



universität
wien

DIPLOMARBEIT / DIPLOMA THESIS

Titel der Diplomarbeit / Title of the Diploma Thesis

„Kausale ARMA-Prozesse und ihr Einsatz
in der Zeitreihenanalyse“

verfasst von / submitted by

Patrick Pröll

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of
Magister der Naturwissenschaft (Mag. rer. nat.)

Wien, 2016 / Vienna, 2016

Studienkennzahl lt. Studienblatt /
degree programme code as it appears on
the student record sheet:

A 190 313 406

Studienrichtung lt. Studienblatt /
degree programme as it appears on
the student record sheet:

Lehramtsstudium
UF Geschichte, Sozialkunde, Polit.Bildg.
UF Mathematik UniStG

Betreut von / Supervisor:

ao. Univ.-Prof. Dr. Karl Auinger

Danksagung

An dieser Stelle möchte ich einigen Personen meinen Dank aussprechen.

Meinem Betreuer, Prof. Dr. Karl Auinger, danke ich für seine Unterstützung, seine hilfreichen Ratschläge, ganz besonders aber für das große Vertrauen, das er mir stets entgegenbrachte und die umfassenden Freiheiten, die er mir sowohl bei der Themenwahl als auch der anschließenden Behandlung des Themas ließ.

Besonderer Dank gilt meiner Familie, insbesondere meinen Eltern, die mich stets uneingeschränkt unterstützten und ohne die ein Studienabschluss wohl kaum möglich gewesen wäre.

Nicht zuletzt gilt mein Dank auch meinen langjährigen Mitbewohnern Peter und Matthias, die meine manchmal auch studienbedingten Launen stets stoisch ertrugen, und auch meinen guten Freunden André und Marlene, die mir besonders gegen Ende meines Studiums stets zur Seite standen und die durch Korrekturlesen einen wichtigen Beitrag zur Fertigstellung dieser Arbeit leisteten.

Eidesstattliche Erklärung

Ich erkläre hiermit an Eides statt, dass ich die vorliegende Arbeit selbstständig und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sind als solche kenntlich gemacht. Die Arbeit wurde bisher in gleicher oder ähnlicher Form keiner anderen Prüfungsbehörde vorgelegt und auch noch nicht veröffentlicht.

Wien,
April 2016

.....

(Patrick Pröll)

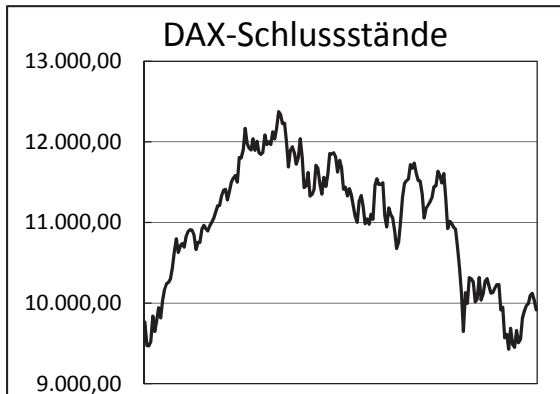
Inhaltsverzeichnis

Vorwort	1
1. Kurze Wiederholung wichtiger mathematischer Konzepte	2
1.1 Der Raum der quadratintegrierbaren, reellen Zufallsvariablen	2
1.2 Konvergenzbegriffe der Stochastik	3
1.3 Der Projektionssatz.....	5
1.4 Schätzfunktionen.....	6
2. Zeitreihen, Stationarität und Autokovarianz.....	7
3. Weißes Rauschen, MA(q)- und AR(p)-Zeitreihen	13
4. Das ARMA-Modell, Kausalität und Invertierbarkeit	21
5. Die Autokorrelationsfunktion kausaler ARMA-Zeitreihen	31
6. Das Komponentenmodell.....	34
6.1 Bestimmung der Trendkomponente	36
6.2 Bestimmung der Saisonalen Komponente	38
7. Die empirische Autokorrelationsfunktion	41
8. Parameterschätzung in ARMA-Modellen.....	45
9. Modellierung mittels ARMA-Zeitreihen	60
9.1 Überprüfen auf einfache MA(q) oder AR(p)-Zeitreihen	60
9.2 Das Akaike-Informationen-Kriterium.....	64
9.3 Überprüfen des Modells.....	66
9.4 Erstellen von Zukunftsprognosen.....	67
10. Anwendungsbeispiele	68
Literaturverzeichnis.....	81
Anhang	82
Zusammenfassung.....	82
Abstract	82

Vorwort

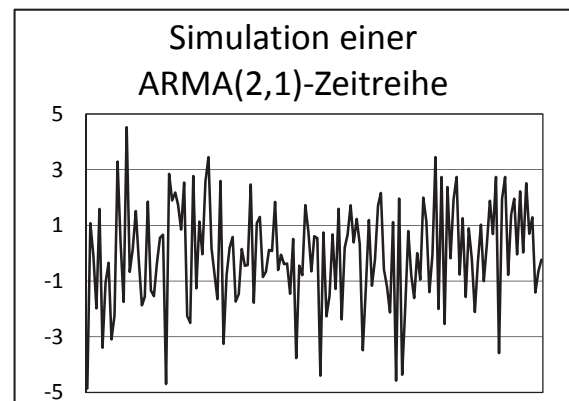
Zeitreihen sind, im Gegensatz zu vielen anderen Beschäftigungsfeldern der Mathematik, etwas, worunter sich auch NichtmathematikerInnen meist etwas vorstellen können. Es handelt sich dabei, salopp gesagt, um in festen zeitlichen Abständen gemessene Daten, die meist in Form eines Diagramms betrachtet werden, wie beispielsweise die täglich aufgezeichnete Entwicklung von Börsenkursen oder die jährlich geschätzte menschliche Gesamtpopulation der Erde.

In beiden genannten Fällen besteht ein Interesse daran, anhand der vorhandenen Daten möglichst verlässliche Zukunftsprognosen zu erstellen, wozu es nötig ist ein mathematisches Modell für die



vorliegende Zeitreihe zu entwickeln. Die Zeitreihenanalyse gehört beispielsweise zu den wichtigsten Werkzeugen der Ökonometrie, findet aber durchaus auch in anderen Wissenschaften wie der Meteorologie, der Geographie und sogar der Medizin Verwendung. Im Diagramm links ist beispielsweise die Zeitreihe der Schlussstände des Deutschen Aktienindex zwischen 14.11.2014 und 02.09.2015 abgebildet, die in Kapitel 10 noch ausgiebig analysiert werden wird.

Obwohl verschiedenste Modellklassen, wie zum Beispiel *ARCH*-, *GARCH*-, oder *ARIMAX*-Prozesse, im Themenfeld der Zeitreihenanalyse Verwendung finden, wird sich diese Arbeit auf eine eingehende Beschäftigung mit sowohl Autoregressiven Zeitreihen als auch mit Moving-Average Zeitreihen beschränken, die schließlich verallgemeinert werden zu dem aus beiden Modellklassen zusammengesetzten *ARMA*-Modell – dem in der angewandten Zeitreihenanalyse wohl wichtigstem Zeitreihentypus.



Dabei wird einerseits genauer auf wichtige Eigenschaften solcher *ARMA*-Prozesse, im Speziellen auf Stationarität, Kausalität und Invertierbarkeit eingegangen werden, aber auch ein genauerer Blick auf wichtige Kenngrößen von Zeitreihen wie die Autokorrelationsfunktion geworfen werden. Erklärtes Ziel dieser Arbeit ist es dem/der LeserIn sowohl grundlegende theoretische Aspekte dieser *ARMA*-Zeitreihen, als auch deren praktische Anwendung bis hin zur Erstellung von Zukunftsprognosen näherzubringen.

Auf dem Weg dorthin musste selbstverständlich ein gewisser Mut zur Lücke aufgebracht werden, weshalb, neben der bereits erwähnten Beschränkung auf eben diese *ARMA*-Prozesse, beispielsweise auch auf eine Behandlung komplexwertiger Zeitreihen und auch auf die Beschäftigung mit dem alternativen Zugang zur Zeitreihenanalyse mittels der Spektralanalyse vollständig verzichtet wurde.

1. Kurze Wiederholung wichtiger mathematischer Konzepte

In diesem Kapitel sollen zum besseren Verständnis der folgenden zeitreihenspezifischen Aussagen und vor allem derer Beweise einige für diese bedeutende Ergebnisse der Stochastik, aber auch anderer mathematischer Teildisziplinen, präsentiert bzw. wiederholt werden. Da diese an und für sich als bekannt vorausgesetzt werden könnten wird hierbei aus Gründen der Sparsamkeit auf Beweise verzichtet werden. Sollte Interesse an diesen bestehen, so seien beispielsweise die Bücher David Meintrup, Stefan Schöffler, Stochastik – Theorie und Anwendungen (Springer, 2005), Peter J. Brockwell, Richard A. Davis, Time Series: Theory and Methods (Springer, 1991) und Karl Bosch, Elementare Einführung in die angewandte Statistik (Vieweg+Teubner, 2010) empfohlen. Der Vollständigkeit halber sei noch erwähnt, dass des Öfteren nur Spezialfälle bekannter Sätze oder Konzepte angeführt sein werden, da dieses Kapitel selbstverständlich spezifisch für diese Arbeit geschrieben wurde.

1.1 Der Raum der quadratintegrierbaren, reellen Zufallsvariablen

Spricht man von Zeitreihen, so spricht man, wie noch zu sehen sein wird, von Familien von Zufallsvariablen. Deshalb ist es sinnvoll, sämtliche Aussagen auf dem Raum $L^2_{\mathbb{R}}(\Omega, \mathcal{A}, \mathbb{P})$, dem Raum der reellen quadratintegrierbaren Zufallsvariablen zu formulieren. Bekanntlich ist dieser Raum definiert als der Raum aller Zufallsvariablen $X: \Omega \rightarrow \mathbb{R}$, für die gilt, dass $\mathbb{E}(X^2) = \int X^2 d\mathbb{P} < \infty$. Bekanntermaßen ist der Raum $L^2_{\mathbb{R}}(\Omega, \mathcal{A}, \mathbb{P})$ ein *reeller Hilbertraum*.

Für solch einen *reellen Hilbertraum* H gilt:

- (i) Es ist für alle Funktionen f, g und h aus H und Koeffizienten a und b aus $\mathbb{R}$ ein Skalarprodukt $\langle f, g \rangle := \int f \cdot g d\mathbb{P}$ und eine Norm $\|f\| = \sqrt{\langle f, f \rangle}$ definiert, für die folgende Aussagen gelten:
 - $\langle f, g \rangle = \langle g, f \rangle$
 - $\langle af + bg, h \rangle = a\langle f, h \rangle + b\langle g, h \rangle$
 - $\langle f, f \rangle \geq 0$ und $\langle f, f \rangle = 0 \Leftrightarrow f = 0$
 - $|\langle f, g \rangle| \leq \|f\| \cdot \|g\|$.
- (ii) Er ist vollständig, d.h. alle Cauchyfolgen in H konvergieren bezüglich der Norm gegen einen Grenzwert, der wiederum in H liegt. (Siehe Def. 1.2.2)

Nachdem es sich beim $L^2_{\mathbb{R}}(\Omega, \mathcal{A}, \mathbb{P})$ um einen Hilbertraum über einem Wahrscheinlichkeitsraum handelt und bekanntermaßen der *Erwartungswert* einer Zufallsvariable X definiert ist durch $\mathbb{E}(X) := \int X d\mathbb{P}$, lassen sich mit dem Skalarprodukt zusammenhängende Ausdrücke durch den Erwartungswert folgendermaßen ausdrücken:

- (iii) $\|X\|^2 = \langle X, X \rangle = \mathbb{E}(X^2)$
- (iv) $\langle X, 1 \rangle = \langle 1, X \rangle = \mathbb{E}(X)$
- (v) $\langle X, Y \rangle = \mathbb{E}(X \cdot Y)$
- (vi) $\text{Var}(X) = \mathbb{E}[(X - \mathbb{E}(X))^2] = \|X - \mathbb{E}(X)\|^2 = \langle X - \mathbb{E}(X), X - \mathbb{E}(X) \rangle$
- (vii) $\text{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}(X)) \cdot (Y - \mathbb{E}(Y))] = \langle X - \mathbb{E}(X), Y - \mathbb{E}(Y) \rangle$
- (viii) $X \in L^2_{\mathbb{R}}(\Omega, \mathcal{A}, \mathbb{P}) \Leftrightarrow \int X^2 d\mathbb{P} = \mathbb{E}(X^2) < \infty$.

Erwähnenswert ist außerdem, dass, weil jedes nach den Kolmogorow-Axiomen definierte Wahrscheinlichkeitsmaß $\mathbb{P}$ ein endliches Maß darstellt, für $1 \leq p \leq q$ der Raum $L^q_{\mathbb{R}}(\Omega, \mathcal{A}, \mathbb{P})$ in den Raum $L^p_{\mathbb{R}}(\Omega, \mathcal{A}, \mathbb{P})$ eingebettet ist, dass also gilt, dass $L^q \subseteq L^p$. Daraus folgt im Speziellen, dass eine Zufallsvariable X aus $L^2_{\mathbb{R}}(\Omega, \mathcal{A}, \mathbb{P})$ stets auch Element des Raums $L^1_{\mathbb{R}}(\Omega, \mathcal{A}, \mathbb{P})$ ist, dass also gilt, dass $\mathbb{E}(X^2) = \int X^2 d\mathbb{P} < \infty \Rightarrow \mathbb{E}(|X|) = \int |X| d\mathbb{P} < \infty$.

Im Folgenden wird mit der verkürzten Schreibweise " L^2 " stets der Raum der reellen, quadratintegrierbaren Zufallsvariablen $L^2_{\mathbb{R}}(\Omega, \mathcal{A}, \mathbb{P})$ gemeint sein – auf komplexwertige Zufallsvariablen – und somit komplexwertige Zeitreihen – wird in dieser Arbeit nicht eingegangen werden.

Es ist noch der Vollständigkeit halber anzumerken, dass es sich bei den Elementen des $L^2_{\mathbb{R}}(\Omega, \mathcal{A}, \mathbb{P})$ genau genommen nicht um Zufallsvariablen, sondern um Äquivalenzklassen von Zufallsvariablen handelt, in denen jeweils alle Zufallsvariablen X mit $\mathbb{E}(X^2) < \infty$, die fast sicher übereinstimmen, zusammengefasst werden.

Hierbei bedeutet „fast sichere“ Übereinstimmung, dass die Zufallsvariablen sich nur über Elementarereignissen, deren Wahrscheinlichkeit 0 ist, unterschiedlich verhalten dürfen – „fast sicher“ ist einfach das wahrscheinlichkeitstheoretische Äquivalent zum aus der Maßtheorie bekannten „fast überall“.

Beispielsweise sind auf stetigen Wahrscheinlichkeitsräumen über den reellen Zahlen die Zufallsvariablen $X(\omega) := \omega$ und $Y(\omega) := \begin{cases} 1 & \text{für } \omega \in \{n \cdot \pi \mid n \in \mathbb{Z}\} \\ \omega & \text{sonst} \end{cases}$ fast sicher äquivalent, da hier einzelnen Punkten wie den Vielfachen von π die Wahrscheinlichkeit 0 zugeordnet ist.

Nachdem diese Unterscheidung zwischen Zufallsvariablen und deren Äquivalenzklassen aber für die Zwecke der Zeitreihenanalyse keine praktischen Implikationen hat, wird dies, der besseren Lesbarkeit geschuldet, in dieser Diplomarbeit geflissentlich ignoriert werden, wie es auch in der gesamten dem Autor bekannten Literatur zu diesem Thema der Fall ist, und einfach von Zufallsvariablen aus L^2 gesprochen werden.

1.2 Konvergenzbegriffe der Stochastik

In der Stochastik existieren zahlreiche Konvergenzbegriffe, von denen hier die zwei bedeutendsten, zusammen mit drei wichtigen Ergebnissen der Integrationstheorie, vorgestellt werden sollen.

Definition 1.2.1 Konvergenz in L^2 - Konvergenz im quadratischen Mittel

Eine Folge von Zufallsvariablen $(X_n)_{n \in \mathbb{N}}$, deren Folgeglieder Elemente des Raums L^2 sind, heißt *konvergent in L^2* bzw. *konvergent im quadratischen Mittel*, wenn eine Zufallsvariable X im selben Raum L^2 existiert, sodass gilt, dass

$$\lim_{n \rightarrow \infty} \|X_n - X\|^2 = 0 \text{ bzw. äquivalent: } \lim_{n \rightarrow \infty} \mathbb{E}[(X_n - X)^2] = 0.$$

Man schreibt auch kurz $X_n \xrightarrow{L^2} X$.

Definition 1.2.2 Cauchyfolge in L^2

Eine Folge von Zufallsvariablen $(X_n)_{n \in \mathbb{N}}$, deren Folgenglieder Elemente des Raums L^2 sind, heißt *Cauchyfolge in L^2* , falls gilt:

$$\forall \varepsilon > 0 \quad \exists n_0 \in \mathbb{N}, \text{ sodass } \forall m, n \geq n_0: \|X_m - X_n\| < \varepsilon \quad \text{bzw. äquivalent: } \sqrt{\mathbb{E}[(X_m - X_n)^2]} < \varepsilon.$$

Anmerkung 1.2.1

Da der L^2 -Raum vollständig ist, genügt es für eine Folge von L^2 -Zufallsvariablen also zu zeigen, dass es sich dabei um eine *Cauchyfolge in L^2* handelt, um ihre Konvergenz in L^2 zu beweisen.

Definition 1.2.3 Fast sichere Konvergenz

Eine Folge von Zufallsvariablen $(X_n)_{n \in \mathbb{N}}$ auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$ heißt *fast sicher konvergent*, wenn eine Zufallsvariable X auf demselben Wahrscheinlichkeitsraum existiert, sodass gilt:

$$\mathbb{P}(\lim_{n \rightarrow \infty} X_n = X) = \mathbb{P}\left\{\omega \in \Omega \mid \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)\right\} = 1.$$

Man schreibt dafür auch kurz $X_n \xrightarrow{f.s.} X$.

Anmerkung 1.2.2

Betrachtet man den Raum $L^2_{\mathbb{R}}(\Omega, \mathcal{A}, \mathbb{P})$, so bedeutet dies anders formuliert, dass eine solche Folge von Zufallsvariablen genau dann fast sicher konvergiert, wenn sie gegen eine Zufallsvariable X konvergiert, deren Werte für alle Elementarereignisse, deren Wahrscheinlichkeit größer als Null ist, reell sind und somit im offenen Intervall $(-\infty, \infty)$ liegen.

Man kann nun für diesen Raum aus $\mathbb{E}\left(\lim_{n \rightarrow \infty} |X_n|\right) = \int \lim_{n \rightarrow \infty} |X_n| d\mathbb{P} < \infty$ folgern, dass $\lim_{n \rightarrow \infty} |X_n|$ auf allen Elementarereignissen, deren Wahrscheinlichkeit größer als Null ist, endliche Werte annimmt, woraus folgt, dass $(|X_n|)_{n \in \mathbb{N}}$ *fast sicher konvergiert*.

