stabilitätsbedingung (8.3) erfüllt ist, die insbesondere impliziert, dass  $a(0) = a_0$  regulär ist.

## 8.1 Reduzierte Form

Multiplikation der Gleichung (8.1) von links mit  $a_0^{-1}$  liefert ein Modell der Form

$$y_t = c^* + a_1^* y_{t-1} + \dots + a_p^* y_{t-p} + d_0^* x_t + \dots + d_s^* x_{t-s} + \epsilon_t \quad (8.5)$$

Das ist die sogenannte *reduzierte Form* (oder *finale Form*) des Mehrgleichungs Modells. Die Stabilitätsbedingung

$$\det(a^*(z)) \neq 0 \quad \forall |z| \leq 1 \quad (8.6)$$

für dieses Modell ist natürlich äquivalent zu (8.3) und man erhält auch dieselbe stationäre Lösung

$$\begin{aligned} y_t &= (a^*)^{-1}(B)c^* + (a^*)^{-1}(B)d^*(B)x_t + (a^*)^{-1}(B)\epsilon_t \\ &= \mu + \sum_{j \geq 0} \bar{d}_j x_{t-j} + \sum_{j \geq 0} \bar{b}_j^* \epsilon_{t-j} \end{aligned} \quad (8.7)$$

wie in (8.4). Die  $\epsilon_t$ 's sind die bedingten Einschnitt-Prognosefehler ( $\epsilon_t = a_0^{-1}u_t$  und  $\bar{b}_j^* = \bar{b}_j a_0$ ). Die Parameter der reduzierten Form sind identifiziert, wenn man folgende Annahmen setzt

1.  $E\epsilon_t \epsilon_t' = \Sigma^* > 0$
2.  $\Gamma_{x,s+1} = \text{Var}(x_t^{s+1}) > 0$

Um diese Behauptung zu zeigen, zerlegen wir die endogenen Variablen in

$$y_t = \hat{y}_t + \tilde{y}_t$$

wobei

$$\begin{aligned} \hat{y}_t &= \mu + \sum_{j \geq 0} \bar{d}_j x_{t-j} = c^* + a_1^* \hat{y}_{t-1} + \dots + a_p^* \hat{y}_{t-p} + d_0^* x_t + \dots + d_s^* x_{t-s} \\ \tilde{y}_t &= \sum_{j \geq 0} \bar{b}_j^* \epsilon_{t-j} = a_1^* \tilde{y}_{t-1} + \dots + a_p^* \tilde{y}_{t-p} + \epsilon_t \end{aligned}$$

Der erste Teil  $\hat{y}_t$  bezeichnet die beste Approximation von  $y_t$  durch eine lineare Funktion der exogenen Variablen ( $x_s, s \in \mathbb{Z}$ ). Der entsprechende Approximationsfehler  $\tilde{y}_t$  ist ein regulärer AR(p) Prozess mit Innovationen  $\epsilon_t$ .

Die reduzierte Form kann als multivariates, inhomogenes Regressions Modell

$$y_t = \beta^* z_t + \epsilon_t$$

dargestellt werden, wobei  $z_t$

$$z_t = \begin{pmatrix} y_{t-1} \\ \vdots \\ y_{t-p} \\ x_t \\ \vdots \\ x_{t-s} \\ 1 \end{pmatrix} = \begin{pmatrix} \tilde{y}_{t-1} \\ \vdots \\ \tilde{y}_{t-p} \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} \hat{y}_{t-1} \\ \vdots \\ \hat{y}_{t-p} \\ x_t \\ \vdots \\ x_{t-s} \\ 1 \end{pmatrix} = \begin{pmatrix} y_{t-1}^p \\ x_t^{s+1} \\ 1 \end{pmatrix} = \begin{pmatrix} \tilde{y}_{t-1}^p \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} \hat{y}_{t-1}^p \\ x_t^{s+1} \\ 1 \end{pmatrix}$$

Hier bezeichnet  $'^*$  die Parameter der reduzierten Form – nicht die Hermit'sch transponierten Matrizen.

Mit der üblichen Notation schreiben wir

$$\begin{aligned} a^*(B) &= I_n - a_1^* B - \dots - a_p^* B^p = a_0^{-1} a(B) \\ d^*(B) &= d_0^* + d_1^* B + \dots + d_s^* B^s = a_0^{-1} d(B) \end{aligned}$$

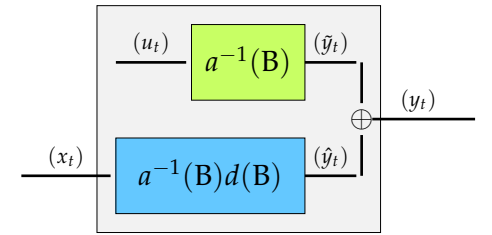


Abbildung 8.1: Block Schema eines interdependenten Mehrgleichungs-Modells (ohne Konstante c).

ein Vektor von *vorbestimmten* Variablen (*predetermined*) ist, da

$$\mathbf{E}z_t \epsilon'_s = 0 \quad \forall s \geq t$$

Die Regression Koeffizienten  $\beta^* = (a_1^*, \dots, a_p^*, d_0^*, \dots, d_s^*, c^*)$  erfüllen daher die Normalgleichungen

$$\mathbf{E}y_t z'_t = \beta^* \mathbf{E}z_t z'_t$$

Wenn  $\mathbf{E}z_t z'_t$  positiv definit ist, dann sind die Parameter eindeutig aus dieser Gleichung bestimmt, d.h. die Parameter sind wie behauptet *identifiziert*. Die Matrix  $\mathbf{E}z_t z'_t$  ist dann und nur dann positiv definit, wenn die Varianz-Kovarianz Matrix

$$\mathbf{Var} \left[ \begin{pmatrix} y_{t-1}^p \\ x_t^s \end{pmatrix} \right] = \mathbf{Var} \left[ \begin{pmatrix} \tilde{y}_{t-1}^p \\ 0 \end{pmatrix} + \begin{pmatrix} \hat{y}_{t-1}^p \\ x_t^s \end{pmatrix} \right] = \begin{pmatrix} \Gamma_{\tilde{y},p} & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} \Gamma_{\hat{y},p} & \mathbf{Cov}(\hat{y}_{t-1}^p, x_t^{s+1}) \\ \mathbf{Cov}(x_t^{s+1}, \hat{y}_{t-1}^p) & \Gamma_{x,s+1} \end{pmatrix}$$

positiv definit ist. Hier haben wir  $y_t = \hat{y}_t + \tilde{y}_t$  und  $\mathbf{Cov}(\hat{y}_t, \tilde{y}_s) = 0$ ,  $\mathbf{Cov}(x_t, \tilde{y}_s) = 0$  für alle  $t, s \in \mathbb{Z}$  benutzt. Das gewünschte Resultat  $\mathbf{E}z_t z'_t > 0$  folgt nun aus der Annahme  $\Sigma^* > 0$  (und damit  $\Gamma_{\tilde{y},p} > 0$ ) und der Annahme  $\Gamma_{x,s+1} > 0$ .

Die Parameter der reduzierten Form können konsistent mit OLS

$$\hat{\beta}_{OLS}^* = \left( \frac{1}{T} \sum_{t=1}^T y_t z'_t \right) \left( \frac{1}{T} \sum_{t=1}^T z_t z'_t \right)^{-1}$$

geschätzt werden, wenn die Stichproben Momente konsistente Schätzer sind. Es ist auch leicht zu zeigen, dass die OLS Schätzer asymptotisch normal verteilt sind (analog zu Theorem 6.7 für den AR Fall), wenn man die “üblichen” Annahmen an die Innovationen ( $\epsilon_t$ ) und den Prozess ( $x_t$ ) stellt.

## 8.2 Identifizierbarkeit der strukturellen Form

Wie schon am Anfang dieses Kapitels, ist die strukturelle Form ohne zusätzliche Restriktionen nicht identifizierbar. Allerdings ist die reduzierte Form (unter den oben gesetzten Annahmen) identifizierbar und daher sind zwei strukturelle Modelle mit Parametern  $a_0$  und  $\beta = (a_1, \dots, a_p, d_0, \dots, d_s, c)$  bzw.  $\tilde{a}_0$  und  $\tilde{\beta} = (\tilde{a}_1, \dots, \tilde{a}_p, \tilde{d}_0, \dots, \tilde{d}_s, \tilde{c})$  dann und nur dann äquivalent zueinander, wenn die entsprechenden reduzierten Formen übereinstimmen. Das heißt, es muss

$$a_0^{-1} \beta = \beta^* = \tilde{a}_0^{-1} \tilde{\beta}$$

gelten. Diese Relation gilt dann und nur dann, wenn eine nicht singuläre Matrix  $T$  existiert, sodass

$$(\tilde{a}_0, \tilde{\beta}) = T(a_0, \beta)$$

und

$$[Ta_0]_{ii} = [\tilde{a}_0]_{ii} = 1 \quad \text{for } i = 1, \dots, n$$

8.8

Natürlich reicht (8.8) nicht aus, um Identifizierbarkeit zu erreichen, d.h. um sicher zu stellen, dass  $T = I_n$  die einzige mögliche Wahl

ist. Wir benötigen zumindest  $n^2 - n = n(n - 1)$  zusätzliche Restriktionen. Typischerweise sind das "Null-Restriktionen", d.h. bestimmte Elemente der Koeffizienten Matrizen werden gleich Null gesetzt. Nehmen wir nun an, dass  $m \geq (n - 1)$  Einträge in der  $i$ -ten Zeile von  $(a_0, \beta)$  a priori auf Null gesetzt werden. Weiters sei  $T_i$  die  $i$ -te Zeile von  $T$  und  $\gamma^{(i)} \in \mathbb{R}^{(n-1) \times m}$  die Teilmatrix von  $(a_0, \beta)$  die man erhält, wenn man die zu den Null-Restriktionen gehörigen Spalten selektiert. Die Null Restriktionen in der  $i$ -ten Zeile von  $(a_0, \beta)$  sind dann äquivalent zu der Gleichung  $T_i \gamma^{(i)} = 0 \in \mathbb{R}^{1 \times m}$ . Per Konstruktion ist die  $i$ -te Zeile von  $\gamma^{(i)}$  gleich Null. Hat nun die Matrix  $\gamma^{(i)}$  Rang  $n - 1$ , so müssen alle Elemente von  $T_i$ , mit Ausnahme des  $i$ -ten Elements, gleich Null sein. Wegen der Normierung  $[Ta_0]_{ii} = 1$  folgt weiter, dass das  $i$ -te Element von  $T_i$  gleich Eins sein muss.

Wenn wir nun in jeder Zeile von  $(a_0, \beta)$  (d.h. in jeder Gleichung)  $m \geq n - 1$  solcher Null Restriktion haben, dann muss  $T = I_n$  gelten und damit folgt wie gewünscht die Identifizierbarkeit der strukturellen Form.