Möchte man auf einem beliebigen Wahrscheinlichkeitsraum hingegen zeigen, dass $(X_n)_{n \in \mathbb{N}}$ fast sicher gegen eine vorgegebene Zufallsvariable X konvergiert, so genügt es trivialerweise zu beweisen, dass $\mathbb{E}\left(\lim_{n \rightarrow \infty} X_n - X\right) = \int \lim_{n \rightarrow \infty} (X_n - X) d\mathbb{P} = 0$ gilt.

Anmerkung 1.2.3

Fast sichere Konvergenz ist nicht äquivalent zu Konvergenz im quadratischen Mittel. Beispielsweise sei $([0,1], \mathcal{B}[0,1], \mathbb{P})$ ein Wahrscheinlichkeitsraum, wobei $\mathbb{P}$ eine (stetige) Gleichverteilung auf $[0,1]$ sei.

Sei $X_n(\omega) := \begin{cases} \sqrt{n}, & \omega \in [0, n^{-1}] \\ 0, & \text{sonst} \end{cases}$, so folgt: $\mathbb{E}\left(\lim_{n \rightarrow \infty} X_n - 0\right) = \int \lim_{n \rightarrow \infty} X_n d\mathbb{P} = \int 0 d\mathbb{P} = 0$,

d.h. X_n konvergiert fast sicher gegen 0. Allerdings gilt auch, dass

$$\lim_{n \rightarrow \infty} \mathbb{E}[(X_n - 0)^2] = \lim_{n \rightarrow \infty} \left((\sqrt{n})^2 \cdot \mathbb{P}(\omega \in [0, n^{-1}]) + 0 \cdot \mathbb{P}(\omega \in (n^{-1}, 1]) \right) = \lim_{n \rightarrow \infty} 1 = 1 \neq 0 ,$$

d.h. X_n konvergiert nicht im quadratischen Mittel gegen 0.

Die folgenden drei wichtigen Sätze machen (unter anderem) Aussagen darüber, unter welchen Bedingungen es möglich ist Erwartungswerte in unendliche Reihen „hineinzuziehen“ – ein Vorgang, der für viele Beweise unerlässlich sein wird.

Satz 1.2.1 Satz von der monotonen Konvergenz

Für eine monoton wachsende Folge von nichtnegativen Zufallsvariablen $(X_n)_{n \in \mathbb{N}}$ auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$ gilt:

$$\lim_{n \rightarrow \infty} \int X_n d\mathbb{P} = \int \lim_{n \rightarrow \infty} X_n d\mathbb{P} \text{ bzw. } \lim_{n \rightarrow \infty} \mathbb{E}(X_n) = \mathbb{E}(\lim_{n \rightarrow \infty} X_n) .$$

Lemma 1.2.1 Lemma von Fatou

Für eine Folge von nichtnegativen Zufallsvariablen $(X_n)_{n \in \mathbb{N}}$ auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$ gilt:

$$\liminf_{n \rightarrow \infty} \int X_n d\mathbb{P} \geq \int \liminf_{n \rightarrow \infty} X_n d\mathbb{P} \text{ bzw. } \liminf_{n \rightarrow \infty} \mathbb{E}(X_n) \geq \mathbb{E} \left(\liminf_{n \rightarrow \infty} X_n \right) .$$

Satz 1.2.2 Satz von der dominierten Konvergenz für L^2 -Zufallsvariablen

Sei $(X_n)_{n \in \mathbb{N}}$ eine Folge von L^2 -Zufallsvariablen. Es gilt: $\lim_{n \rightarrow \infty} \mathbb{E}(X_n) = \mathbb{E}(\lim_{n \rightarrow \infty} X_n)$, wenn folgende Bedingungen erfüllt sind:

- (i) $(X_n)_{n \in \mathbb{N}}$ konvergiert fast sicher oder in L^2 gegen eine Zufallsvariable X
- (ii) $\exists Y \in L^2: \forall n \in \mathbb{N}: |X_n| \leq Y$ fast sicher.

1.3 Der Projektionssatz

Dieser wichtige Satz der Funktionalanalysis wird später bei der Berechnung von partiellen Autokorrelationen und der Erstellung von Zukunftsprognosen eine große Bedeutung haben.

Definition 1.3.1 Orthogonales Komplement

Das *orthogonale Komplement* einer Teilmenge M eines Hilbertraumes H ist jene Teilmenge $M^\perp$ von H , für die gilt:

$$x \in M^\perp \Leftrightarrow \langle x, y \rangle = 0 \quad \forall y \in M .$$

Satz 1.3.1 Projektionssatz

Sei M ein abgeschlossener Unterraum eines Hilbertraumes H und sei $x \in H$, dann gilt

$$(i) \exists! \hat{x} \in M \text{ mit } \|x - \hat{x}\| = \inf_{y \in M} \|x - y\|$$

$$(ii) \hat{x} \in M \text{ und } \|x - \hat{x}\| = \inf_{y \in M} \|x - y\| \Leftrightarrow \hat{x} \in M \text{ und } (x - \hat{x}) \in M^\perp.$$

Die Abbildung $P_M : H \rightarrow M$ mit $P_M(x) = \hat{x}$ heißt Orthogonalprojektion von H auf M .

1.4 Schätzfunktionen

Im Folgenden werden einige wichtige Eigenschaften von Schätzfunktionen vorgestellt, die in jenen Teilen dieser Arbeit, die sich mit statistischen Aspekten der Zeitreihenanalyse beschäftigen, von Bedeutung sein werden.

Definition 1.4.1 Erwartungstreue

Eine Schätzfunktion S für einen Parameter a heißt *erwartungstreu*, wenn gilt:

$$\mathbb{E}(S) = a \quad .$$

Eine Folge von Schätzfunktionen $(S_n)_{n \in \mathbb{N}}$ für einen Parameter a heißt *asymptotisch erwartungstreu*, wenn gilt:

$$\lim_{n \rightarrow \infty} \mathbb{E}(S_n) = a \quad .$$

Definition 1.4.2 Konsistenz

Eine Folge von Schätzfunktionen $(S_n)_{n \in \mathbb{N}}$ für einen Parameter a heißt *konsistent*, wenn gilt:

$$\lim_{n \rightarrow \infty} \mathbb{P}(|S_n - a| > \varepsilon) = 0 \quad \forall \varepsilon > 0 \quad .$$

Satz 1.4.1

Seien für alle n aus $\mathbb{N}$ die Schätzfunktionen S_n für den Parameter a erwartungstreu, dann ist $(S_n)_{n \in \mathbb{N}}$ konsistent, wenn gilt:

$$\lim_{n \rightarrow \infty} \text{Var}(S_n) = \lim_{n \rightarrow \infty} \mathbb{E}[(S_n - a)^2] = 0 \quad \forall n \in \mathbb{N} \quad .$$

2. Zeitreihen, Stationarität und Autokovarianz

In diesem Kapitel soll der Begriff „Zeitreihe“ zunächst einmal klar definiert werden und der wichtige Begriff der Stationarität vorgestellt werden. Darauf aufbauend sollen die linearen Abhängigkeitsstrukturen innerhalb stationärer Zeitreihen genauer untersucht werden, wobei die Begriffe Autokovarianz, Autokorrelation und partielle Autokorrelation definiert werden. Den Abschluss dieses Kapitels bilden zwei Sätze, die zur Beweisführung in folgenden Kapiteln enorme Bedeutung haben werden.

Definition 2.1 Stochastischer Prozess

Ein *stochastischer Prozess* ist eine Familie von auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$ definierten Zufallsvariablen $(X_t : t \in T)$ mit beliebiger Indexmenge T .

Definition 2.2 Zeitreihe, Zentrierte Zeitreihe

Ein stochastischer Prozess $(X_t : t \in T)$ mit diskreter Indexmenge T heißt *Zeitreihe*. Im Speziellen nennt man also die stochastischen Prozesse $X = (X_t : t \in \mathbb{Z})$ und $Y = (Y_t : t \in \mathbb{N})$ *Zeitreihen*. Sei $Z = (X_t : t \in T)$ eine Zeitreihe und gelte für alle t , dass $\mathbb{E}(Z_t) = 0$, so nennt man Z *zentriert*.

Definition 2.3 Strenge Stationarität

Eine Zeitreihe $X := (X_t : t \in \mathbb{Z})$ heißt *streng stationär*, wenn gilt, dass die Zufallsvektoren $(X_1, \dots, X_k)$ und $(X_{1+h}, \dots, X_{k+h})$ für alle k und h aus $\mathbb{Z}$ die selbe (multivariate) Verteilung besitzen.

Definition 2.4 (Schwache) Stationarität

Eine reellwertige Zeitreihe $X := (X_t : t \in \mathbb{Z})$ heißt *schwach stationäre L^2 -Zeitreihe*, oder auch nur *stationäre Zeitreihe*, wenn gilt:

- i) $\mathbb{E}(X_t^2) < \infty \quad \forall t \in \mathbb{Z} \quad \text{d.h.: } X_t \in L^2 \quad \forall t \in \mathbb{Z}$
- ii) $\exists \mu \in \mathbb{R}, \text{ sodass } \mathbb{E}(X_t) = \mu \quad \forall t \in \mathbb{Z}$
- iii) $\text{Cov}(X_{t_1}, X_{t_1+h}) = \text{Cov}(X_{t_2}, X_{t_2+h}) \quad \forall t_1, t_2, h \in \mathbb{Z} .$

Die Funktion $\gamma : \mathbb{Z} \rightarrow \mathbb{R}$ mit $\gamma(h) = \text{Cov}(X_t, X_{t+h}) = \text{Cov}(X_0, X_h)$ heißt *Autokovarianzfunktion*, die Zahl h bezeichnet man als *Lag*.

Für $\gamma(0) > 0$ heißt die Funktion $\rho(h) = \frac{\gamma(h)}{\gamma(0)}$ *Autokorrelationsfunktion*.

Korollar 2.1

Für eine stationäre Zeitreihe X gilt, dass:

- (i) $Var(X_t) = \gamma(0) \quad \forall t \in \mathbb{Z}$
- (ii) $\exists c \in \mathbb{R}$, sodass $\mathbb{E}(X_t^2) = c \quad \forall t \in \mathbb{Z}$
- (iii) $\gamma(h) = \gamma(-h) \quad \forall h \in \mathbb{Z}$
- (iv) $\rho(h) = \rho(-h) \quad \forall h \in \mathbb{Z}$

Ist X zentriert, so gilt außerdem:

- (v) $\gamma(h) = \mathbb{E}(X_t \cdot X_{t+h}) \quad \forall t, h \in \mathbb{Z}$
- (vi) $\gamma(0) = \mathbb{E}(X_t^2) \quad \forall t \in \mathbb{Z}$

Beweis:

- (i) $Var(X_t) = Cov(X_t, X_t) = Cov(X_t, X_{t+0}) = \gamma(0)$
- (ii) $Var(X_t) = \mathbb{E}[(X_t - \mu)^2] = \mathbb{E}(X_t^2) - \mu^2 \Rightarrow \mathbb{E}(X_t^2) = Var(X_t) + \mu^2 = c$
- (iii) $\gamma(h) = Cov(X_t, X_{t+h}) = \mathbb{E}[(X_t - \mu)(X_{t+h} - \mu)] = \mathbb{E}[(X_{t+h} - \mu)(X_t - \mu)] =$
 $= \mathbb{E}[(X_t - \mu)(X_{t-h} - \mu)] = Cov(X_t, X_{t-h}) = \gamma(-h)$
- (iv) $\rho(h) = \frac{\gamma(h)}{\gamma(0)} = \frac{\gamma(-h)}{\gamma(0)} = \rho(-h)$
- (v) $\gamma(h) = Cov(X_t, X_{t+h}) = \mathbb{E}[(X_t - \mu) \cdot (X_{t+h} - \mu)] = \mathbb{E}(X_t \cdot X_{t+h})$
- (vi) $\mathbb{E}(X_t^2) \stackrel{(ii)}{=} Var(X_t) + \mu^2 \stackrel{(i)}{=} \gamma(h) + \mu^2 = \gamma(h)$

■

Anmerkung 2.1

Stationarität ist wohl die „wichtigste“ Eigenschaft, die eine Zeitreihe besitzen kann. Diese Bedeutung lässt sich folgendermaßen erklären:

Salopp ausgedrückt bedeutet Stationarität, dass die linearen Abhängigkeitsstrukturen zwischen den einzelnen Zufallsvariablen einer stationären Zeitreihe in der „Zukunft“ dieselben sein werden wie diejenigen in „Gegenwart“ und „Vergangenheit“.

Ohne Stationarität könnte man niemals (sinnvolle) Schlüsse aus der „Vergangenheit“ auf das „zukünftige“ Verhalten einer Zeitreihe ziehen, was in der Praxis die Analyse von Zeitreihen, speziell wenn das Ziel dabei das Erstellen von Zukunftsprognosen darstellt, als wenig sinnvoll erscheinen ließe.

Obleich strenge Stationarität bedeuten würde, dass sämtliche Abhängigkeitsstrukturen zwischen den Zufallsvariablen zeitinvariant sind, genügt es im Allgemeinen von (schwacher) Stationarität auszugehen, da das Hauptinstrument der Zeitreihenanalyse die Analyse der Autokorrelationsfunktion darstellt, die ohnehin lediglich lineare Zusammenhänge beschreiben kann.

Definition 2.5 Partielle Autokorrelation

Sei $X = (X_t : t \in \mathbb{Z})$ eine stationäre und zentrierte Zeitreihe. Sei $h \geq 1$,
 $H := \text{span}\{X_t, X_{t+1}, \dots, X_{t+h-1}\}$ und P_H die Orthogonalprojektion von L^2 auf H , sodass

$$P_H(X_{t+h}) = d_{h,1}X_{t+h-1} + d_{h,2}X_{t+h-2} + \dots + d_{h,h}X_t,$$

dann nennt man $\pi(h) := d_{h,h}$ die *partielle Autokorrelation von X zum "Lag" h* .

Satz 2.1 Berechnung der partiellen Autokorrelation

Die partiellen Autokorrelationen $\pi(h)$ ergeben sich für $\gamma(0) \neq 0$ als Teile der Lösungen der folgenden linearen Gleichungssysteme, sofern diese eindeutig lösbar sind.

$$\begin{pmatrix} 1 & \rho(1) & \dots & \rho(h-1) \\ \rho(1) & 1 & & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \rho(h-1) & \dots & \dots & 1 \end{pmatrix} \cdot \begin{pmatrix} d_{h,1} \\ d_{h,2} \\ \vdots \\ \pi(h) \end{pmatrix} = \begin{pmatrix} \rho(1) \\ \rho(2) \\ \vdots \\ \rho(h) \end{pmatrix}$$

Beweis

Es gilt: $\langle P_H(X_{t+h}) - X_{t+h}, u \rangle = 0 \quad \forall u \in H$

$$\Rightarrow \langle P_H(X_{t+h}) - X_{t+h}, X_{t+k} \rangle = 0 \quad \forall k \in [0, h-1]$$

$$\Leftrightarrow \langle P_H(X_{t+h}), X_{t+k} \rangle - \langle X_{t+h}, X_{t+k} \rangle = 0 \quad \forall k \in [0, h-1]$$

$$\Leftrightarrow \mathbb{E}(P_H(X_{t+h}) \cdot X_{t+k}) = \mathbb{E}(X_{t+h} \cdot X_{t+k}) \quad \forall k \in [0, h-1]$$

$$\Leftrightarrow \mathbb{E}\left(\sum_{j=1}^h d_{h,j} \cdot X_{t+h-j} \cdot X_{t+k}\right) = \mathbb{E}(X_{t+h} \cdot X_{t+k}) \quad \forall k \in [0, h-1]$$

$$\Leftrightarrow \sum_{j=1}^h d_{h,j} \cdot \mathbb{E}(X_{t+h-j} \cdot X_{t+k}) = \mathbb{E}(X_{t+h} \cdot X_{t+k}) \quad \forall k \in [0, h-1]$$

$$\Leftrightarrow \sum_{j=1}^h d_{h,j} \cdot \gamma(h-j-k) = \gamma(h-k) \quad \forall k \in [0, h-1]$$

$$\Leftrightarrow \sum_{j=1}^h d_{h,j} \cdot \rho(h-j-k) = \rho(h-k) \quad \forall k \in [0, h-1]$$

■

Satz 2.2

Sei $X = (X_t : t \in \mathbb{Z})$ eine beliebige reellwertige Zeitreihe und $(a_n)_{n \in \mathbb{Z}}$ eine beliebige Folge reeller Zahlen, für die gilt, dass

$$(i) \quad \sup_{t \in \mathbb{Z}} \mathbb{E}(|X_t|) < \infty$$

$$(ii) \quad \sum_{j=-\infty}^{\infty} |a_j| < \infty,$$

dann konvergiert folgende Reihe für alle $t \in \mathbb{Z}$ fast sicher absolut:

$$\sum_{j=-\infty}^{\infty} a_j X_{t-j} := \sum_{j=0}^{\infty} a_j X_{t-j} + \sum_{j=1}^{\infty} a_{-j} X_{t+j}.$$

Handelt es sich bei X zusätzlich um eine L^2 -Zeitreihe (also $X_t \in L^2 \quad \forall t \in \mathbb{Z}$), so konvergiert die Reihe im quadratischen Mittel gegen denselben Grenzwert, gegen den sie fast sicher konvergiert.

Beweis

$$\underbrace{\mathbb{E} \left(\sum_{j=-\infty}^{\infty} |a_j| \cdot |X_{t-j}| \right)}_{\text{Satz v.d.mon.Konvergenz}} = \sum_{j=-\infty}^{\infty} |a_j| \cdot \mathbb{E}(|X_{t-j}|) \leq \underbrace{\sup_{t \in \mathbb{Z}} \mathbb{E}(|X_t|)}_{< \infty} \cdot \underbrace{\sum_{j=-\infty}^{\infty} |a_j|}_{< \infty} < \infty ,$$

d.h.: $\sum_{j=-\infty}^{\infty} a_j X_{t-j}$ konvergiert fast sicher absolut.

Für die Konvergenz im quadratischen Mittel genügt es nun zu zeigen, dass die Folge der Partialsummen der Reihe für alle $t \in \mathbb{Z}$ eine Cauchyfolge in L^2 darstellt.

Ist X eine L^2 -Zeitreihe, so folgt $\mathbb{E}(X_t^2) < \infty \quad \forall t \in \mathbb{Z}$.

Seien $m, n \in \mathbb{N}$ mit $0 < m < n$. Dann gilt, dass

$$\begin{aligned} \lim_{m, n \rightarrow \infty} \left\| \sum_{j=m}^n a_j X_{t-j} + \sum_{k=m}^n a_{-k} X_{t+k} \right\|^2 &= \lim_{m, n \rightarrow \infty} \mathbb{E} \left[\left(\sum_{j=m}^n a_j X_{t-j} + \sum_{k=m}^n a_{-k} X_{t+k} \right)^2 \right] = \\ &= \lim_{m, n \rightarrow \infty} \mathbb{E} \left[\left(\sum_{m < |j| < n} a_j X_{t-j} \right)^2 \right] = \lim_{m, n \rightarrow \infty} \mathbb{E} \left(\sum_{m < |j| < n} \sum_{m < |k| < n} a_j \cdot a_k \cdot X_{t-j} \cdot X_{t-k} \right) = \\ &= \lim_{m, n \rightarrow \infty} \sum_{m < |j| < n} \sum_{m < |k| < n} a_j \cdot a_k \cdot \mathbb{E}(X_{t-j} \cdot X_{t-k}) \leq \lim_{m, n \rightarrow \infty} \left[\underbrace{\sup_{t \in \mathbb{Z}} \mathbb{E}(X_t^2)}_{< \infty} \cdot \underbrace{\left(\sum_{m < |j| < n} a_j \right)^2}_{\rightarrow 0} \right] = 0 , \end{aligned}$$

wobei $\sum_{m < |j| < n} a_j$ gegen 0 konvergiert, weil nach Voraussetzung $\sum_{j=-\infty}^{\infty} |a_j|$ konvergiert, somit auch $\sum_{j=-\infty}^{\infty} a_j$ konvergiert und deshalb nach dem Cauchy Kriterium die Folge der Partialsummen dieser Reihe eine Cauchyfolge sein muss.