**Beispiel 8.1.** Das Klein Modell I:

$$\begin{aligned} C_t &= c_1 + c_1 Q_t + c_2 Q_{t-1} + c_3 W_t + u_{1t} \\ I_t &= i_0 + i_1 Q_t + i_2 Q_{t-1} + i_3 K_{t-1} + u_{2t} \\ W_t^{\text{pr}} &= w_0 + w_1 E_t + w_2 E_{t-1} + w_3 (t - t_0) + u_{3t} \\ Y_t &= C_t + I_t + G_t - T_t^{\text{ind}} \\ Q_t &= Y_t - W_t \\ K_t &= K_{t-1} + I_t \\ W_t &= W_t^{\text{pr}} + W_t^{\text{G}} \\ E_t &= Y_t + T_t^{\text{ind}} - W_t^{\text{G}} \end{aligned}$$

Die endogenen Variablen sind

$C_t$  ... consumption  
 $I_t$  ... investments  
 $W_t^{\text{pr}}$  ... wages in the private sector  
 $Y_t$  ... income  
 $Q_t$  ... profits  
 $K_t$  ... capital stock  
 $W_t$  ... wages  
 $E_t$  ... net social product of the private sector

und die exogenen Variablen sind

$W_t^{\text{G}}$  ... wages in the governmental sector  
 $T_t^{\text{ind}}$  ... indirect taxes  
 $G_t$  ... government expenditures  
 $t - t_0$  ... time

Die endogenen und die exogenen Variablen werden in die beiden

In der Praxis überprüft man oft nur die Zahl der Null Restriktionen, da die "Rang-Bedingung" ( $\gamma^{(i)}$  hat Rang  $(n - 1)$ ) generisch erfüllt ist, wenn die Position der Nullen kein Widerspruch zu dieser Rang Bedingung ist. Man spricht dann auch von einer "Zähl-Bedingung".

Das Klein Modell beinhaltet fünf Bilanz Gleichungen (also Gleichungen mit bekannten Koeffizienten, die ohne Fehler erfüllt sind). Das ist typisch für Mehrgleichungs Modelle.

Vektoren  $y_t$  und  $x_t$  zusammengefasst

$$y_t = \begin{pmatrix} C_t \\ I_t \\ W_t^{\text{Pr}} \\ Y_t \\ Q_t \\ K_t \\ W_t \\ E_t \end{pmatrix} \quad \text{und} \quad x_t = \begin{pmatrix} W_t^{\text{G}} \\ T_t^{\text{ind}} \\ G_t \\ t - t_0 \end{pmatrix}$$

Die Ordnungen sind  $p = 1$  und  $s = 0$  und die Parameter des Modells sind

$$a_0 = \begin{pmatrix} 1 & 0 & 0 & 0 & -c_1 & 0 & -c_3 & 0 \\ 0 & 1 & 0 & 0 & -i_1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & -w_1 \\ -1 & -1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 1 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & 0 & 1 \end{pmatrix}, \quad a_1 = \begin{pmatrix} 0 & 0 & 0 & 0 & c_2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & i_2 & i_3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & w_2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$d_0 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & w_3 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 \end{pmatrix} \quad \text{and} \quad c = \begin{pmatrix} c_0 \\ i_0 \\ w_0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

Das Klein Modell I erfüllt die (Zähl-) Bedingungen für die Identifizierbarkeit der strukturellen Form, da die Diagonal Elemente von  $a_0$  gleich Eins sind und es in jeder Zeile von  $(c, a_0, a_1, d_0)$  mindestens  $(n - 1) = 7$  a priori Nullen gibt.  $\square$

### 8.3 Schätzung der strukturellen Form

Die strukturelle Form basiert auf ökonomischen Überlegungen und die Parameter der strukturellen Form haben daher auch eine entsprechende ökonomische Interpretation. Die Parameter der reduzierten Form hingegen hängen nicht linear von den Parametern der strukturellen Form ab und haben keine direkte ökonomische Interpretation. Daher ist man vor allem an den Parametern der strukturellen Form interessiert. Leider ist eine direkte Schätzung dieser Parameter mit OLS nicht zielführend, da OLS verzerrte Schätzungen liefert, wie folgendes einfache Beispiel zeigt.

**Beispiel 8.2.** Ein einfaches Keynesianisches Modell

$$C_t = \beta Y_t + u_{1t} \quad \boxed{8.9}$$

$$Y_t = C_t + G_t \quad \boxed{8.10}$$



mit

- $C_t \dots$  Consumption (endogenous)
- $Y_t \dots$  (disposable) income (endogenous)
- $G_t \dots$  private investments and government expenditures (exogenous)

Die strukturelle Form dieses Modells ist

$$\begin{pmatrix} 1 & -\beta \\ -1 & 1 \end{pmatrix} \begin{pmatrix} C_t \\ Y_t \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} G_t + \begin{pmatrix} u_{1t} \\ 0 \end{pmatrix}$$

und die reduzierte Form ist

$$\begin{pmatrix} C_t \\ Y_t \end{pmatrix} = \begin{pmatrix} \frac{\beta}{1-\beta} \\ \frac{1}{1-\beta} \end{pmatrix} G_t + \begin{pmatrix} \frac{1}{1-\beta} \\ \frac{1}{1-\beta} \end{pmatrix} u_{1t}$$

Das Einkommen  $Y_t$  ist korreliert mit den Störungen  $u_{1t}$  der Konsumgleichung, da

$$\mathbf{E}Y_t u_{1t} = \mathbf{E}\left(\frac{\beta}{1-\beta}G_t + \frac{1}{1-\beta}u_{1t}\right)u_{1t} = \frac{1}{1-\beta}\mathbf{E}u_{1t}^2 \neq 0$$

Die exogenen Variablen und die Störungen sind unkorreliert ( $\mathbf{E}G_t u_s = 0$ )!

Daher liefert OLS einen verzerrten und nicht-konsistenten Schätzer für  $\beta$  in der ersten Gleichung von (8.9)! Es gilt

$$\text{plim}(\hat{\beta} - \beta) = \text{plim}\left(\frac{1}{T} \sum Y_t^2\right)^{-1} \underbrace{\text{plim} \frac{1}{T} \sum Y_t u_t}_{\mathbf{E}Y_t u_t} \neq 0$$

□

Wir betrachten nun den allgemeinen Fall und schreiben dazu die  $i$ -te Gleichung folgendermaßen an:

$$y_{it} = y_t^{(i)} \alpha^{(i)} + z_t^{(i)} \beta^{(i)} + u_{it}$$

Terme, die zu den a Priori Nullen gehören, lassen wir dabei weg, d.h.  $y_t^{(i)}$  ist ein Teilvektor von  $y_t'$  und  $z_t^{(i)}$  ist ein Teilvektor des Vektors  $z_t' = (y_{t-1}', \dots, y_{t-p}', x_t', \dots, x_{t-s}', 1)$  der vorbestimmten Variablen. Die (Spalten-) Vektoren  $\alpha^{(i)}$  und  $\beta^{(i)}$  enthalten die entsprechenden Parameter der strukturellen Form. Der OLS Schätzer für diese Parameter (der  $i$ -ten Gleichung) ist gleich

Der Einfachheit halber betrachten wir nur den Fall von "Nullen". Andere Restriktion, wie z.B. der Fall, dass ein Koeffizient gleich Eins sein muss, werden wir nicht diskutieren.

$$\begin{pmatrix} \hat{\alpha}^{(i)} \\ \hat{\beta}^{(i)} \end{pmatrix} = \left[ \frac{1}{T} \sum_{t=1}^T \begin{pmatrix} (y_t^{(i)})' \\ (z_t^{(i)})' \end{pmatrix} \begin{pmatrix} y_t^{(i)} & z_t^{(i)} \end{pmatrix} \right]^{-1} \left[ \frac{1}{T} \sum_{t=1}^T \begin{pmatrix} (y_t^{(i)})' \\ (z_t^{(i)})' \end{pmatrix} y_{it} \right]$$

und die Schätzfehler sind

$$\begin{pmatrix} \hat{\alpha}^{(i)} - \alpha^{(i)} \\ \hat{\beta}^{(i)} - \beta^{(i)} \end{pmatrix} = \left[ \frac{1}{T} \sum_{t=1}^T \begin{pmatrix} (y_t^{(i)})' \\ (z_t^{(i)})' \end{pmatrix} \begin{pmatrix} y_t^{(i)} & z_t^{(i)} \end{pmatrix} \right]^{-1} \left[ \frac{1}{T} \sum_{t=1}^T \begin{pmatrix} (y_t^{(i)})' \\ (z_t^{(i)})' \end{pmatrix} u_{it} \right]$$

Wenn man annimmt, dass die Stichproben-Momente konsistente Schätzer der Populations-Momente sind, dann folgt

$$\text{plim}_{T \rightarrow \infty} \begin{pmatrix} \hat{\alpha}^{(i)} - \alpha^{(i)} \\ \hat{\beta}^{(i)} - \beta^{(i)} \end{pmatrix} = \left[ \mathbf{E} \begin{pmatrix} (y_t^{(i)})' y_t^{(i)} & (y_t^{(i)})' z_t^{(i)} \\ (z_t^{(i)})' y_t^{(i)} & (z_t^{(i)})' z_t^{(i)} \end{pmatrix} \right]^{-1} \left[ \mathbf{E} \begin{pmatrix} (y_t^{(i)})' u_{it} \\ (z_t^{(i)})' u_{it} \end{pmatrix} \right]$$

Der Vektor  $z_t^{(i)}$  enthält nur vorbestimmte (predetermined) Variable d.h.  $\mathbf{E}u_s z_t^{(i)} = 0$  für alle  $s > t$  und daher insbesondere auch  $\mathbf{E}u_{it} z_t^{(i)} = 0$ . Allerdings folgt aus der Gleichung (8.4), dass

$$\mathbf{E}y_t u_t' = \underbrace{\mathbf{E}\hat{y}_t u_t'}_{=0} + \mathbf{E}\tilde{y}_t u_t' = \sum_{j \geq 0} \bar{b}_j u_{t-j} u_t' = \bar{b}_0 \Sigma = a_0^{-1} \Sigma \quad \boxed{8.11}$$

und daher ist  $\mathbf{E}u_{it}(y_t^{(i)})$  im Allgemeinen nicht gleich Null. Der OLS Schätzer ist also asymptotisch verzerrt. Der wesentliche Grund für diese Verzerrung ist, dass die "Regressoren"  $y_t^{(i)}$  auch von den gegenwärtigen Störungen  $u_t$  abhängen.