Das bedeutet nun, dass die Folge der Partialsummen der Reihe $\sum_{j=-\infty}^{\infty} a_j X_{t-j}$ eine Cauchyfolge in L^2 ist und somit diese Reihe im quadratischen Mittel konvergiert.

Nun muss noch die fast sichere Gleichheit der beiden Grenzwerte gezeigt werden.

Sei S der Grenzwert im L^2 -Sinne, dann gilt:

$$\mathbb{E} \left[\lim_{n \rightarrow \infty} \underbrace{\left(S - \sum_{j=-n}^n a_j X_{t-j} \right)^2}_{> 0 \quad \forall \omega \in \Omega : \mathbb{P}(\omega) > 0} \right] = \mathbb{E} \left[\liminf_{n \rightarrow \infty} \left(S - \sum_{j=-n}^n a_j X_{t-j} \right)^2 \right] \stackrel{\text{Fatou}}{\leq} \liminf_{n \rightarrow \infty} \mathbb{E} \left(S - \sum_{j=-n}^n a_j X_{t-j} \right)^2 .$$

Dabei kann man den Limes dem Limes Inferior gleichsetzen, weil man weiß, dass die Reihe $\sum_{j=-\infty}^{\infty} a_j X_{t-j}$ fast sicher konvergiert und somit auch $S - \sum_{j=-\infty}^{\infty} a_j X_{t-j}$ fast sicher konvergieren muss.

Die rechte Seite dieser Ungleichung ist nun, nachdem $\sum_{j=-n}^n a_j X_{t-j}$ im quadratischen Mittel für $n \rightarrow \infty$ gegen S konvergiert, gleich 0. Die linke Seite der Ungleichung kann klarerweise nicht kleiner als 0 sein, also muss sie genau 0 sein.

Also konvergiert $(S - \sum_{j=-n}^n a_j X_{t-j})^2$ fast sicher gegen 0, woraus wiederum folgen muss, dass $S - \sum_{j=-n}^n a_j X_{t-j}$ fast sicher gegen 0 und somit $\sum_{j=-n}^n a_j X_{t-j}$ fast sicher gegen S konvergiert, woraus die fast sichere Gleichheit der beiden Grenzwerte folgt. ■

Lemma 2.1

Sei $X = (X_t : t \in \mathbb{Z})$ eine L^2 -Zeitreihe und sei die Reihe $\sum_{j=-\infty}^{\infty} a_j X_{t-j}$ fast sicher absolut konvergent, so gilt

$$\mathbb{E} \left(\sum_{j=-\infty}^{\infty} a_j X_{t-j} \right) = \sum_{j=-\infty}^{\infty} a_j \mathbb{E}(X_{t-j}) .$$

Beweis:

Es gilt $\sum_{j=-\infty}^{\infty} a_j X_{t-j} = \sum_{j=0}^{\infty} a_j X_{t-j} + \sum_{j=1}^{\infty} a_{-j} X_{t+j} = \lim_{n \rightarrow \infty} \sum_{j=0}^n a_j X_{t-j} + \lim_{n \rightarrow \infty} \sum_{j=1}^n a_{-j} X_{t+j} .$

Nachdem $\sum_{j=-\infty}^{\infty} a_j X_{t-j}$ fast sicher absolut konvergiert müssen auch $\sum_{j=0}^{\infty} a_j X_{t-j}$ und $\sum_{j=1}^{\infty} a_{-j} X_{t+j}$ fast sicher absolut konvergieren und damit natürlich auch fast sicher konvergieren. Aus dieser fast sicheren absoluten Konvergenz, jeweils gegen die L^2 -Zufallsvariablen Y_1 und Y_2 , folgt

- $\forall n \in \mathbb{N}: |\sum_{j=0}^n a_j X_{t-j}| \leq \sum_{j=0}^n |a_j X_{t-j}| = Y_1$ fast sicher
- $\forall n \in \mathbb{N}: |\sum_{j=1}^n a_{-j} X_{t+j}| \leq \sum_{j=1}^n |a_{-j} X_{t+j}| = Y_2$ fast sicher.

Deshalb kann man den Satz von der dominierten Konvergenz für L^2 -Zufallsvariablen (Satz 1.2.2) anwenden und erhält

$$\begin{aligned} \mathbb{E}(\sum_{j=-\infty}^{\infty} a_j X_{t-j}) &= \mathbb{E} \left(\lim_{n \rightarrow \infty} \sum_{j=0}^n a_j X_{t-j} \right) + \mathbb{E} \left(\lim_{n \rightarrow \infty} \sum_{j=1}^n a_{-j} X_{t+j} \right) = \\ &= \lim_{n \rightarrow \infty} \sum_{j=0}^n a_j \mathbb{E}(X_{t-j}) + \lim_{n \rightarrow \infty} \sum_{j=1}^n a_{-j} \mathbb{E}(X_{t+j}) = \\ &= \sum_{j=0}^{\infty} a_j \mathbb{E}(X_{t-j}) + \sum_{j=-\infty}^{-1} a_j \mathbb{E}(X_{t-j}) = \sum_{j=-\infty}^{\infty} a_j \mathbb{E}(X_{t-j}) . \end{aligned}$$

Satz 2.3

Sei $X = (X_t : t \in \mathbb{Z})$ eine stationäre Zeitreihe mit Autokovarianzfunktion γ . Sei $(a_j)_{j \in \mathbb{Z}}$ eine Folge reeller Koeffizienten, für die gilt, dass $\sum_{j=-\infty}^{\infty} |a_j| < \infty$, dann gilt für alle t aus $\mathbb{Z}$ folgendes:

- (i) Die Reihe $\sum_{j=-\infty}^{\infty} a_j X_{t-j}$ konvergiert fast sicher absolut.
- (ii) Die Reihe $\sum_{j=-\infty}^{\infty} a_j X_{t-j}$ konvergiert fast sicher und im quadratischen Mittel gegen denselben Grenzwert Y_t .
- (iii) Die so erhaltene Zeitreihe $Y := (Y_t : t \in \mathbb{Z})$ ist stationär.
- (iv) Ist X zentriert, so ist auch Y zentriert.
- (v) Die Zeitreihe Y besitzt die Autokovarianzfunktion

$$\gamma_Y(h) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} a_j a_k \gamma(h - k + j) .$$

Beweis:

Aufgrund der Definition von Stationarität ist X eine L^2 -Zeitreihe, somit auch eine L^1 -Zeitreihe und es gilt somit für alle t , dass $\mathbb{E}(|X_t|) < \infty$ und somit trivialerweise auch $\sup_{t \in \mathbb{Z}} \mathbb{E}(|X_t|) < \infty$.

Somit sind die Bedingungen von Satz 2.2 erfüllt, woraus einerseits für $\sum_{j=-\infty}^{\infty} a_j X_{t-j}$ für alle t die fast sichere und die L^2 -Konvergenz gegen den jeweils selben Grenzwert Y_t aus L^2 folgt, andererseits aber auch die fast sichere absolute Konvergenz der Reihe hervorgeht. Somit ist (i) und (ii) bewiesen.

Für die Zeitreihe $Y = (Y_t : t \in \mathbb{Z})$ gilt nun

$$\mathbb{E}(Y_t) = \mathbb{E}\left(\sum_{j=-\infty}^{\infty} a_j X_{t-j}\right) \stackrel{\text{Lemma 2.1}}{=} \sum_{j=-\infty}^{\infty} a_j \mathbb{E}(X_{t-j}) = \underbrace{\mathbb{E}(X_t)}_{=\mu \forall t \in \mathbb{Z}} \cdot \underbrace{\sum_{j=-\infty}^{\infty} a_j}_{< \infty} = c \in \mathbb{R} \quad \forall t \in \mathbb{Z}.$$

Wenn X zentriert ist, so gilt offensichtlich für alle t aus $\mathbb{Z}$, dass $\mathbb{E}(X_t) = 0$ und somit auch $\mathbb{E}(Y_t) = 0$, womit (iv) bewiesen ist (unter der Voraussetzung, dass (iii) stimmt, wurde doch Zentriertheit nur für stationäre Zeitreihen definiert). Weiters gilt

$$\begin{aligned} \mathbb{E}(Y_t \cdot Y_{t+h}) &= \mathbb{E}\left[\left(\sum_{j=-\infty}^{\infty} a_j X_{t-j}\right)\left(\sum_{k=-\infty}^{\infty} a_k X_{t-k+h}\right)\right] = \mathbb{E}\left(\sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} a_j a_k X_{t-j} X_{t-k+h}\right) = \\ &\stackrel{\text{Lemma 2.1}}{=} \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} a_j a_k \underbrace{\mathbb{E}(X_{t-j} \cdot X_{t-k+h})}_{\mathbb{E}(XY) = \text{Cov}(X,Y) + \mathbb{E}(X)\mathbb{E}(Y)} = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} a_j a_k [\gamma(h-k+j) + \mathbb{E}(X_t)^2]. \end{aligned}$$

Anzumerken ist, dass man Lemma 2.1 hier anwenden darf, weil bekanntlich das Produkt zweier absolut konvergenter Reihen wiederum absolut konvergent sein muss. Insgesamt gilt nun also

$$\begin{aligned} \text{Cov}(Y_{t+h}, Y_t) &= \mathbb{E}(Y_t \cdot Y_{t+h}) - \mathbb{E}(Y_t) \cdot \mathbb{E}(Y_{t+h}) = \\ &= \left(\sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} a_j a_k [\gamma(h-k+j) + \mathbb{E}(X_t)^2]\right) - \mathbb{E}(X_t)^2 \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} a_j a_k = \\ &= \underbrace{\sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} a_j a_k \gamma(h-k+j)}_{\text{unabhängig von } t}. \end{aligned}$$

Somit sind nun einerseits alle Kriterien von Stationarität nachgewiesen worden und damit (iii) bewiesen, andererseits hat man auch die gewünschte Autokovarianzfunktion erhalten, womit auch die Korrektheit von (v) nachgewiesen wurde. ■

Anmerkung 2.2

Auch wenn die Aussagen der vorangegangenen Sätze auf den ersten Blick recht „trocken“ wirken mögen, so sind doch Satz 2.3, der ja auf Satz 2.2 aufbaut, und auch Lemma 2.1 von fundamentaler Bedeutung für die Beweisführungen der nächsten Kapitel.

3. Weißes Rauschen, $MA(q)$ - und $AR(p)$ -Zeitreihen

In diesem Kapitel werden erstmals konkrete Arten von Zeitreihen behandelt, allen voran das elementar wichtige Modell des Weißen Rauschens, aber auch *Moving Average*- und *Autoregressive*-Zeitreihen. Dabei wird auf den Spezialfall der $AR(1)$ -Zeitreihen näher eingegangen werden um einige Problemstellungen des Kapitels 4 bereits im Vorfeld etwas näher zu beleuchten.

Definition 3.1 Weißes Rauschen

Eine reellwertige Zeitreihe $e = (e_t : t \in \mathbb{Z})$ heißt *Weißes Rauschen*, wenn sie folgende Bedingungen erfüllt:

- (i) $\mathbb{E}(e_t) = 0 \quad \forall t \in \mathbb{Z}$
- (ii) $0 < \text{Var}(e_t) := \sigma_e^2 < \infty \quad \forall t \in \mathbb{Z}$
- (iii) $\underbrace{\mathbb{E}(e_{t_1} \cdot e_{t_2}) = 0}_{d.h. \text{unkorreliert}} \quad \forall t_1 \neq t_2; \quad t_1, t_2 \in \mathbb{Z} \quad .$

Korollar 3.1

Ein Weißes Rauschen ist immer stationär und zentriert.

Beweis

$$\forall t \in \mathbb{Z} : \mathbb{E}(e_t) = 0$$

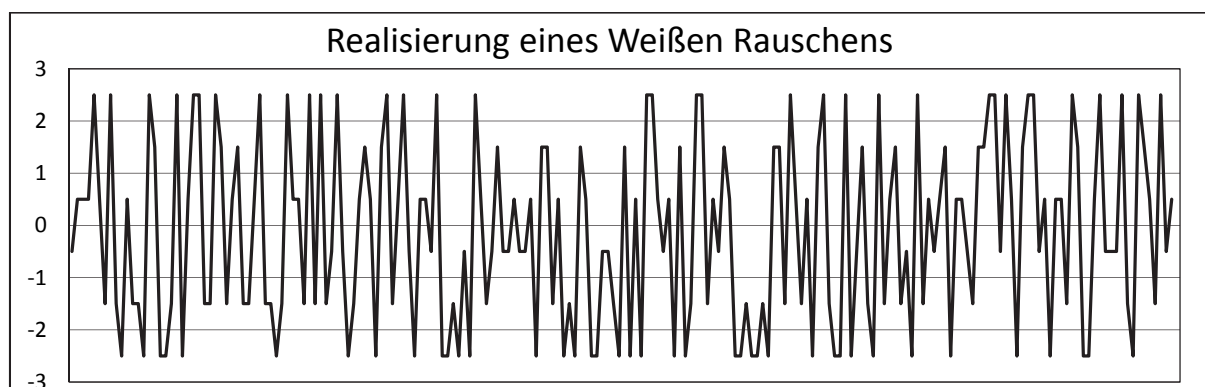
$$\forall t \in \mathbb{Z} : \mathbb{E}(e_t^2) = \text{Var}(e_t) < \infty$$

$$\text{Cov}(e_t, e_{t+h}) = \begin{cases} \sigma_e^2, & h = 0 \\ 0, & \text{sonst} \end{cases} \quad \Rightarrow \text{Cov ist unabhängig von } t$$

■

Beispiel 3.1

Sei e jene Zeitreihe, die entsteht, wenn man immer wieder einen gewöhnlichen 6-seitigen Würfel wirft und jedes Ergebnis abzüglich des Erwartungswertes (= 3,5) notiert. Es handelt sich bei e um ein – in diesem Fall nicht nur unkorreliertes, sondern sogar stochastisch unabhängiges – Weißes Rauschen, dessen Graph, wie man sehen kann, keine Abhängigkeitsstrukturen erkennen lässt.



Definition 3.2 Moving-Average-Zeitreihen

Sei $e = (e_t : t \in \mathbb{Z})$ ein Weißes Rauschen. Eine Zeitreihe $X = (X_t : t \in \mathbb{Z})$ heißt *Moving Average Zeitreihe der Ordnung q* oder kurz *MA(q)-Zeitreihe*, wenn sie für alle t aus $\mathbb{Z}$ die folgende Gleichung erfüllt:

$$X_t = e_t + b_1 e_{t-1} + \cdots + b_q e_{t-q} \quad b_1, \dots, b_q \in \mathbb{R}; \quad q \in \mathbb{N}; \quad b_q \neq 0 .$$

Satz 3.1

Eine MA(q)-Zeitreihe ist immer stationär und zentriert.

Beweis:

Sei $b_0 := 1$ und σ_e^2 die Varianz des zugrundeliegenden Weißen Rauschens, dann gilt für alle t aus $\mathbb{Z}$, dass $\mathbb{E}(X_t) = \mathbb{E}(e_t + b_1 e_{t-1} + \cdots + b_q e_{t-q}) = \underbrace{\mathbb{E}(e_t)}_{=0} + \underbrace{b_1 \mathbb{E}(e_{t-1})}_{=0} + \cdots + \underbrace{b_q \mathbb{E}(e_{t-q})}_{=0} = 0$.

Außerdem gilt Folgendes:

$$\begin{aligned} \text{Cov}(X_t, X_{t+h}) &= \mathbb{E}(X_t \cdot X_{t+h}) = \mathbb{E} \left[\left(\sum_{j=0}^q b_j \cdot e_{t-j} \right) \cdot \left(\sum_{k=0}^q b_k \cdot e_{t-k+h} \right) \right] = \\ &= \mathbb{E} \left(\sum_{j=0}^q \sum_{k=0}^q b_j \cdot b_k \cdot e_{t-j} \cdot e_{t-k+h} \right) = \sum_{j=0}^q \sum_{k=0}^q b_j \cdot b_k \cdot \underbrace{\mathbb{E}(e_{t-j} \cdot e_{t-k+h})}_{\substack{\sigma_e^2, & j=k-h \\ 0, & \text{sonst}}} = \\ &= \begin{cases} \sigma_e^2 \cdot \sum_{j=0}^{q-h} b_j \cdot b_{j+h}, & h \leq q \\ 0, & \text{sonst} \end{cases} \Rightarrow \text{Cov unabhängig von } t . \end{aligned}$$

Insbesondere folgt, dass $\mathbb{E}(X_t^2) = \sigma_e^2 \cdot \sum_{j=0}^q b_j^2 < \infty$.

■

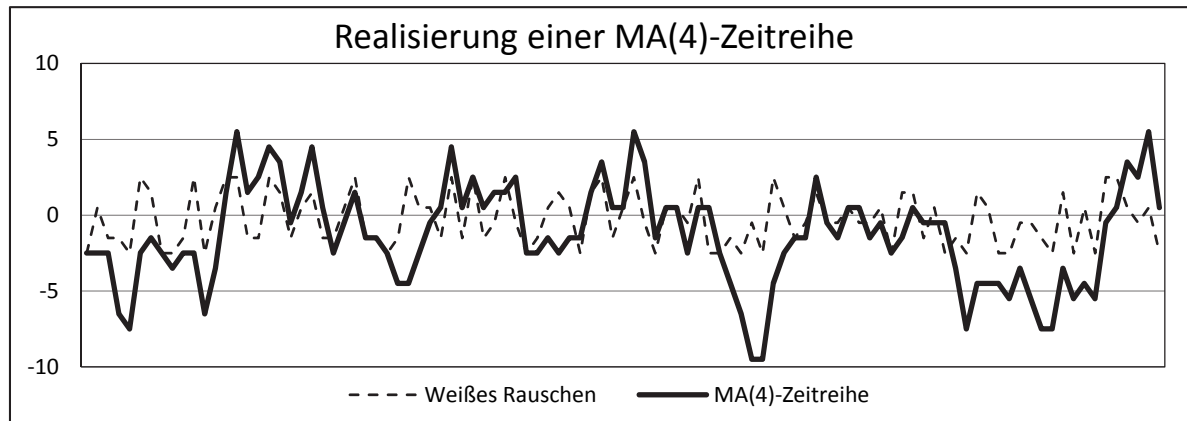
Anmerkung 3.1

Man hätte den vorigen Satz selbstverständlich auch wesentlich kürzer mittels Satz 2.3 beweisen können. Das Motiv an dieser direkteren Beweisführung ist jenes, dem/der LeserIn nochmals die Kriterien für Stationarität vor Augen zu führen. Außerdem hat man durch diesen Beweis sogleich die Gleichung der Autokovarianzfunktion von MA(q)-Zeitreihen erhalten, deren wesentliches Merkmal ist, dass sie für Lags größer als q nur noch den Wert 0 annimmt.