### Schätzung mit Hilfe der reduzierten Form

Im Fall von genau  $(n - 1)$  Null Restriktionen in jeder Gleichung (*just identified case*) gibt es eine eins-zu-eins Beziehung zwischen den Parametern der strukturellen Form und den Parametern der reduzierten Form. Sei  $\theta$  der Vektor der Parameter der strukturellen Form (ohne Nullen und ohne die Diagonalelemente von  $a_0$ ) und  $\theta^* = \text{vec}(a_1^*, \dots, a_p^*, d_0^*, \dots, d_s^*, c^*)$  der Vektor der Parameter der reduzierten Form, dann existiert eine bijektive Abbildung  $g(\cdot)$  sodass  $\theta^* = g(\theta)$  und  $\theta = g^{-1}(\theta^*)$ . In diesem Fall ist  $\hat{\theta} = g^{-1}(\hat{\theta}_{OLS}^*)$  ein konsistenter Schätzer für  $\theta$ , wobei  $\hat{\theta}_{OLS}^*$  die OLS Schätzung der Parameter  $\theta^*$  der reduzierten Form bezeichnet.

Üblicherweise hat man aber viel mehr Restriktionen, als für die Identifizierbarkeit notwendig wären (*over identified case / über-identifizierte Fall*). Siehe z.B. das oben vorgestellte Klein Modell. Die Menge  $\Theta_r^*$  aller Parameter der reduzierten Form, die die a priori Bedingungen erfüllen, ist daher eine sehr "dünne" Teilmenge der Menge  $\Theta^*$  aller Parameter  $\theta^*$ . Der OLS Schätzer  $\hat{\theta}_{OLS}^*$  wird daher (fast sicher) nicht in der Menge  $\Theta_r^*$  enthalten sein, obwohl der wahre Parameter Vektor  $\theta^*$  ein Element von  $\Theta_r^*$  ist und  $\hat{\theta}_{OLS}^*$  ein konsistenter Schätzer von  $\theta^*$  ist. Daher existiert (fast sicher) kein Parameter Vektor  $\hat{\theta}$  (der strukturellen Form), der alle Restriktionen erfüllt und der dem OLS Schätzer  $\hat{\theta}_{OLS}^*$  entspricht.

### Rekursive Systeme:

$$y_{it} = y_t^{(i)} \alpha^{(i)} + z_t^{(i)} \beta^{(i)} + u_{it}$$

Wenn die Varianz-Kovarianzmatrix  $\Sigma = \mathbf{E}u_t u_t'$  der Fehler diagonal ist und  $a_0$  (nach einer geeigneten Umordnung der endogenen Variablen) eine untere Dreiecksmatrix ist, dann liefert OLS konsistente Schätzer:

1. Gleichung:  $y_{1t} = z_t^{(1)} \beta^{(1)} + u_{1t}$
2. Gleichung:  $y_{2t} = -a_{0,21} y_{1t} + z_t^{(2)} \beta^{(2)} + u_{2t}$
3. Gleichung:  $y_{3t} = -a_{0,31} y_{1t} - a_{0,32} y_{2t} + z_t^{(3)} \beta^{(3)} + u_{3t}$
- .....

Mit der Annahmen  $\mathbf{E}x_{t-j} u_t' = 0 \forall j$  folgt  $\mathbf{E}y_{t-j} u_t' = \mathbf{E}\hat{y}_{t-j} u_t' + \mathbf{E}\tilde{y}_{t-j} u_t' = 0$  für alle  $j > 0$ , da  $\hat{y}_{t-j}$  eine Funktion der exogenen Variablen  $x_s$ ,  $s \leq t-j$  ist und da  $\tilde{y}_{t-j}$  nur von vergangenen Störungen  $u_s$ ,  $s \leq t-j$  abhängt.

Der über identifizierte Fall hat aber *statistische* Vorteile, da i.A. natürlich die Schätzung der unbekannten Parameter umso einfacher wird, je mehr a priori Information man über die Parameter hat. Das Problem ist nur, wie man einen (konsistenten) Schätzer konstruiert, der dieses a priori Wissen über die Parameter auch effizient ausnützt.

Falls einige  $a_{0,ij}$ ,  $i > j$  gleich Null sind, dann lässt man die entsprechenden Terme natürlich weg.

Nach Gleichung (8.11) und mit den obigen Annahmen folgt dass  $Ey_t u_t' = a_0^{-1} \Sigma$  eine untere Dreiecksmatrix ist. Daher gilt  $Ey_{kt} u_{it} = 0$  für  $k < i$  und alle Regressoren der  $i$ -ten Gleichung sind unkorreliert zu den entsprechenden Störungen  $u_{it}$ . Daraus folgt nun, dass die Parameter der  $i$ -ten Gleichung konsistent mit OLS geschätzt werden.

### 2-stufige Kleinstquadrate Schätzung (2SLS)

Die asymptotische Verzerrung des OLS Schätzers wird von den Regressoren  $y_t^{(i)}$  verursacht. Daher liegt es nahe, geeignete Instrumente  $w_t^{(i)}$  für  $y_t^{(i)}$  zu benutzen. Diese Instrumente müssen folgende Bedingungen erfüllen

$$\text{plim} \frac{1}{T} \sum_{t=1}^T \begin{pmatrix} (w_t^{(i)})' w_t^{(i)} & (w_t^{(i)})' z_t^{(i)} \\ (z_t^{(i)})' w_t^{(i)} & (z_t^{(i)})' z_t^{(i)} \end{pmatrix} \text{ existiert und ist nicht singulär}$$

und

$$\text{plim} \frac{1}{T} \sum_{t=1}^T \begin{pmatrix} (w_t^{(i)})' u_{it} \\ (z_t^{(i)})' u_{it} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Dann sind die Instrumenten-Variablen Schätzer

$$\begin{pmatrix} \hat{\alpha}_{IV}^{(i)} \\ \hat{\beta}_{IV}^{(i)} \end{pmatrix} = \left[ \frac{1}{T} \sum_{t=1}^T \begin{pmatrix} (w_t^{(i)})' w_t^{(i)} & (w_t^{(i)})' z_t^{(i)} \\ (z_t^{(i)})' w_t^{(i)} & (z_t^{(i)})' z_t^{(i)} \end{pmatrix} \right]^{-1} \left[ \frac{1}{T} \sum_{t=1}^T \begin{pmatrix} (w_t^{(i)})' y_{it} \\ (z_t^{(i)})' y_{it} \end{pmatrix} \right]$$

(schwach) konsistent.

Der Vektor  $y_t^{(i)}$  kann durch  $y_t^{(i)} = y_t' S^{(i)}$  dargestellt werden, wobei  $S^{(i)}$  eine geeignete "Selektions-Matrix" ist. Man kann zeigen, dass die Instrumente  $w_t^{(i)} = w_t' S^{(i)}$ , wobei

$$w_t = c^* + a_1^* y_{t-1} + \dots + a_p^* y_{t-p} + d_0^* x_t + \dots + d_s^* x_{t-s} \quad \boxed{8.12}$$

diese Bedingungen erfüllen (wenn das System die oben diskutierten Bedingungen für die Identifizierbarkeit erfüllt). Natürlich sind die Parameter der reduzierten Form nicht bekannt und  $w_t^{(i)}$  kann nicht berechnet werden. Um berechenbare Schätzer zu erhalten, ersetzt man nun die unbekannten Parameter der reduzierten Form durch ihre OLS Schätzer (die ja konsistent sind). Sei

$$\hat{w}_t = \hat{c}^* + \hat{a}_1^* y_{t-1} + \dots + \hat{a}_p^* y_{t-p} + \hat{d}_0^* x_t + \dots + \hat{d}_s^* x_{t-s} \quad \boxed{8.13}$$

dann setzt man  $\hat{w}_t^{(i)} = \hat{w}_t' S^{(i)}$ . Unter geeigneten Annahmen liefert diese Konstruktion Instrumente, die die oben angegebenen Bedingungen für konsistente Instrumenten-Variablen Schätzung erfüllen.

Zusammengefasst läuft die 2-stufige Kleinstquadrate Schätzung also in folgenden zwei Schritten ab:

1. Berechne die OLS Schätzer  $\hat{c}^*, \hat{a}_1^*, \dots, \hat{a}_p^*, \hat{d}_0^*, \dots, \hat{d}_s^*$  für die Parameter der reduzierten Form
2. Berechne die Instrumente  $\hat{w}_t^{(i)} = \hat{w}_t' S^{(i)} = (\hat{c}^* + \hat{a}_1^* y_{t-1} + \dots + \hat{a}_p^* y_{t-p} + \hat{d}_0^* x_t + \dots + \hat{d}_s^* x_{t-s})' S^{(i)}$  und damit die IV Schätzer für die Parameter der strukturellen Form.

### 8.4 SVAR Modelle

In diesem Abschnitt betrachten wir (vor allem der einfachen Notation halber) keine exogene Variablen  $x_t$  und keine Konstante 1. Die reduzierte Form des Mehrgleichungs-Modells ist daher ein "klassische" AR Modell

$$y_t = a_1^* y_{t-1} + \dots + a_p^* y_{t-p} + \epsilon_t \quad 8.14$$

und wegen der Stabilitätsannahme (8.6) hat der Prozess  $(y_t)$  eine kausale MA Darstellung

$$y_t = (a^*)^{-1}(B)\epsilon_t = \sum_{j \geq 0} \bar{b}_j^* \epsilon_{t-j}$$

und die  $\epsilon_t$ 's sind die Innovationen von  $(y_t)$  mit einer Varianz-Kovarianzmatrix  $E\epsilon_t \epsilon_t' = \Sigma^*$ . Das Mehrgleichungsmodell wird nun als sogenanntes *strukturelles VAR Modell (SVAR)* geschrieben

$$a_0(y_t - a_1^* y_{t-1} - \dots - a_p^* y_{t-p}) = b_0 \eta_t \quad 8.15$$

wobei  $b_0 \eta_t = u_t = a_0 \epsilon_t$ . Man setzt voraus, dass  $\text{Var}(\eta_t) = I_n$ , d.h. man interpretiert  $\eta_t$  als Vektor von *orthogonalisierten* Schocks. Die Strukturparameter des Mehrgleichungsmodells  $(a_0, a_1, \dots, a_p)$  und  $\Sigma = \text{Var}(u_t)$  und die Parameter  $(a_0, a_1^*, \dots, a_p^*, b_0)$  dieses SVAR Modells sind durch folgende Gleichungen verknüpft:

$$\begin{aligned} a_0 &= a_0 \\ a_0 a_i^* &= a_i \text{ for } i = 1, \dots, p \\ b_0 b_0' &= \Sigma = a_0 \Sigma^* a_0' \end{aligned}$$

Natürlich sind die Parameter des SVAR Modells ohne weitere Restriktionen nicht identifiziert. Ohne zusätzlichen Annahmen kann  $a_0$  eine beliebige nicht singuläre Matrix sein und  $b_0$  muss nur die Bedingung

$$b_0 b_0' = a_0 \Sigma^* a_0'$$

erfüllen. (D.h.  $b_0$  und  $O b_0$ , wobei  $O$  eine orthogonal Matrix ist, sind äquivalent zueinander.) Man benötigt also zumindest  $n^2 + n^2 - n(n+1)/2$  Restriktionen, um Identifizierbarkeit zu erhalten. Mit anderen Worten, wir benötigen ein eindeutiges Verfahren, um die orthogonalisierte Schocks  $\eta_t = b_0^{-1} \epsilon_t$  zu konstruieren. Siehe auch die Diskussion über die orthogonalisierte Impulsantwort (und die Zerlegung der Prognosefehler-Varianz) in Abschnitt 6.6.