Während die einzelnen Zufallsvariablen des elementaren Zeitreihenmodells des weißen Rauschens untereinander unkorreliert sind, ist dies offensichtlich bei den *Moving Average Zeitreihen* nicht der Fall. Hier haben zufällige Ereignisse kurz- bzw. mittelfristige Nachwirkungen auf zukünftige Werte.

Beispiel 3.2

Man betrachte das in Beispiel 3.1 beschriebene Weiße Rauschen. Anstatt nun direkt die erwürfelten Werte abzüglich des Erwartungswertes zu notieren, addiert man nun jeweils die vorhergegangenen 4 Werte zum aktuellen Würfelergebnis – jeweils abzüglich 3,5 – hinzu. Man erhält so die $MA(4)$ -Zeitreihe X , gegeben durch $X_t = e_t + e_{t-1} + e_{t-2} + e_{t-3} + e_{t-4}$.



Definition 3.3 Autoregressive Zeitreihen

Sei $e = (e_t : t \in \mathbb{Z})$ ein Weißes Rauschen. Eine Zeitreihe $X = (X_t : t \in \mathbb{Z})$ heißt *Autoregressive Zeitreihe der Ordnung p* oder kurz $AR(p)$ -Zeitreihe, wenn sie für alle t aus $\mathbb{Z}$ die folgende Gleichung erfüllt:

$$X_t = e_t + a_1 X_{t-1} + \dots + a_p X_{t-p} \quad a_1, \dots, a_p \in \mathbb{R}; a_p \neq 0$$

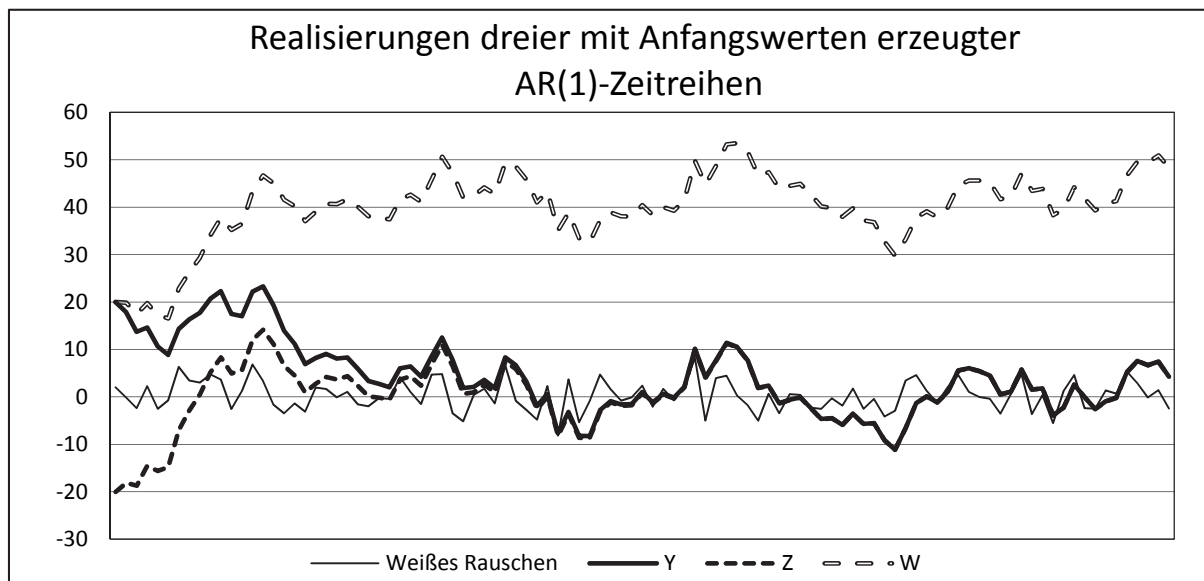
Anmerkung 3.2

Zu jedem gegebenen Weißen Rauschen kann man durch willkürliches Festlegen von „Anfangswerten“ $X_0, \dots, X_p$, seien sie deterministisch oder auch mit dem Weißen Rauschen unkorrelierte Zufallsvariablen, problemlos verschiedene $AR(p)$ -Zeitreihen erzeugen. Im Allgemeinen sind diese – im Gegensatz zu $MA(q)$ -Zeitreihen- jedoch nicht stationär.

Setzt man beispielsweise für eine $AR(1)$ -Zeitreihe, gegeben durch die Gleichung $X_t = 0,8X_{t-1} + e_t$ mit $N(0,3^2)$ -verteilterm Weißem Rauschen e , einfach für eine diese Gleichung erfüllende Zeitreihe Y den deterministischen Anfangswert $Y_0 := 20$ und für eine andere, dieselbe Gleichung erfüllende Zeitreihe Z den Anfangswert $Z_0 := -20$, so ergeben sich zwei offensichtlich sehr verschiedene $AR(1)$ -Zeitreihen, die nach Definition nicht stationär sind, wie man am Diagramm am Ende dieser Anmerkung sehen wird.

Man sieht, dass zu Beginn der Zeitreihen eben durch das Setzen der Anfangswerte die Stationarität gestört wird, sich die so generierten Zeitreihen aber immer mehr annähern. Das lässt den Verdacht aufkommen, dass wohl zu jedem Weißen Rauschen und zu jeder $AR(1)$ -Gleichung eine eindeutig bestimmte stationäre Zeitreihe existiert, was später noch bewiesen werden wird. Betrachtet man nun die Gleichung $X_t = X_{t-1} + e_t$ mit demselben Weißen Rauschen und setzt für die diese

erfüllende $AR(1)$ -Zeitreihe W den Anfangswert $W_0 = 20$ fest, so erhält man eine Zeitreihe, die ganz offensichtlich nicht stationär beginnt und sich auch nicht an eine stationäre Zeitreihe annähert.



Die folgenden beiden Sätze beschäftigen sich nun mit der Frage, für welche Koeffizienten a überhaupt stationäre $AR(1)$ -Zeitreihen existieren und ob diese stationären Lösungen eindeutig bestimmt sind.

Satz 3.2

Es existieren keine stationären $AR(1)$ -Zeitreihen, die die Gleichung $X_t = e_t + aX_{t-1}$ für $|a| = 1$ erfüllen.

Beweis:

1.Fall: $a = 1$

Seien $r \in \mathbb{Z}$ beliebig und o.B.d.A. $t \in \mathbb{Z}$ mit $t > r$, dann kann man X_t durch X_r folgendermaßen ausdrücken: $X_t = X_{t-1} + e_t = \dots = X_r + \sum_{j=r+1}^t e_j$. Angenommen X wäre stationär, dann gilt

$$\begin{aligned}
 \text{Var}(X_t) &= \mathbb{E}(X_t^2) - \mu^2 = \mathbb{E}\left[\left(X_r + \sum_{j=r+1}^t e_j\right)^2\right] - \mu^2 = \\
 &= \mathbb{E}(X_r^2) + 2\mathbb{E}\left(X_r \cdot \sum_{j=r+1}^t e_j\right) + \mathbb{E}\left[\left(\sum_{j=r+1}^t e_j\right)^2\right] - \mu^2 = \\
 &= \text{Var}(X_t) + \mu^2 + 2\mathbb{E}\left(X_r \cdot \sum_{j=r+1}^t e_j\right) + \sum_{j=r+1}^t \sum_{k=r+1}^t \underbrace{\mathbb{E}(e_j \cdot e_k)}_{\substack{\sigma_e^2, & j=k \\ 0, & \text{sonst}}} - \mu^2 = \\
 &= \text{Var}(X_t) + 2\mathbb{E}\left(X_r \cdot \sum_{j=r+1}^t e_j\right) + (t - r - 1) \cdot \sigma_e^2 = \\
 &= \text{Var}(X_t) + (t - r - 1) \cdot \sigma_e^2 + 2\mu \underbrace{\mathbb{E}\left(\sum_{j=r+1}^t e_j\right)}_{=0} + 2 \underbrace{\text{Cov}\left(X_r, \sum_{j=r+1}^t e_j\right)}_{=0} .
 \end{aligned}$$

Also muss gelten, dass $\sigma_e^2 = 0$, was nach Def. 3.1 nicht möglich ist und somit einen Widerspruch darstellt. Drückt man hingegen X_r durch X_t aus mittels $X_r = X_{r+1} - e_r = \dots = X_t - \sum_{j=r}^{t-1} e_j$ so stößt man auf denselben Widerspruch.

2.Fall: $a = -1$

Analog mittels $X_t = (-1)^{t-r} X_r + \sum_{j=r+1}^t (-1)^{t-j} e_j$.

■

Satz 3.3

Für gegebenes Weißes Rauschen $e = (e_t : t \in \mathbb{Z})$ und eine reelle Zahl a mit $0 \neq |a| \neq 1$, gibt es stets genau eine stationäre $AR(1)$ -Zeitreihe mit Koeffizient a . Diese ist immer zentriert.

Beweis:

1.Fall: $|a| < 1$

$$\begin{aligned} X_t &= e_t + aX_{t-1} = e_t + a(e_{t-1} + aX_{t-2}) = \dots = e_t + ae_{t-1} + \dots + a^q e_{t-q} + a^{q+1} X_{t-(q+1)} = \\ &= a^{q+1} X_{t-(q+1)} + \sum_{j=0}^q a^j e_{t-j} \Leftrightarrow X_t - \sum_{j=0}^q a^j e_{t-j} = a^{q+1} X_{t-(q+1)} \quad \forall t \in \mathbb{Z}, \forall q \in \mathbb{N} \end{aligned}$$

Aus Satz 2.3 weiß man nun, dass $\sum_{j=0}^{\infty} a^j e_{t-j}$ wegen $\sum_{j=0}^{\infty} |a^j| = \frac{1}{1-|a|} < \infty$ sowohl in L^2 als auch fast sicher gegen den selben Grenzwert konvergiert. Nun muss noch überprüft werden, ob es sich bei diesem, Stationarität von X vorausgesetzt, für alle t jeweils um X_t handelt.

Angenommen $X = (X_t : t \in \mathbb{Z})$ wäre stationär, so muss gelten, dass

$$\begin{aligned} \left\| X_t - \sum_{j=0}^q a^j e_{t-j} \right\|^2 &= \|a^{q+1} X_{t-(q+1)}\|^2 = \mathbb{E} \left[(a^{q+1} X_{t-(q+1)})^2 \right] = \mathbb{E} (a^{2q+2} X_{t-(q+1)}^2) = \\ &= a^{2q+2} \mathbb{E} (X_{t-(q+1)}^2) \stackrel{\text{stationär}}{=} a^{2q+2} \mathbb{E} (X_t^2). \end{aligned}$$

$$\text{Also gilt: } \lim_{q \rightarrow \infty} \left\| X_t - \sum_{j=0}^q a^j e_{t-j} \right\|^2 = \lim_{q \rightarrow \infty} \underbrace{a^{2q+2}}_{|a| < 1} \underbrace{\mathbb{E} (X_t^2)}_{< \infty} = 0 \Rightarrow \sum_{j=0}^{\infty} a^j e_{t-j} \xrightarrow{L^2, f.s.} X_t \quad \forall t \in \mathbb{Z}.$$

Es wurde also bewiesen: Wenn es eine stationäre $AR(1)$ -Zeitreihe für gegebenes Weißes Rauschen und $|a| < 1$ gibt, so ist diese eindeutig bestimmt durch $X_t = \sum_{j=0}^{\infty} a^j e_{t-j}$. Nun muss noch gezeigt werden, dass $\sum_{j=0}^{\infty} a^j e_{t-j}$ auch tatsächlich eine stationäre und zentrierte $AR(1)$ -Zeitreihe mit Koeffizient a ist. Es gilt:

$$\begin{aligned}
X_t - aX_{t-1} &= \lim_{q \rightarrow \infty} \left[\left(\sum_{j=0}^q a^j e_{t-j} \right) - a \left(\sum_{k=0}^{q-1} a^k e_{t-k-1} \right) \right] = \lim_{q \rightarrow \infty} \left[\left(\sum_{j=0}^q a^j e_{t-j} \right) - \left(\sum_{k=1}^q a^k e_{t-k} \right) \right] = \\
&= \lim_{q \rightarrow \infty} \left[e_t + \left(\sum_{j=1}^q a^j e_{t-j} \right) - \left(\sum_{k=1}^q a^k e_{t-k} \right) \right] = \lim_{q \rightarrow \infty} e_t = e_t,
\end{aligned}$$

d.h.: $X = (X_t : t \in \mathbb{Z})$ mit $X_t = \sum_{j=0}^{\infty} a^j e_{t-j}$ ist für $|a| < 1$ eine $AR(1)$ -Zeitreihe.

Nachdem das Weiße Rauschen e eine zentrierte, stationäre Zeitreihe ist, folgt nun direkt aus Satz 2.3, dass X eine stationäre und zentrierte Zeitreihe ist.

2.Fall: $|a| > 1$

$$\begin{aligned}
X_{t+1} &= aX_t + e_{t+1} \Leftrightarrow X_t = \frac{1}{a}(X_{t+1} - e_{t+1}) = -\frac{1}{a}e_{t+1} - \frac{1}{a^2}e_{t+2} - \dots - \frac{1}{a^q}e_{t+q} + \frac{1}{a^q}X_{t+q} = \\
&= \frac{1}{a^q}X_{t+q} - \sum_{j=1}^q \frac{1}{a^j}e_{t+j} \Leftrightarrow X_t - \left(-\sum_{j=1}^q \frac{1}{a^j}e_{t+j} \right) = \frac{1}{a^q} \cdot X_{t+q} \quad \forall t \in \mathbb{Z}, \forall q \in \mathbb{N}
\end{aligned}$$

Wiederum ist aus Satz 2.3 klar, dass $-\sum_{j=1}^{\infty} \frac{1}{a^j}e_{t+j}$ fast sicher und in L^2 gegen den selben Grenzwert konvergiert, und wieder soll untersucht werden ob es sich unter der Annahme von Stationarität hierbei für alle t aus $\mathbb{Z}$ um X_t handelt.

Angenommen $X = (X_t : t \in \mathbb{Z})$ wäre stationär, so muss gelten, dass

$$\begin{aligned}
\left\| X_t - \left(-\sum_{j=1}^q \frac{1}{a^j}e_{t+j} \right) \right\|^2 &= \left\| \frac{1}{a^q}X_{t+q} \right\|^2 = \mathbb{E} \left[\left(\frac{1}{a^q}X_{t+q} \right)^2 \right] = \mathbb{E} \left(\frac{1}{a^{2q}}X_{t+q}^2 \right) = \left(\frac{1}{a} \right)^{2q} \cdot \mathbb{E}(X_t^2) \\
\Rightarrow \lim_{q \rightarrow \infty} \left\| X_t - \left(-\sum_{j=1}^q \frac{1}{a^j}e_{t+j} \right) \right\|^2 &= \lim_{q \rightarrow \infty} \left(\frac{1}{a} \right)^{2q} \cdot \mathbb{E}(X_t^2) = 0 \Rightarrow -\sum_{j=1}^{\infty} \frac{1}{a^j}e_{t+j} \xrightarrow{L^2, f.s.} X_t.
\end{aligned}$$

Wenn es eine stationäre $AR(1)$ -Zeitreihe für gegebenes Weißes Rauschen und $|a| > 1$ gibt, so ist diese also eindeutig für alle t aus $\mathbb{Z}$ bestimmt durch $X_t = -\sum_{j=1}^{\infty} \frac{1}{a^j}e_{t+j}$. Nun muss wiederum gezeigt werden, dass $-\sum_{j=1}^{\infty} \frac{1}{a^j}e_{t+j}$ auch tatsächlich eine stationäre $AR(1)$ -Zeitreihe zum Koeffizienten a ist. Es gilt, dass

$$\begin{aligned}
X_t - aX_{t-1} &= \lim_{q \rightarrow \infty} \left[\left(-\sum_{j=1}^q \frac{1}{a^j}e_{t+j} \right) - a \cdot \left(-\sum_{j=1}^{q+1} \frac{1}{a^j}e_{t+j-1} \right) \right] = \\
&= \lim_{q \rightarrow \infty} \left[\left(-\sum_{j=1}^q \frac{1}{a^j}e_{t+j} \right) - \left(-\sum_{j=1}^{q+1} \frac{1}{a^{j-1}}e_{t+j-1} \right) \right] = \lim_{q \rightarrow \infty} \left[\left(-\sum_{j=1}^q \frac{1}{a^j}e_{t+j} \right) + \left(\sum_{j=0}^q \frac{1}{a^j}e_{t+j} \right) \right] = \\
&= \lim_{q \rightarrow \infty} \left[e_t + \sum_{j=1}^q \frac{1}{a^j}e_{t+j} - \sum_{j=1}^q \frac{1}{a^j}e_{t+j} \right] = \lim_{q \rightarrow \infty} e_t = e_t.
\end{aligned}$$

Nun muss noch die Stationarität und Zentriertheit bewiesen werden. Nachdem $-\sum_{j=1}^{\infty} \frac{1}{a^j} e_{t+j}$ fast sicher absolut konvergiert, darf die Summationsreihenfolge geändert werden. Es gilt nun

$$-\sum_{j=1}^{\infty} \frac{1}{a^j} e_{t+j} = -\sum_{j=-\infty}^{-1} \frac{1}{a^{-j}} e_{t-j} \quad \text{und} \quad \sum_{j=-\infty}^{-1} \left| \frac{1}{a^{-j}} \right| = \sum_{j=1}^{\infty} \left| \frac{1}{a} \right|^j < \infty.$$

Somit kann man nun Satz 2.3 anwenden, woraus Stationarität und Zentriertheit folgt. ■

Anmerkung 3.3

Für praktische Anwendungen sind $AR(1)$ -Zeitreihen mit $|a| > 1$ kaum brauchbar und werden deshalb meist nicht weiter beachtet. Der Grund dafür liegt darin, dass diese Zeitreihen, die ja gegeben sind durch die Gleichung $X_t = -\sum_{j=1}^{\infty} \frac{1}{a^j} e_{t+j}$, durch Ergebnisse in der „Zukunft“ beeinflusst sind, sie sind, wie später noch genauer definiert werden wird, nicht *kausal*.

Nochmals klarzustellen ist Folgendes: Die Gleichung $X_t = a \cdot X_{t-1} + e_t$ wird für jedes Weiße Rauschen und für alle reellen Koeffizienten von unendlich vielen Zeitreihen erfüllt (Man beachte Anmerkung 3.2). Erst durch die Forderung von Stationarität definiert diese Gleichung, $|a| \neq 1$ vorausgesetzt, eindeutig eine zugehörige Zeitreihe.

Die Existenz von stationären $AR(p)$ -Zeitreihen höherer Ordnung p soll im folgenden Kapitel als Sonderfall einer weiteren Klasse von Zeitreihen untersucht werden, den so genannten *ARMA*-Zeitreihen.

Satz 3.4

Sei $X = (X_t: t \in \mathbb{Z})$ die stationäre $AR(1)$ -Zeitreihe zum Weißen Rauschen e und zum Koeffizienten a mit Autokovarianzfunktion γ , Autokorrelationsfunktion ρ und partieller Autokorrelationsfunktion π , dann gilt:

- (i) Für $|a| < 1$ gilt $\gamma(h) = \sigma_e^2 \cdot \frac{a^h}{1-a^2}$ und $\rho(h) = a^h$
- (ii) Für $|a| > 1$ gilt $\gamma(h) = \sigma_e^2 \cdot \frac{a^{-h}}{a^2-1}$ und $\rho(h) = a^{-h}$

Beweis:

- (i) Aus dem Beweis von Satz 3.3 weiß man, dass sich die Zeitreihe durch $X_t = \sum_{j=0}^{\infty} a^j e_{t-j}$ ausdrücken lässt. Nach Satz 2.3 lässt sich nun die Autokovarianzfunktion berechnen:

$$\begin{aligned} \gamma(h) &= \lim_{n \rightarrow \infty} \left(\sum_{j=0}^n \sum_{k=0}^n a^j \cdot a^k \cdot \underbrace{\gamma_e(h-k+j)}_{\substack{\sigma_e^2 \text{ für } j=k-h \\ 0 \text{ sonst}}} \right) = \lim_{n \rightarrow \infty} \begin{cases} \sum_{j=0}^{n-h} a^j \cdot a^{j+h} \cdot \sigma_e^2, & h \leq n \\ 0, & \text{sonst} \end{cases} = \\ &= \sigma_e^2 \cdot \sum_{j=0}^{\infty} a^{2j} a^h = \sigma_e^2 \cdot a^h \cdot \sum_{j=0}^{\infty} a^{2j} = \sigma_e^2 \cdot \frac{a^h}{1-a^2}. \end{aligned}$$

Für die Autokorrelationsfunktion gilt nun, dass $\rho(h) = \frac{\gamma(h)}{\gamma(0)} = \frac{\sigma_e^2 \cdot \frac{a^h}{1-a^2}}{\sigma_e^2 \cdot \frac{a^0}{1-a^2}} = a^h$.

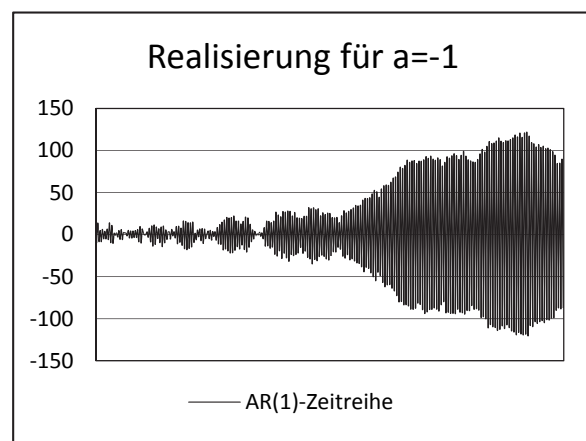
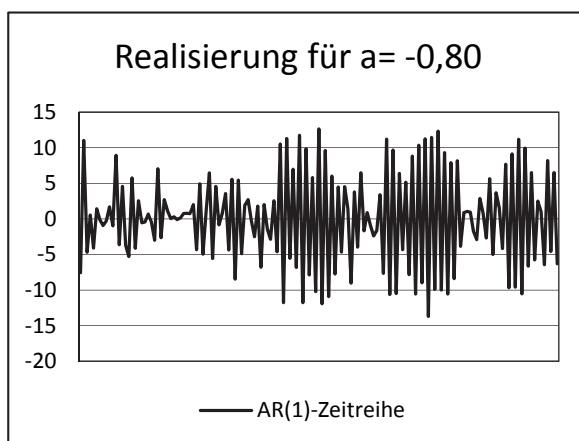
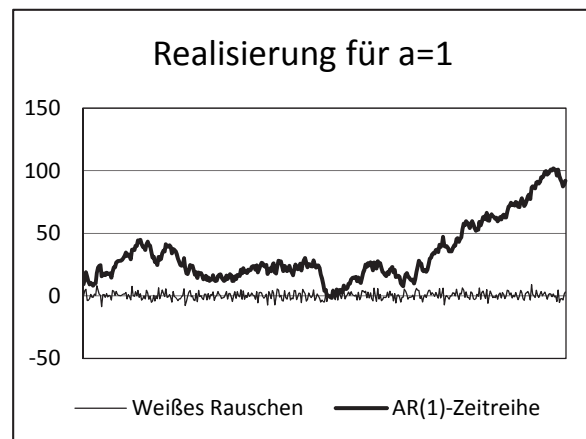
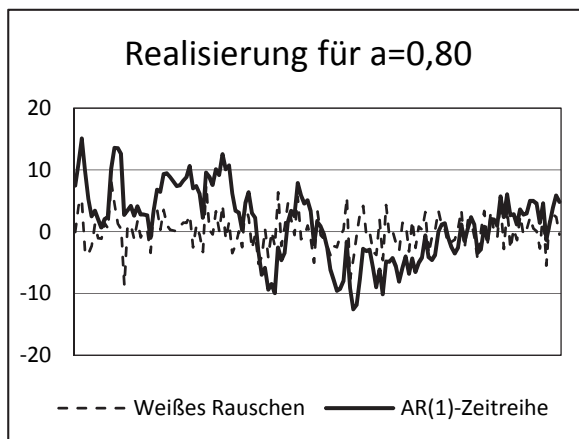
- (ii) Gemäß des Beweises von Satz 3.3 lässt sich die Zeitreihe durch $X_t = -\sum_{j=1}^{\infty} \frac{1}{a^j} e_{t+j}$ ausdrücken und so umformen, dass man Satz 2.3 anwenden kann. Der Rest des Beweises erfolgt wie für (i) durch einfaches Ausrechnen.

■

Beispiel 3.3

Zum besseren Verständnis der soeben so ausführlich behandelten $AR(1)$ -Zeitreihen seien hier Diagramme mehrerer Realisierungen solcher Zeitreihen über dem stets selben $N(0,3^2)$ -verteilten Weißen Rauschen mit wechselndem Koeffizienten a präsentiert. Man sieht recht deutlich den Einfluss des Koeffizienten auf den Verlauf des Graphen der zugehörigen Zeitreihe. Je näher a von unten an die 1 heranrückt, desto stärker weicht die Zeitreihe vom jeweiligen Wert des Weißen Rauschens ab, je näher a von oben an -1 heranrückt, desto stärker oszilliert sie.

Für die „nicht-stationären“ Koeffizienten $a = \pm 1$, hier für einen größeren Zeitraum geplottet, sieht man deutlich das Ergebnis von Satz 3.2: Der Erwartungswert ist nicht konstant bzw. steigt die Varianz mit fortlaufender Zeit, wodurch Stationarität in beiden Fällen nicht gegeben sein kann.



4. Das ARMA-Modell, Kausalität und Invertierbarkeit

In diesem Kapitel wird nun, wie bereits angekündigt, die vielleicht wichtigste Klasse von Zeitreihen in der angewandten Zeitreihenanalyse, die der so genannten *ARMA*-Zeitreihen, vorgestellt werden um einen soliden theoretischen Hintergrund für die späteren Kapitel, die von statistischen Verfahren handeln werden, zu geben. Zusätzlich zu diesen *ARMA*-Zeitreihen werden auch noch zwei weitere zentrale Begriffe der Zeitreihenanalyse vorgestellt und untersucht werden, nämlich einerseits *Kausalität*, ein Begriff, der bereits bei der Behandlung von *AR*(1)-Zeitreihen einen ersten Auftritt hatte, und andererseits *Invertierbarkeit*.

Definition 4.1 *ARMA*(p, q)-Zeitreihen

Sei $e = (e_t : t \in \mathbb{Z})$ ein Weißes Rauschen. Eine Zeitreihe $X = (X_t : t \in \mathbb{Z})$ heißt *Autoregressive Moving Average Zeitreihe der Ordnungen p und q* , oder kurz *ARMA*(p, q)-*Zeitreihe*, wenn sie für alle t aus $\mathbb{Z}$ folgende Gleichung, die auch *ARMA-Gleichung* genannt wird, erfüllt:

$$X_t = \left(\sum_{j=1}^p a_j X_{t-j} \right) + \left(\sum_{k=0}^q b_k e_{t-k} \right) \text{ mit } b_0 = 1; a_1, \dots, a_p, b_1, \dots, b_q \in \mathbb{R}; a_p, b_q \neq 0; p, q \in \mathbb{N}_0.$$

Anmerkung 4.1

Offensichtlich setzt sich eine *ARMA*(p, q)-Zeitreihe, wie schon aus dem Namen ersichtlich ist, aus einer *AR*(p)- und einer *MA*(q)-Zeitreihe zusammen. Eine *ARMA*(0, q)-Zeitreihe erfüllt klar ersichtlich die Gleichung $X_t = e_t + b_1 e_{t-1} + \dots + b_q e_{t-q}$ und ist somit eine *MA*(q)-Zeitreihe.

Gleichermaßen ist eine *ARMA*($p, 0$)-Zeitreihe, die ja $X_t = a_1 X_{t-1} + \dots + a_p X_{t-p} + e_t$ erfüllt, eine *AR*(p)-Zeitreihe und eine *ARMA*(0,0)-Zeitreihe $X_t = e_t$ trivialerweise einfach ein Weißes Rauschen. Insofern kann man insbesondere *AR*(p)- und *MA*(q)-Zeitreihen als Spezialfälle der Klasse der *ARMA*(p, q)-Modelle betrachten, weshalb alle allgemeinen Aussagen über diese übergeordnete Klasse selbstverständlich auch Gültigkeit für diese Sonderfälle haben.

Um Missverständnisse zu vermeiden ist noch zu erwähnen, dass in Teilen der Literatur zur Zeitreihenanalyse Stationarität bereits in der Definition von *ARMA*-Zeitreihen vorausgesetzt wird, was in dieser Arbeit nicht der Fall ist, was eventuell zu Verwirrungen führen könnte, aber keine tatsächlichen Auswirkungen haben wird.

Definition 4.2 Backshift-Operator, *AR*- und *MA*-Polynom

Der Operator B , definiert durch $B^j \cdot X_t = X_{t-j} \quad \forall j \in \mathbb{N}_0$ heißt *Backshift-Operator*.

Sei $\Phi(x) := -\sum_{j=0}^p a_j x^j$ mit $a_0 = -1$ und $\Psi(x) := \sum_{j=0}^q b_j x^j$ mit $b_0 = 1$, so kann man die *ARMA*-Gleichung auch vereinfacht schreiben mittels

$$\Phi(B) \cdot X_t = \Psi(B) \cdot e_t.$$

Dabei bezeichnet man $\Phi(x)$ als *Autoregressives Polynom*, oder kurz *AR-Polynom*, und $\Psi(x)$ als *Moving-Average-Polynom*, oder kurz *MA-Polynom*.

Definition 4.3 Kausalität

Eine $ARMA(p, q)$ -Zeitreihe $X = (X_t : t \in \mathbb{Z})$ heißt *kausal*, wenn eine Folge reeller Koeffizienten $(\tilde{a}_j)_{j \in \mathbb{N}_0}$ mit $\sum_{j=0}^{\infty} |\tilde{a}_j| < \infty$ existiert, sodass gilt:

$$X_t = \sum_{j=0}^{\infty} \tilde{a}_j e_{t-j} \quad \forall t \in \mathbb{Z}$$

Eine $ARMA$ -Zeitreihe ist also kausal, wenn man sie als " $MA(\infty)$ "-Zeitreihe darstellen kann, oder, salopp ausgedrückt, wenn ihre Zufallsvariablen nicht von zukünftigen Werten des Weißen Rauschens abhängen.

Korollar 4.1

Eine kausale $ARMA(p, q)$ -Zeitreihe ist immer stationär und zentriert.

Beweis:

Die Aussage folgt unmittelbar aus Definition 4.3, Korollar 3.1 und Satz 2.3 . ■

Satz 4.1

Sei $X = (X_t : t \in \mathbb{Z})$ eine stationäre $ARMA(p, q)$ -Zeitreihe, die die Gleichung $\Phi(B) \cdot X_t = \Psi(B) \cdot e_t$ erfüllt und für die das AR- und das MA-Polynom keine gemeinsamen komplexen Nullstellen haben, dann gilt:

- (i) X ist genau dann kausal, wenn keine komplexen Nullstellen des AR-Polynoms $\Phi(x)$ existieren, deren Betrag ≤ 1 ist.
- (ii) Die Koeffizienten der Darstellung von X mittels $X_t = \sum_{j=0}^{\infty} \tilde{a}_j e_{t-j}$ sind eindeutig bestimmt durch

$$\sum_{j=0}^{\infty} \tilde{a}_j x^j = \frac{\Psi(x)}{\Phi(x)} \quad \forall x \in \mathbb{C}: |x| \leq 1 .$$

Beweis:

" $\Leftarrow$ " Gelte also für alle Nullstellen von Φ , dass deren Betrag größer als 1 ist. Dann existiert ein $\varepsilon > 0$, sodass der Ausdruck $\frac{1}{\Phi(x)}$ für alle komplexen Zahlen x mit $|x| \leq 1 + \varepsilon$ definiert ist.

$\frac{1}{\Phi(x)}$ lässt sich nun in Faktoren $\frac{1}{-a_p} \cdot \frac{1}{x-N_1} \cdot \dots \cdot \frac{1}{x-N_p}$ zerlegen, wobei $N_1, \dots, N_p$ die Nullstellen von $\Phi(x)$ bezeichnen. Jeder der Faktoren $\frac{1}{x-N_j}$ lässt sich jeweils als Potenzreihe um 0 mit

Konvergenzradius $|N_j|$ darstellen. Das Produkt dieser Potenzreihen, multipliziert mit dem konstanten

Faktor $\frac{1}{-a_p}$, lässt sich wiederum als Potenzreihe darstellen mit Konvergenzradius $R \geq \min_j |N_j|$.

Nachdem gilt, dass $\min_j |N_j| > 1 + \varepsilon$, folgt, dass die so entstandene Potenzreihe zumindest für $|x| < 1 + \varepsilon$ gegen $\frac{1}{\Phi(x)}$ konvergiert. Es existiert also eine Potenzreihe $c(x)$, für die gilt, dass

$$c(x) = \frac{1}{\Phi(x)} = \sum_{j=0}^{\infty} c_j x^j \quad \forall x \in \mathbb{C} \text{ mit } |x| < 1 + \varepsilon.$$

Aus der Konvergenz der Potenzreihe folgt nun, dass die Folge ihrer Summanden eine Nullfolge sein muss, d.h. es gilt $\lim_{j \rightarrow \infty} c_j x^j = 0$ und damit selbstverständlich auch $\lim_{j \rightarrow \infty} c_j \left(1 + \frac{\varepsilon}{2}\right)^j = 0$.

Es folgt daraus, dass es ein reelles K mit $0 < K < \infty$ geben muss, sodass gilt, dass

$$\begin{aligned} \left| c_j \cdot \underbrace{\left(1 + \frac{\varepsilon}{2}\right)^j}_{>0} \right| < K &\Leftrightarrow |c_j| \cdot \left(1 + \frac{\varepsilon}{2}\right)^j < K \Leftrightarrow |c_j| < \frac{K}{\left(1 + \frac{\varepsilon}{2}\right)^j} \quad \forall j \in \mathbb{N}_0 \\ \Rightarrow \sum_{j=0}^{\infty} |c_j| &< \sum_{j=0}^{\infty} \frac{K}{\left(1 + \frac{\varepsilon}{2}\right)^j} = K \cdot \sum_{j=0}^{\infty} \underbrace{\left(\frac{1}{1 + \frac{\varepsilon}{2}}\right)^j}_{<1} = K \cdot \left(\frac{1}{1 - \frac{1}{1 + \frac{\varepsilon}{2}}}\right) = K \cdot \frac{2}{\varepsilon} + K < \infty. \end{aligned}$$

Trivialerweise folgt außerdem: $\Phi(x) \cdot c(x) = 1 \quad \forall x \in \mathbb{C} \text{ mit } |x| < 1 + \varepsilon$.

Nachdem die Zeitreihe $X = (X_t : t \in \mathbb{Z})$ nach Voraussetzung stationär ist, gilt nach Satz 2.3, dass $c(B) \cdot X_t$ sowohl fast sicher absolut als auch im quadratischen Mittel (gegen jeweils denselben Grenzwert) konvergiert. Man darf also auf beide Seiten ihrer *ARMA*-Gleichung $\Phi(B) \cdot X_t = \Psi(B) \cdot e_t$ den Operator $c(B)$ anwenden und erhält

$$\Phi(B) \cdot X_t = \Psi(B) \cdot e_t \Leftrightarrow \underbrace{c(B) \cdot \Phi(B)}_{=1} \cdot X_t = c(B) \cdot \Psi(B) \cdot e_t \Leftrightarrow X_t = c(B) \cdot \Psi(B) \cdot e_t.$$

Nachdem $c(x) \cdot \Psi(x)$, eine Potenzreihe multipliziert mit einem Polynom, offensichtlich wiederum eine Potenzreihendarstellung besitzt, existiert eine Folge von Koeffizienten $(\tilde{a}_j)_{j \in \mathbb{N}_0}$, sodass gilt:

- $X_t = c(B) \cdot \Psi(B) \cdot e_t = \frac{\Psi(B)}{\Phi(B)} e_t = \sum_{j=0}^{\infty} \tilde{a}_j e_{t-j} \quad \forall t \in \mathbb{Z}$
- $\sum_{j=0}^{\infty} |\tilde{a}_j| < \infty$, was man auf dieselbe Weise folgern kann wie schon zuvor für $\sum_{j=0}^{\infty} |c_j| < \infty$.

Das bedeutet nun, dass X kausal ist, wobei $(\tilde{a}_j)_{j \in \mathbb{N}_0}$ gegeben ist durch:

$$\frac{\Psi(x)}{\Phi(x)} = \sum_{j=0}^{\infty} \tilde{a}_j x^j \quad \forall x \in \mathbb{C} \text{ mit } |x| < 1 + \varepsilon$$

Damit ist, unter der Voraussetzung, dass die Bedingungen in (i) erfüllt sind, (ii) bewiesen.

" $\Rightarrow$ " Sei X eine kausale Zeitreihe, dann existiert eine Folge von Koeffizienten $(\tilde{a}_j)_{j \in \mathbb{N}_0}$, sodass gilt:

$X_t = \sum_{j=0}^{\infty} \tilde{a}_j e_{t-j} \quad \forall t \in \mathbb{Z} \text{ mit } \sum_{j=0}^{\infty} |\tilde{a}_j| < \infty$. Sei $\tilde{a}(x) := \sum_{j=0}^{\infty} \tilde{a}_j x^j$, dann gilt, dass

$$|\tilde{a}(x)| = \left| \sum_{j=0}^{\infty} \tilde{a}_j x^j \right| < \sum_{j=0}^{\infty} |\tilde{a}_j| |x|^j < \sup |\tilde{a}_j| \cdot \sum_{j=0}^{\infty} |x|^j = \sup |\tilde{a}_j| \cdot \frac{1}{1 - |x|} < \infty \quad \text{für } |x| < 1.$$

Offensichtlich ist $|\tilde{a}(x)|$ für $|x| > 1$ divergent, d.h.: $\tilde{a}(x)$ hat Konvergenzradius 1.