Die Parameter  $(a_1^*, \dots, a_p^*)$  und  $\Sigma^*$  der reduzierten Form sind identifiziert.

**A-MODELL:** Hier verlangt man, dass  $b_0$  eine Diagonalmatrix (mit positiven Diagonal Elementen) ist. Weiters sollen die Diagonalelemente von  $a_0$  gleich Eins sein. Um Identifizierbarkeit zu erreichen, benötigt man also noch  $n(n-1)/2$  zusätzliche Bedingungen. Diese Bedingungen sind so zu wählen, dass

$$a_0 \Sigma^* a_0' = \Sigma = b_0 b_0' = \text{diag}(b_{0,11}^2, \dots, b_{0,nn}^2)$$

eine eindeutige Lösung hat. Verlangt man z.B., dass (nach einer geeigneten Permutation der Variablen  $y_{it}$ ) die Matrix  $a_0$  eine untere Dreiecksmatrix ist, dann ist das SVAR Modell identifiziert. Dieser Fall entspricht dem oben diskutierten Fall eines rekursiven Systems.

**B-MODELL:** Hier verlangen wir, dass  $a_0 = I_n$  und daher sind noch mindestens  $n(n-1)/2$  zusätzliche Restriktionen für  $b_0$  notwendig. Diese Restriktionen sind so zu wählen, dass

$$a_0 \Sigma^* a_0' = \Sigma^* = b_0 b_0'$$

eine eindeutige Lösung hat. Beispielsweise könnte man verlangen, dass  $b_0$  eine Dreiecksmatrix ist, d.h. man bestimmt  $b_0$  durch die Cholesky Zerlegung von  $\Sigma^*$ .

**AB-MODELL:** Hier werden Restriktionen für  $a_0$  und  $b_0$  benutzt.

Der Prozess  $y_t$  kann dargestellt werden als

$$y_t = \sum_{j \geq 0} \bar{b}_j^o \eta_{t-j}$$

wobei  $\bar{b}_j^o = \bar{b}_j^* b_0$  und insbesondere

$$\bar{b}_0^o = a_0^{-1} b_0$$

Diese Darstellung zeigt, dass Restriktionen für  $a_0$  und  $b_0$  insbesondere Auswirkungen auf den unmittelbaren Effekt der Schocks auf die Variablen  $y_t$  hat. Daher nennt man SVAR Modelle mit diesen Restriktionen (A-Modell, B-Modell und AB-Modell) auch oft "short-run" SVAR Modelle.

Eine weitere Möglichkeit Identifizierbarkeit zu erhalten sind Restriktionen für "long-run" Effekte:

**"LONG-RUN"-BEDINGUNGEN:** Ein Maß für den "long-run" Effekt der Schocks ist die Matrix

$$c_0 := \sum_{j \geq 0} \bar{b}_j^o = (a_0(I_n - a_1^* - \dots - a_p^*))^{-1} b_0 \quad \boxed{8.16}$$

Das  $ij$ -te Element dieser Matrix ist ein Maß für den kumulativen Effekt eines Schocks in der  $j$ -ten Gleichung auf die  $i$ -te Variable  $y_{it}$ , das heißt

$$c_{0,ij} = \frac{\partial (\sum_{s=t}^{\infty} y_{is})}{\partial \eta_{jt}}$$

A-priori Restriktionen an die Matrix  $c_0$  implizieren also Bedingungen für den "long-run" Effekt der Schocks. Verlangt man z.B.  $c_{0,ij} = 0$ , dann heißt das, dass der  $j$ -te Schock nur einen "transienten" Einfluss auf die  $i$ -te Variable hat.

Natürlich kann man auch Restriktionen an die Matrizen  $a_0$ ,  $b_0$  und  $c_0$  mischen.

### *Impuls Antwort*

Das (identifizierte) SVAR Modell liefert eine eindeutige Konstruktion für orthogonalisierte Schocks and daher auch für eine entsprechende Impulsantwort ( $\bar{b}_{k,ij}^o = \frac{\partial y_{it+k}}{\partial \eta_{jt}} \mid k \geq 0$ ). Wenn das Modell exogene Variable ( $x_{1t}, \dots, x_{st}$ ) enthält, kann man auch die entsprechende Impulsantwort der endogenen Variablen bezüglich dieser exogenen Variablen betrachten, d.h. ( $\partial y_{it+k} / \partial x_{jt} \mid k \geq 0$ ).

### *Schätzen und Testen*

Wenn das SVAR Modell durch  $n^2 + n^2 - n(n+1)/2$  Restriktionen identifizierbar gemacht wird (just-identified case), dann können SVAR Parameter Schätzer einfach aus (konsistenten) Schätzern der Parameter der reduzierten Form berechnet werden. Falls es mehr Restriktionen gibt (over-identified case), dann kann man z.B. maximum Likelihood Schätzer verwenden.