Nun gilt $\Psi(B) \cdot e_t = \Phi(B) \cdot X_t = \Phi(B) \cdot \tilde{a}(B) \cdot e_t$.

Wiederum wird eine Potenzreihe mit einem Polynom multipliziert, weswegen dieses Produkt durch eine Potenzreihe mit minimalem Konvergenzradius 1 ausgedrückt werden kann. Untersucht man das Verhalten am Rand des Konvergenzbereiches, so erhält man

$$\Phi(x) \cdot \tilde{a}(x) \leq \underbrace{\left(\sum_{k=0}^p |a_k \cdot (\pm 1)^k| \right)}_{< \infty} \cdot \underbrace{\left(\sum_{j=0}^{\infty} |\tilde{a}_j \cdot (\pm 1)^j| \right)}_{< \infty} < \infty.$$

Die Potenzreihe $\Phi(x)\tilde{a}(x)$ konvergiert also auch absolut für $|x| = 1$. Deshalb kann man definieren:

$$\delta(x) := \Phi(x) \cdot \tilde{a}(x) = \sum_{j=0}^{\infty} \delta_j x^j \quad \forall x \in \mathbb{C} \text{ mit } |x| \leq 1.$$

Man erhält somit

$$\begin{aligned} \Psi(B) \cdot e_t &= \Phi(B) \cdot X_t \Leftrightarrow \Psi(B) \cdot e_t = \Phi(B) \cdot \tilde{a}(B) \cdot e_t \Leftrightarrow \Psi(B) \cdot e_t = \delta(B) \cdot e_t \Leftrightarrow \\ &\Leftrightarrow \sum_{j=0}^q b_j e_{t-j} = \sum_{j=0}^{\infty} \delta_j e_{t-j} \Leftrightarrow e_{t-k} \sum_{j=0}^q b_j e_{t-j} = e_{t-k} \sum_{j=0}^{\infty} \delta_j e_{t-j} \Rightarrow \\ &\Rightarrow \mathbb{E} \left(\sum_{j=0}^q b_j e_{t-j} e_{t-k} \right) = \mathbb{E} \left(\sum_{j=0}^{\infty} \delta_j e_{t-j} e_{t-k} \right) \Leftrightarrow \sum_{j=0}^q b_j \underbrace{\mathbb{E}(e_{t-j} e_{t-k})}_{=\begin{cases} \sigma_e^2, & j=k \\ 0, & \text{sonst} \end{cases}} = \sum_{j=0}^{\infty} \delta_j \underbrace{\mathbb{E}(e_{t-j} e_{t-k})}_{=\begin{cases} \sigma_e^2, & j=k \\ 0, & \text{sonst} \end{cases}}. \end{aligned}$$

Das „Hineinziehen“ des Erwartungswertes in die Reihe auf der rechten Seite ist folgendermaßen zu argumentieren: Man hat gesehen, dass $\delta(x)$ auch für $x = 1$ absolut konvergiert, weshalb $\sum_{j=0}^{\infty} \delta_j$ absolut konvergieren muss. Nach Satz 2.3 konvergiert nun $\sum_{j=0}^{\infty} \delta_j e_{t-j}$ absolut (sowohl fast sicher als auch in L^2) und damit trivialerweise auch $e_{t-k} \cdot \sum_{j=0}^{\infty} \delta_j e_{t-j} = \sum_{j=0}^{\infty} \delta_j e_{t-j} e_{t-k}$. Darum darf man hier Lemma 2.1 anwenden.

Man erhält nun

- $\forall k \leq q: b_k \sigma_e^2 = \delta_k \sigma_e^2 \Leftrightarrow b_k = \delta_k$
- $\forall k > q: 0 = \delta_k \underbrace{\sigma_e^2}_{>0} \Leftrightarrow \delta_k = 0$.

Somit gilt also: $\Psi(x) = \delta(x) = \Phi(x) \cdot \tilde{a}(x) \quad \forall x \in \mathbb{C} \text{ mit } |x| \leq 1$.

Indirekt angenommen $\Phi(x)$ hätte nun eine Nullstelle x_0 mit $|x_0| \leq 1$, dann müsste gelten, dass

$|\tilde{a}(x)| < \infty \quad \forall x \in \mathbb{C} \text{ mit } |x| \leq 1 \Rightarrow \Phi(x_0) \cdot \tilde{a}(x_0) = 0$ und, dass

$\Psi(x) = \Phi(x) \cdot \tilde{a}(x) \quad \forall x \in \mathbb{C} \text{ mit } |x| \leq 1 \text{ und } \Phi(x_0) \cdot \tilde{a}(x_0) = 0 \Rightarrow \Psi(x_0) = 0$.

Also müssten $\Psi(x)$ und $\Phi(x)$ die gemeinsame Nullstelle x_0 besitzen, was einen Widerspruch zur Voraussetzung darstellt, womit nun auch (i) bewiesen ist.

■

Anmerkung 4.2

Der vorhergegangene Satz hat nun zwar festgelegt, dass, gibt es unter den gegebenen Voraussetzungen an die *ARMA*-Gleichung tatsächlich eine stationäre Zeitreihe, die diese erfüllt, diese kausal sein muss. Noch nicht behandelt wurde die Frage, ob es denn überhaupt solche stationären *ARMA*-Zeitreihen gibt, eine Lücke, die der folgende Satz 4.2 schließen soll. Satz 4.3 soll dann abschließend die Bedingungen an das *AR*- und das *MA*-Polynom verallgemeinern um auch Aussagen bzgl. Kausalität für *ARMA*-Gleichungen machen zu können, deren *AR*- und *MA*-Polynome gemeinsame Nullstellen besitzen, was ja in Satz 4.1 explizit ausgeschlossen war.

Satz 4.2

Für jedes Weiße Rauschen und jede *ARMA*-Gleichung, für die das *AR*- und das *MA*-Polynom keine gemeinsamen komplexen Nullstellen haben, und für die keine Nullstellen des *AR*-Polynoms $\Phi(x)$ existieren, deren Betrag ≤ 1 ist, gilt folgende Aussage:

Es existiert genau eine stationäre *ARMA*-Zeitreihe $X = (X_t : t \in \mathbb{Z})$, die die gegebene *ARMA*-Gleichung für dieses Weiße Rauschen erfüllt. Diese ist kausal.

Beweis:

Aus Satz 4.1 weiß man, dass, wenn es eine solche stationäre *ARMA*-Zeitreihe X gibt, diese eindeutig bestimmt und kausal sein muss, das heißt, es kann zu jeder solchen *ARMA*-Gleichung nicht mehr als eine diese erfüllende stationäre Zeitreihe existieren.

Im Beweis von Satz 4.1 hat man gesehen, dass man für eine *ARMA*-Gleichung, die die gegebenen Voraussetzungen erfüllt, stets für alle x mit $|x| \leq 1$ die Potenzreihe $\tilde{a}(x) = \frac{\Psi(x)}{\Phi(x)}$ definieren kann, die immer die Bedingung $\sum_{j=0}^{\infty} |\tilde{a}_j| < \infty$ erfüllt.

Deshalb ist es nun ohne weiteres möglich für jedes Weiße Rauschen und für jede solche *ARMA*-Gleichung eine Zeitreihe $Y = (Y_t : t \in \mathbb{Z})$ mit $Y_t := \tilde{a}(B)e_t = \sum_{j=0}^{\infty} \tilde{a}_j e_{t-j} \quad \forall t \in \mathbb{Z}$ zu definieren.

Nun muss noch nachgewiesen werden, dass diese Zeitreihe Y auch tatsächlich stets eine stationäre *ARMA*-Zeitreihe ist, die die gegebene *ARMA*-Gleichung für das vorgegebene Weiße Rauschen erfüllt. Zunächst ist nach Satz 2.3 ersichtlich, dass Y stationär sein muss. Durch Einsetzen von $X_t = \tilde{a}(B)e_t$ in die *ARMA*-Gleichung erhält man $\Phi(B) \cdot Y_t = \Phi(B) \cdot \tilde{a}(B)e_t = \Psi(B) \cdot e_t$.

Also erfüllt Y diese *ARMA*-Gleichung und ist somit eine stationäre *ARMA*-Zeitreihe. ■

Definition 4.4 Reduzierte *ARMA*-Gleichung

Existieren gemeinsame Nullstellen $N_0, \dots, N_k$ des *AR*- und des *MA*-Polynoms einer *ARMA*-Gleichung $\Phi(B) \cdot X_t = \Psi(B) \cdot e_t$, dann nennt man die folgende Gleichung die zugehörige *reduzierte ARMA-Gleichung*:

$$\frac{\Phi(B)}{\prod_{j=0}^k (B - N_j)} \cdot X_t = \frac{\Psi(B)}{\prod_{j=0}^k (B - N_j)} \cdot e_t \quad .$$

Seien $\tilde{\Phi}(x) := \frac{\Phi(x)}{\prod_{j=0}^k (x - N_j)}$ und $\tilde{\Psi}(x) := \frac{\Psi(x)}{\prod_{j=0}^k (x - N_j)}$, so bezeichnet man $\tilde{\Phi}(x)$ als *reduziertes AR-Polynom* und $\tilde{\Psi}(x)$ als *reduziertes MA-Polynom*.

Existieren keine gemeinsamen Nullstellen von $\Phi(x)$ und $\Psi(x)$, so nennt man $\Phi(B) \cdot X_t = \Psi(B) \cdot e_t$ *reduzierte ARMA-Gleichung*, $\Phi(x)$ *reduziertes AR-* und $\Psi(x)$ *reduziertes MA-Polynom*.

Satz 4.3

Für jedes Weiße Rauschen und jede *ARMA*-Gleichung, deren reduziertes *AR*-Polynom keine Nullstellen besitzt, deren Betrag ≤ 1 ist, gilt folgende Aussage:

Es existiert genau eine stationäre *ARMA*-Zeitreihe $X = (X_t : t \in \mathbb{Z})$, die die gegebene *ARMA*-Gleichung für dieses Weiße Rauschen erfüllt. Diese ist kausal.

Wenn nicht alle Nullstellen des *AR*- und des *MA*-Polynoms gleich sind, so sind die Koeffizienten der Darstellung von X mittels $X_t = \sum_{j=0}^{\infty} \tilde{a}_j e_{t-j}$ eindeutig bestimmt durch

$$\sum_{j=0}^{\infty} \tilde{a}_j x^j = \frac{\tilde{\Psi}(x)}{\tilde{\Phi}(x)} \quad \forall x \in \mathbb{C}: |x| \leq 1 .$$

Dabei bezeichnet $\tilde{\Phi}(x)$ das reduzierte *AR*-Polynom und $\tilde{\Psi}(x)$ das reduzierte *MA*-Polynom.

Beweis:

Seien $N_0, \dots, N_k$ alle gemeinsamen Nullstellen von $\Phi(x)$ und $\Psi(x)$, so gilt für die eindeutig bestimmten reduzierten Polynome $\tilde{\Phi}(x)$ und $\tilde{\Psi}(x)$, dass

$$\Phi(x) = \tilde{\Phi}(x) \cdot \prod_{j=0}^k (x - N_j) \text{ und } \Psi(x) = \tilde{\Psi}(x) \cdot \prod_{j=0}^k (x - N_j) .$$

Setzt man dies nun in die *ARMA*-Gleichung ein erhält man

$$\begin{aligned} \Phi(B) \cdot X_t = \Psi(B) \cdot e_t &\Leftrightarrow \tilde{\Phi}(B) \cdot X_t \cdot \prod_{j=0}^k (B - N_j) = \tilde{\Psi}(B) \cdot e_t \cdot \prod_{j=0}^k (B - N_j) \Leftrightarrow \\ &\Leftrightarrow \tilde{\Phi}(B) \cdot X_t = \tilde{\Psi}(B) \cdot e_t . \end{aligned}$$

Das bedeutet: Eine Zeitreihe erfüllt eine *ARMA*-Gleichung genau dann, wenn sie deren reduzierte *ARMA*-Gleichung erfüllt. Nachdem nun $\tilde{\Phi}(x)$ und $\tilde{\Psi}(x)$ keine gemeinsamen Nullstellen besitzen können und das reduzierte *AR*-Polynom nach Voraussetzung keine Nullstellen besitzt, deren Betrag ≤ 1 ist, kann man auf die reduzierte *ARMA*-Gleichung nun direkt Satz 4.2 anwenden. ■

Anmerkung 4.3

Genauso wie für *AR*(1)-Zeitreihen gibt es auch stationäre *ARMA*-Zeitreihen, die nicht kausal sind, nämlich dann, wenn das *AR*-Polynom keine Nullstellen mit Betrag ≥ 1 besitzt.

Ebenfalls analog zu den *AR*(1)-Zeitreihen existieren überhaupt keine stationären *ARMA*-Zeitreihen, wenn das reduzierte *AR*-Polynom mindestens eine Nullstelle mit Betrag 1 besitzt.

Wie bereits bei den *AR*(1)-Zeitreihen kurz angesprochen wurde, sind für die Praxis in erster Linie

kausale $ARMA$ -Zeitreihen interessant, weswegen hier auf die diesbezüglichen Sätze und Beweise verzichtet wird.

Definition 4.5 Invertierbarkeit

Eine $ARMA(p, q)$ -Zeitreihe $X = (X_t : t \in \mathbb{Z})$ heißt *invertierbar*, wenn eine Folge reeller Koeffizienten $(\tilde{b}_j)_{j \in \mathbb{N}_0}$ existiert mit $\sum_{j=0}^{\infty} |\tilde{b}_j| < \infty$, sodass gilt, dass

$$e_t = \sum_{j=0}^{\infty} \tilde{b}_j X_{t-j} \quad \forall t \in \mathbb{Z} .$$

Eine $ARMA$ -Zeitreihe ist also invertierbar, wenn sich die einzelnen Zufallsvariablen e_t des Weißen Rauschens eindeutig durch vergangene Elemente von X bestimmen lassen, bzw. wenn sich X als " $AR(\infty)$ "-Zeitreihe darstellen lässt.

Satz 4.4

Sei $X = (X_t : t \in \mathbb{Z})$ eine stationäre $ARMA(p, q)$ -Zeitreihe, die die Gleichung $\Phi(B) \cdot X_t = \Psi(B) \cdot e_t$ erfüllt, dann gilt:

- (i) X ist genau dann invertierbar, wenn das zugehörige reduzierte MA -Polynom $\tilde{\Psi}(x)$ keine Nullstellen besitzt, deren Betrag ≤ 1 ist
- (ii) Die Koeffizienten der Darstellung von X mittels $e_t = \sum_{j=0}^{\infty} \tilde{b}_j X_{t-j}$ sind durch das reduzierte AR -Polynom $\tilde{\Phi}(x)$ und das reduzierte MA -Polynom $\tilde{\Psi}(x)$ bestimmt mittels

$$\sum_{j=0}^{\infty} \tilde{b}_j x^j = \frac{\tilde{\Phi}(x)}{\tilde{\Psi}(x)} \quad \forall x \in \mathbb{C}: |x| \leq 1 .$$

Beweis:

Haben das AR - und das MA -Polynom keine gemeinsamen Nullstellen und vertauscht man im Beweis von Satz 4.1 einfach $\Psi(x)$ mit $\Phi(x)$, so folgt der Rest der Beweisführung im Wesentlichen analog. Existieren wiederum gemeinsame Nullstellen, so erfolgt der Beweis der Aussage im Wesentlichen analog zum Beweis von Satz 4.3.

■

Anmerkung 4.4

Erfüllt eine $ARMA$ -Gleichung sowohl die Bedingungen von Satz 4.3 als auch jene von Satz 4.4, so existiert offensichtlich wiederum genau eine stationäre Zeitreihe X , die diese $ARMA$ -Gleichung erfüllt, wobei X dann klarerweise kausal und invertierbar ist.

Wie in späteren Kapiteln noch ersichtlich werden wird, wird bei der Suche nach einem passenden $ARMA$ -Modell zu einer real beobachteten Zeitreihe dessen Koeffizienten (auch) anhand der (empirischen) Autokorrelationsfunktion geschätzt. Setzt man Invertierbarkeit und Kausalität – und die damit verbundenen Anforderungen an die Polynome $\Phi(x)$ und $\Psi(x)$ – nicht voraus, so kann dies

aufgrund fehlender Eindeutigkeit der Autokorrelationsfunktion zu Problemen bei der richtigen Modellwahl führen, wie die folgenden beiden Beispiele zeigen sollen.

Beispiel 4.1

Sei X eine beliebige $MA(1)$ -Zeitreihe, d.h. gelte, dass $X_t = e_t + b \cdot e_{t-1}$, dann ist nach dem Beweis von Satz 3.1 ersichtlich, dass die Autokorrelationsfunktion von X folgendermaßen bestimmt ist:

$$\rho(h) = \begin{cases} 1, & h = 0 \\ \frac{b}{1+b^2}, & h = 1 \\ 0, & \text{sonst} \end{cases} . \quad \text{Es gilt nun aber: } \frac{b}{1+b^2} = \frac{\frac{b}{b^2}}{\frac{1+b^2}{b^2}} = \frac{\left(\frac{1}{b}\right)}{1+\left(\frac{1}{b}\right)^2} .$$

Die Autokorrelationsfunktion von $MA(1)$ -Zeitreihen mit Koeffizient b ist also die gleiche wie jene von $MA(1)$ -Zeitreihen mit Koeffizient $\frac{1}{b}$, nur durch Kenntnis der Autokorrelationsfunktion lässt sich also keine Entscheidung darüber treffen, welche der beiden Zeitreihen für die Modellierung der vorliegenden Realisierung besser geeignet ist.

Setzt man aber bei der Modellsuche für den Koeffizienten bereits Invertierbarkeit voraus, so kann man von vornherein einen – nämlich den für das weitere Vorgehen „ungünstigeren“ - der beiden Koeffizienten ausschließen.

Beispiel 4.2

Dieselbe Problematik betrifft auch die Autokorrelationsfunktion von $AR(p)$ -Zeitreihen.

Betrachtet man beispielsweise die stationäre $AR(1)$ -Zeitreihe X , die die Gleichung $X_t = e_t + aX_{t-1}$ erfüllt, so ist gemäß Satz 3.4 für $|a| < 1$ deren Autokorrelationsfunktion bestimmt durch

$$|a| > 1 \Rightarrow \rho(h) = a^h \quad \text{und} \quad |a| < 1 \Rightarrow \rho(h) = a^{-h} .$$

Das bedeutet nun, dass für jeden Koeffizienten a mit $|a| \neq 1$ die $AR(1)$ -Zeitreihen Y und Z , die die Gleichungen $Y_t = e_t + aY_{t-1}$ bzw. $Z_t = e_t + \frac{1}{a}Z_{t-1}$ erfüllen, die selbe Autokorrelationsfunktion besitzen, weshalb es sinnvoll ist, die Erfüllung der Bedingungen von Satz 4.3 für das AR -Polynom und damit eben auch Kausalität vorauszusetzen, um somit von vornherein den „ungünstigeren“ der beiden Koeffizienten ausschließen zu können.

Satz 4.5

$MA(q)$ -Zeitreihen sind immer kausal, $AR(p)$ -Zeitreihen immer invertierbar.

Beweis: trivial.

■

Satz 4.6

- (i) Sei X eine kausale $ARMA$ -Zeitreihe mit der Darstellung $X_t = \sum_{j=0}^{\infty} \tilde{a}_j e_{t-j}$, so fallen die Koeffizienten $\tilde{a}_j$ geometrisch ab, das bedeutet:

$$\exists K > 0 ; \exists c > 1 : |\tilde{a}_j| < K \cdot c^{-j} \quad \forall j \geq 0$$

- (ii) Sei X eine invertierbare $ARMA$ -Zeitreihe mit der Darstellung $e_t = \sum_{j=0}^{\infty} \tilde{b}_j X_{t-j}$, so fallen die Koeffizienten $\tilde{b}_j$ geometrisch ab, das bedeutet:

$$\exists K > 0 ; \exists c > 1 : |\tilde{b}_j| < K \cdot c^{-j} \quad \forall j \geq 0$$

Beweis:

- (i) Gemäß dem Beweis von Satz 4.1 weiß man, dass ein $\delta > 0$ existiert, sodass die Potenzreihe $\sum_{j=0}^{\infty} \tilde{a}_j x^j$ für alle x mit $|x| \leq 1 + \delta$ konvergiert. Es muss also offensichtlich im speziellen für $x = 1 + \delta$ gelten, dass $\lim_{j \rightarrow \infty} [\tilde{a}_j \cdot (1 + \delta)^j] = 0$ sein muss.

Es muss nun also ein $K > 0$ existieren, sodass für alle $j \geq 0$ gilt, dass:

$$K > \left| \tilde{a}_j \cdot \underbrace{(1 + \delta)^j}_{>0} \right| = |\tilde{a}_j| \cdot \left(\underbrace{1 + \delta}_{=:c} \right)^j \Leftrightarrow |\tilde{a}_j| < K \cdot c^{-j} .$$

- (ii) Analog mittels des Beweises von Satz 4.4. ■

Satz 4.7 Bestimmung der Koeffizienten $\tilde{a}_j$ und $\tilde{b}_j$

Seien $a_j := 0$ für $j \notin [0, p]$, $b_j := 0$ für $j \notin [0, q]$, und $\tilde{a}_j := \tilde{b}_j := 0$ für $j < 0$, dann gilt:

- (i) Sei X eine kausale $ARMA(p, q)$ -Zeitreihe, die die Gleichung $\Phi(B) \cdot X_t = \Psi(B) \cdot e_t$ erfüllt. Dann gilt $\tilde{a}_j = b_j + \sum_{k=1}^p \tilde{a}_{j-k} a_k \quad \forall j \in \mathbb{N}_0$.
- (ii) Sei X eine invertierbare $ARMA(p, q)$ -Zeitreihe, die die Gleichung $\Phi(B) \cdot X_t = \Psi(B) \cdot e_t$ erfüllt. Dann gilt $\tilde{b}_j = -a_j - \sum_{k=1}^q \tilde{b}_{j-k} b_k \quad \forall j \in \mathbb{N}_0$.

Beweis:

Sei o.B.d.A die $ARMA$ -Gleichung $\Phi(B) \cdot X_t = \Psi(B) \cdot e_t$ bereits reduziert.

- (i) Nach Satz 4.3 gilt für alle $x \in \mathbb{C}$ mit $|x| \leq 1$, dass

$$\begin{aligned} \tilde{a}(x) &= \frac{\Psi(x)}{\Phi(x)} \Leftrightarrow \tilde{a}(x) \cdot \Phi(x) = \Psi(x) \Leftrightarrow \left(\sum_{j=0}^{\infty} \tilde{a}_j x^j \right) \cdot \left(- \sum_{k=0}^p a_k x^k \right) = \sum_{j=0}^q b_j x^j \Leftrightarrow \\ &\Leftrightarrow - \sum_{j=0}^{\infty} \left(\sum_{k=0}^j \tilde{a}_{j-k} a_k \right) x^j = \sum_{j=0}^q b_j x^j \Leftrightarrow - \sum_{j=0}^q \left(\sum_{k=0}^j \tilde{a}_{j-k} a_k \right) x^j = \sum_{j=0}^q b_j x^j . \end{aligned}$$

Durch Koeffizientenvergleich erhält man nun für alle j aus $\mathbb{Z}$, dass

$$- \sum_{k=0}^p \tilde{a}_{j-k} a_k = - \tilde{a}_j \cdot \underbrace{a_0}_{=-1} - \sum_{k=1}^p \tilde{a}_{j-k} a_k = b_j \Leftrightarrow \tilde{a}_j = b_j + \sum_{k=1}^p \tilde{a}_{j-k} a_k .$$

(ii) Nach Satz 4.4 gilt für alle $x \in \mathbb{C}$ mit $|x| \leq 1$, dass

$$\begin{aligned} \tilde{b}(x) = \frac{\Phi(x)}{\Psi(x)} &\Leftrightarrow \tilde{b}(x) \cdot \Psi(x) = \Phi(x) \Leftrightarrow \left(\sum_{j=0}^{\infty} \tilde{b}_j x^j \right) \cdot \left(\sum_{k=0}^q b_k x^k \right) = - \sum_{k=0}^p a_k x^k \Leftrightarrow \\ &\Leftrightarrow \sum_{j=0}^{\infty} \left(\sum_{k=0}^j \tilde{b}_{j-k} b_k \right) x^j = - \sum_{k=0}^p a_k x^k \Leftrightarrow \sum_{j=0}^p \left(\sum_{k=0}^j \tilde{b}_{j-k} b_k \right) x^j = - \sum_{k=0}^p a_k x^k . \end{aligned}$$

Wiederum durch Koeffizientenvergleich erhält man für alle j aus $\mathbb{Z}$, dass

$$\sum_{k=0}^j \tilde{b}_{j-k} b_k = \underbrace{\tilde{b}_j b_0}_{=1} + \sum_{k=1}^j \tilde{b}_{j-k} b_k = -a_j \Leftrightarrow \tilde{b}_j = -a_j - \sum_{k=1}^j \tilde{b}_{j-k} b_k .$$

■

Beispiel 4.3

Gegeben sei die $ARMA(3,2)$ -Gleichung $X_t = \frac{1}{2}X_{t-1} - \frac{1}{4}X_{t-2} + \frac{1}{8}X_{t-3} + e_t + \frac{1}{9}e_{t-2}$.

Die Nullstellen von $\Phi(x) = 1 - \frac{1}{2}x + \frac{1}{4}x^2 - \frac{1}{8}x^3$ sind 2 und $\pm 2i$, deren Betrag offensichtlich größer als 1 ist. Die Nullstellen von $\Psi(x) = 1 + \frac{1}{9}x^2$ wiederum sind $\pm 3i$ und ihr Betrag ist ebenfalls größer als 1. Außerdem haben $\Phi(x)$ und $\Psi(x)$ keine gemeinsame Nullstellen, also ist die $ARMA$ -Gleichung bereits reduziert. Es existiert also zu dieser Gleichung für jedes Weiße Rauschen e eine eindeutig bestimmte kausale und invertierbare $ARMA(3,2)$ -Zeitreihe X . Die Koeffizienten der Darstellung von X durch $X_t = \sum_{j=0}^{\infty} \tilde{a}_j e_{t-j}$ ergeben sich nun nach Satz 4.7 folgendermaßen:

$\tilde{a}_0$	$= b_0 = 1$
$\tilde{a}_1$	$= b_1 + a_1 = 0 + a_1 = \frac{1}{2}$
$\tilde{a}_2$	$= b_2 + \tilde{a}_1 a_1 + \tilde{a}_0 a_2 = \frac{1}{9} + \frac{1}{2} \cdot \frac{1}{2} + 1 \cdot \left(-\frac{1}{4}\right) = \frac{1}{9}$
$\tilde{a}_j \quad \forall j > 2$	$= \underbrace{b_j}_{=0} + \tilde{a}_{j-1} a_1 + \tilde{a}_{j-2} a_2 + \tilde{a}_{j-3} a_3 = \frac{\tilde{a}_{j-1}}{2} - \frac{\tilde{a}_{j-2}}{4} + \frac{\tilde{a}_{j-3}}{8}$

Zusätzlich sollen nun die Koeffizienten $\tilde{b}_j$ der Darstellung der Zeitreihe durch $e_t = \sum_{j=0}^{\infty} \tilde{b}_j X_{t-j}$ bestimmt werden. Man erhält wiederum gemäß Satz 4.7 :

$\tilde{b}_0$	$= -a_0 = 1$
$\tilde{b}_1$	$= -a_1 - \tilde{b}_0 b_1 = -\frac{1}{2}$
$\tilde{b}_2$	$= -a_2 - \tilde{b}_1 b_1 - \tilde{b}_0 b_2 = \frac{5}{36}$
$\tilde{b}_3$	$= -a_3 - \tilde{b}_2 b_1 - \tilde{b}_1 b_2 = -\frac{5}{72}$
$\tilde{b}_j \quad \forall j > 3$	$= \underbrace{-a_j}_{=0} - \tilde{b}_3 \underbrace{b_1}_{=0} - \tilde{b}_{j-2} b_2 = -\frac{\tilde{b}_{j-2}}{9}$

Beispiel 4.4

Gegeben sei die $ARMA(2,1)$ -Gleichung $X_t = -\frac{1}{4}X_{t-1} + \frac{1}{8}X_{t-2} + e_t - \frac{1}{4}e_{t-1}$.

Die Nullstellen von $\Phi(x) = 1 + \frac{1}{4}x - \frac{1}{8}x^2$ sind 4 und -2 , die einzige Nullstelle von $\Psi(x) = 1 - \frac{1}{4}x$ ist 4.

Das AR - und das MA -Polynom haben also die gemeinsame Nullstelle 4, weshalb man die $ARMA$ -Gleichung erst reduzieren muss mittels

$$\begin{aligned} X_t &= -\frac{1}{4}X_{t-1} + \frac{1}{8}X_{t-2} + e_t - \frac{1}{4}e_{t-1} \Leftrightarrow \left(1 + \frac{1}{4}B - \frac{1}{8}B^2\right)X_t = \left(1 - \frac{1}{4}B\right)e_t \Leftrightarrow \\ &\Leftrightarrow -\frac{1}{8}(B-4)(B+2)X_t = -\frac{1}{4}(B-4)e_t \Leftrightarrow -\frac{1}{8}(B+2)X_t = -\frac{1}{4}e_t \Leftrightarrow (B+2)X_t = 2e_t \Leftrightarrow \\ &\Leftrightarrow 2X_t + X_{t-1} = 2e_t \Leftrightarrow X_t = -\frac{1}{2}X_{t-1} + e_t. \end{aligned}$$

Man kann also diese $ARMA(2,1)$ -Zeitreihe durch eine $AR(1)$ -Zeitreihe mit Koeffizient $a = -\frac{1}{2}$ ausdrücken. Da $\tilde{\Phi}(x) = 1 + \frac{1}{2}x$ nun lediglich die Nullstelle -2 besitzt und deren Betrag offensichtlich nicht ≤ 1 ist, existiert nun zu jedem Weißen Rauschen eine eindeutig bestimmte stationäre Zeitreihe X , die kausal ist. Nachdem es sich um eine $AR(1)$ -Zeitreihe handelt, ist die Zeitreihe außerdem invertierbar.

Wie im Beweis von Satz 3.3 bereits ausgeführt wurde, kann man jeden $AR(1)$ -Prozess mit $|a| < 1$ ausdrücken durch $X_t = \sum_{j=0}^{\infty} a^j e_{t-j}$, wodurch ersichtlich ist, dass $\tilde{a}_j = a^j \forall j \in \mathbb{N}_0$.

Die Folge der $\tilde{b}_j$ lässt sich nun sehr direkt bestimmen. Es gilt

$$X_t = -\frac{1}{2}X_{t-1} + e_t \Leftrightarrow e_t = X_t + \frac{1}{2}X_{t-1} \Leftrightarrow e_t = \sum_{j=0}^{\infty} \tilde{b}_j X_{t-j} \quad \text{für } \tilde{b}_0 = 1, \tilde{b}_1 = \frac{1}{2} \text{ und } \tilde{b}_j = 0 \forall j > 1.$$

5. Die Autokorrelationsfunktion kausaler $ARMA$ -Zeitreihen

Bevor sich diese Arbeit nun, nachdem das $ARMA$ -Modell vorgestellt wurde, den eher statistischen, anwendungsbezogeneren Aspekten der Zeitreihenanalyse widmet, soll noch kurz das Augenmerk näher auf bisher nicht behandelte Aspekte der Autokorrelationsfunktion kausaler $ARMA$ -Zeitreihen gelegt werden. Diese wird später eine Grundlage bilden, aus der Modelle für real vorliegende Realisierungen von Zeitreihen entwickelt werden können.

Satz 5.1 Bestimmung der Autokorrelationsfunktion

Sei X eine kausale $ARMA$ -Zeitreihe, gegeben für alle t aus $\mathbb{Z}$ durch die Gleichung $X_t = \sum_{j=0}^{\infty} \tilde{a}_j e_{t-j}$, so ist ihre Autokovarianzfunktion γ eindeutig bestimmt durch

$$\gamma(h) = \sigma_e^2 \cdot \sum_{j=0}^{\infty} \tilde{a}_j \tilde{a}_{j+|h|}.$$

Beweis:

Nachdem gemäß Korollar 3.1 das Weiße Rauschen e eine stationäre Zeitreihe sein muss, kann man Satz 2.3(v) anwenden um die Autokovarianzfunktion von X zu bestimmen. Sei γ_e die Autokovarianzfunktion von e , so erhält man

$$\begin{aligned}\gamma(h) &= \gamma(|h|) = \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \tilde{a}_j \tilde{a}_k \gamma_e(|h| - k + j) = \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \tilde{a}_j \tilde{a}_k \cdot \begin{cases} \sigma_e^2, & |h| - k + j = 0 \\ 0, & \text{sonst} \end{cases} = \\ &= \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \tilde{a}_j \tilde{a}_k \cdot \begin{cases} \sigma_e^2, & k = j + |h| \\ 0, & \text{sonst} \end{cases} = \sigma_e^2 \cdot \sum_{j=0}^{\infty} \tilde{a}_j \tilde{a}_{j+|h|} .\end{aligned}$$

■

Satz 5.2 Geometrische Abnahme der Autokovarianzfunktion

Sei X eine kausale $ARMA$ -Zeitreihe, so nehmen die Werte ihrer Autokovarianzfunktion γ geometrisch ab, das heißt es gilt

$$\exists L > 0 ; \exists c > 1 : |\gamma(h)| < L \cdot c^{-|h|} .$$

Beweis:

Aus Satz 4.6 weiß man, dass für die Koeffizienten $\tilde{a}_j$ der Darstellung der Zeitreihe X durch $X_t = \sum_{j=0}^{\infty} \tilde{a}_j e_{t-j}$ eine Konstante $K > 0$ und eine Zahl $c > 1$ existieren, sodass für alle $j > 0$ gilt, dass $|\tilde{a}_j| < K \cdot c^{-j}$. Gemäß Satz 5.1 erhält man nun

$$\begin{aligned}|\gamma(h)| &= \left| \sigma_e^2 \cdot \sum_{j=0}^{\infty} \tilde{a}_j \tilde{a}_{j+|h|} \right| < \sigma_e^2 \cdot \sum_{j=0}^{\infty} |\tilde{a}_j| \cdot |\tilde{a}_{j+|h|}| < \sigma_e^2 \cdot \sum_{j=0}^{\infty} K \cdot c^{-j} \cdot K \cdot c^{-j-|h|} = \\ &= K^2 \cdot \sigma_e^2 \cdot \sum_{j=0}^{\infty} c^{-2j} \cdot c^{-|h|} = \underbrace{\left(K^2 \cdot \sigma_e^2 \cdot \sum_{j=0}^{\infty} \left(\frac{1}{c} \right)^{2j} \right)}_{:=L (<\infty)} \cdot c^{-|h|} = L \cdot c^{-|h|} .\end{aligned}$$

■

Beispiel 5.1

Gegeben sei die $ARMA(3,2)$ -Gleichung $X_t = \frac{1}{2}X_{t-1} - \frac{1}{4}X_{t-2} + \frac{1}{8}X_{t-3} + e_t + \frac{1}{9}e_{t-2}$ mit dem $N(0,2^2)$ -verteilten Weißen Rauschen e . Aus Beispiel 4.3 weiß man, dass es eine kausale (und invertierbare) Zeitreihe X gibt, die diese Gleichung erfüllt. Außerdem wurden dort bereits die Koeffizienten $\tilde{a}_j$ für X bestimmt. Berechnet man nun beispielsweise die ersten 8 Koeffizienten numerisch, so ergeben sich folgende Werte:

$\tilde{a}_0$	$\tilde{a}_1$	$\tilde{a}_2$	$\tilde{a}_3$	$\tilde{a}_4$	$\tilde{a}_5$	$\tilde{a}_6$	$\tilde{a}_7$
1	0,5	0,1	0,05	0,0625	0,03125	0,00694	0,003472

Eine gute Näherung an die Werte der Autokovarianzfunktion von X erhält man nun mittels

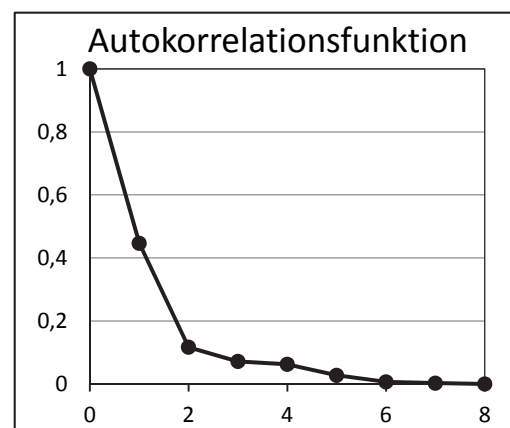
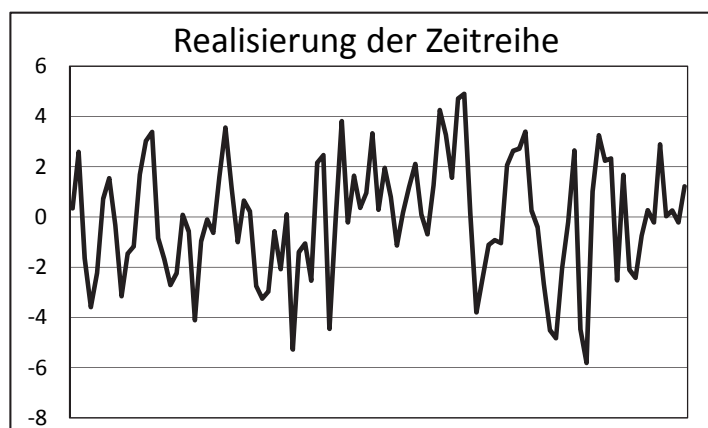
$$\gamma(h) \approx \sigma_e^2 \cdot \sum_{j=0}^7 \tilde{a}_j \tilde{a}_{j+|h|} = 4 \cdot \sum_{j=0}^7 \tilde{a}_j \tilde{a}_{j+|h|}, \text{ wobei } \tilde{a}_j := 0 \forall j > 7.$$

Man erhält so die folgenden Näherungswerte:

$\gamma(0)$	$\gamma(1)$	$\gamma(2)$	$\gamma(3)$	$\gamma(4)$	$\gamma(5)$	$\gamma(6)$	$\gamma(7)$	$\gamma(h > 7)$
5,0815	2,2696	0,5924	0,3635	0,3164	0,1404	0,0347	0,0139	0

Aus diesen Werten kann man nun mittels $\rho(h) = \frac{\gamma(h)}{\gamma(0)}$ sehr einfach Näherungswerte für die Autokorrelationsfunktion berechnen und erhält:

$\rho(0)$	$\rho(1)$	$\rho(2)$	$\rho(3)$	$\rho(4)$	$\rho(5)$	$\rho(6)$	$\rho(7)$	$\rho(h > 7)$
1	0,4466	0,1166	0,0715	0,0623	0,0276	0,0068	0,0027	0



Satz 5.3

Die Autokorrelationsfunktion einer $MA(q)$ -Zeitreihe verschwindet für Lags größer als q .