Da diese Schätzer (unter Standard Annahmen) asymptotisch normal verteilt sind, kann man auch die typischen Test Verfahren (z.B. Wald Test) verwenden, um lineare Hypothesen an die Parameter zu testen. Insbesondere kann man auch testen, ob die “over-identifying” Restriktionen gültig sind.

Oft ist die Frage, ob eine Variable “vorbestimmt” oder sogar exogen ist, nicht so einfach mit a priori Überlegungen zu beantworten. In diesem Fall kann man z.B. einen Hausmann Test verwenden, um zu überprüfen, ob eine Variable exogen ist oder nicht.

## 9

### *Literaturverzeichnis*

P.J. Brockwell and R.A. Davis. *Time Series: Theory and Methods*. Springer series in statistics. Springer-Verlag, New York, 2nd edition, 1991.

Manfred Deistler and Wolfgang Scherrer. *Modelle der Zeitreihenanalyse*. Birkhäuser, 2018.

J.D. Hamilton. *Time Series Analysis*. Princeton, NJ: Princeton University Press, 1994.

E. J. Hannan and M. Deistler. *The Statistical Theory of Linear Systems*. John Wiley & Sons, New York, 1988.

Fumio Hayashi. *Econometrics*. Princeton Univ. Press, 2000.

Helmut Lütkepohl. *Palgrave Handbook of Econometrics, Volume 1 Econometric Theory*, chapter Vector Autoregressive Models, pages 477–510. Palgrave MacMillan, 2007.





# 10

## Index

- AR Prozess, 63
  - Autokovarianzfunktion, 71
  - Prognose, 70
  - spektrale Dichte, 72
- AR System, 63
  - Ordnungsschätzung, 83
  - Schätzung, 75
- ARMA Prozess, 89
  - Prognose, 94
  - spektrale Dichte, 92
- ARMA System, 89
  - ML Schätzung, 95
  - Ordnungsschätzung, 97
- Autokorrelationsfunktion, 7
  - partielle, 84
- Autokovarianzfunktion, 6, 7
  - Schätzung der, 14
- Auto-Spektrum, 45
- Backshift Operator, 51
- Filter
  - BIBO, 52
  - gain, 60
  - Hintereinanderausführung, 54
  - Inverse eines, 55
  - $I_1$ , 52
  - Transferfunktion, 58
  - $z$ -Transformation, 53
- Fourier-Transformation
  - zeitdiskrete, 33
- Frequenzbereich eines stat. Prozesses, 43
- Granger Kausalität, 72
- Hilbertraum, 19
- Impulsantwort, 74
- Informationskriterien
  - AR Ordnung, 84
  - ARMA Ordnung, 97
- Informationskriterium
  - AIC, 84, 97
  - AIC korrigiertes, 97
  - BIC, 84, 97
  - HQ, 84
- Innovationen, 26
- Kohärenz, 46
- Konvergenz im quadratischen Mittel, 8
- Kreuzkovarianzfunktion, 8
- Kreuz-Spektrum, 45
- Kronecker Produkt, 87
- MA Prozess, 9
  - Autokovarianzfunktion, 10
  - spektrale Dichte, 48
- MA( $\infty$ ) Prozess
  - spektrale Dichte, 48
- MA( $\infty$ ) Prozess, 10
  - Autokovarianzfunktion, 10
- Martingaldifferenz, 17
- Mehrgleichungsmodelle, 99
  - reduzierte Form, 101
  - strukturelle Form, 100
- Mittelwertfunktion, 5
- MSE, 19
- Parseval, 34
- Periodogramm, 35
- positiv semidefinite Folge, 8
- Produktfilter, 54
- Prognose
  - einfache exponentielle Glättung, 30
  - endliche Vergangenheit, 22
  - mittels Wold-Darstellung, 29
  - MSE optimale, 19
  - naive, 29
  - unendliche Vergangenheit, 25
- Prognosegleichungen, 23
- Prognosehorizont, 22
- Projektion
  - orthogonale, 20
- Projektionssatz, 20
- Prozess
  - ergodisch, 17
  - harmonischer, 36
  - IID, 9
  - integriert, 11
  - mit orthogonalen Inkrementen, 38
  - Mittelwert-ergodisch, 12
  - regulärer, 26
  - schwach stationär, 7
  - singulärer, 26
  - stochastischer, 5
  - strikt stationär, 7
  - white noise, 9
- random walk, 11
- Riesz-Fisher Theorem, 9
- Schätzung
  - der Autokovarianzfunktion, 14
  - des Erwartungswertes, 12
- Spektraldarstellung
  - der Autokovarianzfunktion, 43
  - stationärer Prozesse, 42
- spektrale Dichte, 43
  - rational, 93
- spektrale Verteilungsfunktion, 43
- Spektral-Prozess, 42
- stochastisches Integral, 38, 40
- Strukturelles VAR Modell (SVAR), 108
- Test
  - Box-Pierce, 15
  - Breusch-Godfrey, 87
  - Ljung-Box, 15
  - Portmanteau, 86
- unimodular, 57
- weißes Rauschen, 9

spektrale Dichte, 47  
Wold Zerlegung, 26

Yule-Walker  
Gleichungen, 71  
Schätzer, 75

Zeitbereich eines stat. Prozesses, 21  
Zeitreihe, 5