Beweis:

Siehe den Beweis von Satz 3.1. ■

Satz 5.4

Die partielle Autokorrelationsfunktion einer kausalen $AR(p)$ -Zeitreihe verschwindet für Lags größer als p .

Beweis:

Sei X die kausale Zeitreihe, die die $AR(p)$ -Gleichung $X_t - \sum_{j=1}^p a_j X_{t-j} = e_t$ für alle $t \in \mathbb{Z}$ erfüllt. Sei außerdem für beliebiges $h > 0$ der Raum H definiert durch $H := \text{span}\{X_t, X_{t+1}, \dots, X_{t+h-1}\}$ und sei $Y \in H$ beliebig.

Aufgrund der Kausalität von X gilt nun:

$$0 = \langle e_{t+h}, Y \rangle = \langle X_{t+h} - \sum_{j=1}^p a_j X_{t+h-j}, Y \rangle \Rightarrow \sum_{j=1}^p a_j X_{t+h-j} = P_H(X_{t+h}) .$$

Insbesondere ergibt sich somit für $h > p$, dass

$$\begin{aligned} P_H(X_{t+h}) &= d_{h,1}X_{t+h-1} + \dots + d_{h,h-1}X_{t+1} + \pi(h)X_t = \\ &= a_1X_{t+h-1} + \dots + a_pX_{t+h-p} + 0 \cdot X_{t+h-p-1} + \dots + 0 \cdot X_t . \end{aligned}$$

Hieraus folgt trivialerweise, dass für $h > p$ die partielle Autokorrelation $\pi(h)=0$ sein muss. ■

6. Das Komponentenmodell

In diesem und den folgenden Kapiteln wird es sich nun um die Entwicklung geeigneter *ARMA*-Modelle zur Beschreibung von in der Realität auftretenden Zeitreihen drehen. Dabei ist es offensichtlich, dass zu untersuchende Zeitreihen das Kriterium der Stationarität erfüllen sollten, da ansonsten Zukunftsprognosen nicht sinnvoll möglich wären. Des Weiteren ist es auch klar ersichtlich, dass die zu behandelnden Zeitreihen jedenfalls kausal und idealerweise auch invertierbar sein sollten.

Nun ist es aber im Allgemeinen so, dass solche auf Beobachtungen und Messungen basierenden Zeitreihen schon auf den ersten Blick als nicht-stationär erkennbar sind. In den meisten Fällen liegt dies daran, dass stochastische Zeitreihenmodelle wie das bereits ausführlich vorgestellte *ARMA*-Modell so genannte reine nicht-deterministische Modelle sind, während Ereignisse und Zustände in der Realität meist nicht nur vom (gewissen Regeln folgenden) Zufall abhängig sind, sondern auch von äußeren Begebenheiten beeinflusst werden.

So lässt sich oftmals schon mit bloßem Auge bei Betrachtung des Graphen einer Zeitreihe ein Trend feststellen. Ebenfalls zu beobachten ist in sehr vielen Fällen ein zyklisches Verhalten. Ein Beispiel hierfür wäre etwa eine monatliche Aufzeichnung der Anzahl der Übernachtungen in der österreichischen Tourismusbranche. Diese zeigt einen der allgemeinen Wirtschaftslage geschuldeten Trend auf der einen und klare zyklische Strukturen – im Winter und Sommer gibt es stets mehr Übernachtungen als im Herbst – auf der anderen Seite, was Stationarität ausschließt, bleibt ja weder der Erwartungswert noch die Varianz konstant. Dieses Kapitel wird sich nun mit Methoden beschäftigen die Trendkomponente wie auch die zyklische Komponente aus einer Zeitreihe herauszufiltern um eine „Restzeitreihe“ zu erhalten, der man Stationarität unterstellen kann und für die es sinnvoll ist, nach sie möglichst gut beschreibenden kausalen (und invertierbaren) *ARMA*-Modellen zu suchen.

Definition 6.1 Realisierung einer Zeitreihe

Sei $Y := (Y_t : t \in \mathbb{Z})$ eine beliebige Zeitreihe und sei $n \geq 1$, so nennt man $\widehat{Y}_{[n]} := (\widehat{Y}_t : t \in [1, n])$ eine *Realisierung* der Zeitreihe Y vom *Umfang* n , wenn gilt, dass $\widehat{Y}_t$ für alle $t \in [1, n]$ eine Realisierung von Y_t ist. In diesem Fall nennt man Y auch die $\widehat{Y}_{[n]}$ zugrunde liegende Zeitreihe.

Definition 6.2 Komponentenmodell

Sei $Y := (Y_t : t \in \mathbb{Z})$ eine beliebige Zeitreihe, so nennt man die Gleichung $Y_t = T_t + S_t + X_t$ das *Additive Komponentenmodell* von Y und $Y_t = T_t \cdot S_t \cdot X_t$ das *Multiplikative Komponentenmodell* von Y , wobei man die Zeitreihe $T := (T_t : t \in \mathbb{Z})$ die *Trendkomponente* von Y , die Zeitreihe $S := (S_t : t \in \mathbb{Z})$ die *Saisonale Komponente* von Y und die Zeitreihe $X := (X_t : t \in \mathbb{Z})$ die *Stochastische Komponente* von Y nennt.

Anmerkung 6.1

Es kann neben der Existenz von Trend- und/oder Saisonalkomponenten auch noch einen anderen Grund für Nicht-Stationarität einer vorliegenden Realisierung einer Zeitreihe geben, nämlich, dass zufällige Ereignisse nicht-verschwindende Auswirkungen auf die Zukunft haben, oder anders ausgedrückt, dass das *AR*-Polynom der zugrunde liegenden *ARMA*-Zeitreihe eine Nullstelle besitzt, deren Betrag (sehr nahe bei) 1 ist.

In diesem Fall wird die Zeitreihe so lange durch Differenzenbildung bearbeitet, bis das Ergebnis stationär ist, man nennt solche Zeitreihen auch *ARIMA*-Zeitreihen. Um festzustellen, ob man nun mit dem Komponentenmodell oder mit dem *ARIMA*-Modell arbeiten sollte, existieren mehrere Testverfahren, die feststellen sollen, ob eben solch eine „*Unit Root*“ – also eine Nullstelle, deren Betrag genau (oder sehr nahe bei) 1 ist – existiert. Zu nennen wären hierbei u.a. der *Augmented Dickey-Fuller-Test* oder auch der *Kwiatkowski-Phillips-Schmidt-Shin-Test*. Auf diese wird hier aber aus Gründen des Umfangs der Arbeit nicht näher eingegangen werden, weshalb im Folgenden stets die Anwendbarkeit des Komponentenmodelles vorausgesetzt werden und die Möglichkeit der Existenz einer „*Unit Root*“ ignoriert werden wird.

Anmerkung 6.2

Der Sinn der Aufteilung einer Zeitreihe in ihre deterministischen und nicht-deterministischen Komponenten liegt darin begründet, dass, bereinigt man die Zeitreihe um die Trend- und die Saisonalkomponente, nur noch die stochastische Komponente übrig bleibt, die man meistens mittels

einer *ARMA*-Zeitreihe modellieren kann. Die Entscheidung, welches der beiden Komponentenmodelle auf eine vorliegende Realisierung einer Zeitreihe anwendbar ist, lässt sich recht einfach treffen: Man führt zunächst das additive Modell aus und betrachtet die nach der Bereinigung um die Trend- und die Saisonale Komponente entstandene stochastische Zeitreihe. Weist diese keine wachsende Varianz auf, so ist das additive Modell gut dafür geeignet, wächst oder sinkt die Varianz hingegen signifikant mit der Zeit, so sollte man das multiplikative Modell anwenden.

Durch Logarithmieren der vorliegenden Werte der ursprünglichen Realisierung der Zeitreihe erhält man, wendet man das multiplikative Modell an, $\log(Y_t) = \log(T_t) + \log(S_t) + \log(X_t)$, also wiederum ein additives Modell, wobei die Varianz von $\log(Y_t)$ konstant bleibt.

Zur Berechnung der Trend- und der Saisonalen Komponente existieren zahlreiche Möglichkeiten, von denen hier nur einige besonders wichtige exemplarisch angeführt werden.

6.1 Bestimmung der Trendkomponente

6.1.1 Polynomiale Regression mit der Methode der kleinsten Quadrate

Sei $\widehat{Y}_{[n]}$ eine Realisierung der Zeitreihe Y . Bei dieser Methode versucht man $\widehat{Y}_{[n]}$ mittels einer Polynomfunktion $T_t \approx f(t) := a_0 + a_1 t + a_2 t^2 + \dots + a_p t^p$ möglichst gut abzuschätzen. Dafür verwendet man die Methode der kleinsten Quadrate, das bedeutet, man sucht Koeffizienten $a_0, \dots, a_p$, sodass die Summe der quadrierten Abstände zwischen den Werten von f und $\widehat{Y}_{[n]}$ möglichst gering ist. Dieser quadratische Abstand ist nun für alle $t \in [1, n]$ offensichtlich gegeben durch $(\widehat{Y}_t - f(t))^2 = (\widehat{Y}_t - a_0 - a_1 t - \dots - a_p t^p)^2$ und die Summe der beobachteten Abstandsquadrate dementsprechend durch $d^2(a_0, \dots, a_p) = \sum_{t=1}^n (Y_t - a_0 - a_1 t - \dots - a_p t^p)^2$. Um das Minimum der Funktion d^2 zu erhalten differenziert man nun d^2 nach allen Variablen und setzt diese Ableitungen jeweils gleich 0:

$$\frac{\partial d^2}{\partial a_0} = \sum_{t=1}^n 2 \cdot (-1) \cdot (\widehat{Y}_t - a_0 - a_1 t - \dots - a_p t^p) = -2 \sum_{t=1}^n \widehat{Y}_t - a_0 - a_1 t - \dots - a_p t^p = 0$$

$$\frac{\partial d^2}{\partial a_1} = \sum_{t=1}^n 2 \cdot (-t) \cdot (\widehat{Y}_t - a_0 - a_1 t - \dots - a_p t^p) = -2 \sum_{t=1}^n t \cdot (\widehat{Y}_t - a_0 - a_1 t - \dots - a_p t^p) = 0$$

⋮

$$\frac{\partial d^2}{\partial a_p} = \sum_{t=1}^n 2 \cdot (-t^p) \cdot (\widehat{Y}_t - a_0 - a_1 t - \dots - a_p t^p) = -2 \sum_{t=1}^n t^p \cdot (\widehat{Y}_t - a_0 - a_1 t - \dots - a_p t^p) = 0 \quad .$$

Man erhält so das lineare Gleichungssystem

$$\begin{aligned}
 n \cdot a_0 + a_1 \left(\sum_{t=1}^n t \right) + a_2 \left(\sum_{t=1}^n t^2 \right) + \dots + \left(\sum_{t=1}^n t^p \right) &= \sum_{t=1}^n \hat{Y}_t \\
 a_0 \left(\sum_{t=1}^n t \right) + a_1 \left(\sum_{t=1}^n t^2 \right) + a_2 \left(\sum_{t=1}^n t^3 \right) + \dots + \left(\sum_{t=1}^n t^{p+1} \right) &= \sum_{t=1}^n t \cdot \hat{Y}_t \\
 \vdots \\
 a_0 \left(\sum_{t=1}^n t^p \right) + a_1 \left(\sum_{t=1}^n t^{p+1} \right) + a_2 \left(\sum_{t=1}^n t^{p+2} \right) + \dots + \left(\sum_{t=1}^n t^{2p} \right) &= \sum_{t=1}^n t^p \cdot Y_t .
 \end{aligned}$$

Durch Lösen dieses Gleichungssystems erhält man nun für geeignetes p eine Schätzung für die Trendkomponente T der Zeitreihe Y . Während es selbstverständlich möglich ist durch Erhöhung der Ordnung der gesuchten Polynomfunktion diese immer genauer an die tatsächlichen Werte anzupassen, sollte davon möglichst abgesehen werden, da hierbei stets die Gefahr des so genannten „Overfittings“ besteht – berechnet man im Extremfall für $\hat{Y}_{[n]}$ eine polynomiale Trendfunktion der Ordnung $n - 1$, so bleibt nach der Bereinigung der Zeitreihe um die Trendkomponente nur noch eine Zeitreihe übrig, deren Werte stets konstant 0 sind. Es ist bei der Trendbereinigung – und auch bei der Bereinigung um saisonale Anteile – also durchaus Fingerspitzengefühl gefragt.

6.1.2 Exponentielle Regression mit der Methode der kleinsten Quadrate

Bei manchen Zeitreihen ist es sinnvoll, deren Trend nicht als Polynomfunktion, sondern als Exponentialfunktion $T_t \approx f(t) := a \cdot e^{\lambda t}$ zu beschreiben. Man geht hierbei ähnlich vor wie bei einer Polynomfunktion ersten Grades, jedoch muss man die Exponentialfunktion erst einmal mittels Logarithmieren linearisieren. Es ergibt sich

$$\log[d^2(a, \lambda)] = \sum_{t=1}^n (\log(\hat{Y}_t) - \log(a) - \lambda t)^2 .$$

Um das Minimum dieser Funktion, die ja trivialerweise auch das Minimum der Funktion $d^2(a, \lambda)$ ist, zu erhalten leitet man wieder partiell nach a und nach λ ab und setzt die Ableitungen 0. Man erhält dadurch das lineare Gleichungssystem:

$$n \cdot \log(a) + \lambda \cdot \sum_{t=1}^n t = \sum_{t=1}^n \log(\hat{Y}_t)$$

$$\log(a) \cdot \left(\sum_{t=1}^n t \right) + \lambda \cdot \left(\sum_{t=1}^n t^2 \right) = \sum_{t=1}^n t \cdot \log(\hat{Y}_t)$$

Durch das Lösen dieses Gleichungssystems erhält man die Werte λ und $\log(a)$, woraus wiederum a errechnet werden kann. Abschließend muss man bei dieser Methode – im Gegensatz zur polynomialen Regression – die so erhaltene trendbereinigte Zeitreihe noch zentrieren, indem man von sämtlichen Werten der so erhaltenen Zeitreihenrealisierung $(\widehat{Y-T})_{[n]}$ deren empirischen Erwartungswert subtrahiert, man rechnet also folgendermaßen:

$$(\widehat{Y-T})'_t = (\widehat{Y-T})_t - \frac{1}{n} \cdot \sum_{i=1}^n (\widehat{Y-T})_i \quad \forall t \in [0,1]$$

6.2 Bestimmung der Saisonalen Komponente

Um die Saisonale Komponente S einer Zeitreihe Y abschätzen zu können muss man sich zunächst Gedanken über die Periodenlänge möglicher zyklischer Schwankungen machen. Diese sind meist entweder anhand des Graphen der bereits trendbereinigten Realisierung der Zeitreihe erkennbar, oder aber bereits durch die Häufigkeit der Aufzeichnungen – bei monatlicher Datenerfassung wird man meist die Periodenlänge 12, bei täglicher oftmals die Periodenlänge 7 wählen – ersichtlich.

Sei $(\widehat{Y-T})_{[n]}$ eine zentrierte Realisierung der trendbereinigten Zeitreihe $Y - T$, d die verwendete Periodenlänge, m die Anzahl der komplett beobachteten Perioden und $k := n - d \cdot m$, so definiert man für alle $q \in [1, d]$:

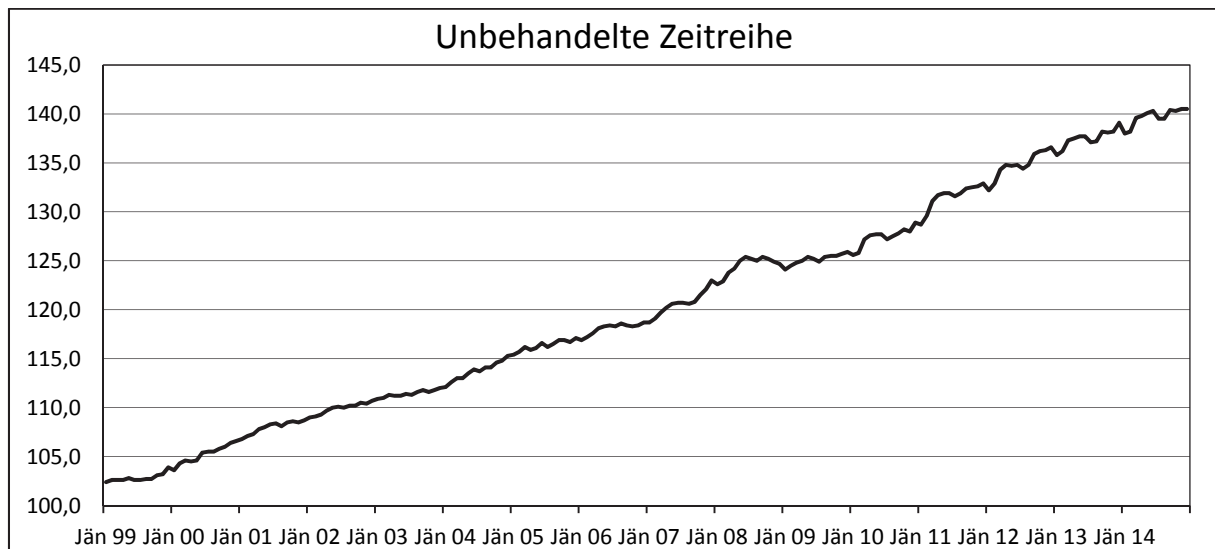
$$s_q := \frac{1}{m} \sum_{j=0}^{m-1} (\widehat{Y-T})_{k+q+j \cdot d}$$

Für alle $q \neq [1, d]$ hingegen definiert man $s_q := s_{m \cdot d + k} := s_k$

Man erhält somit eine Schätzung für den Saisonalen Anteil S mittels $S_t \approx s_t$.

Beispiel 6.1 Österreichischer Verbraucherpreisindex

Für dieses Beispiel wurde die Entwicklung des monatlichen österreichischen Verbraucherpreisindex mit der Basis des Jahres 1996 näher betrachtet. Der Datensatz hierfür wurde entnommen von http://statistik.at/web_de/statistiken/wirtschaft/preise/verbraucherpreisindex_vpi_hvpi/index.html. Es liegen monatliche Daten von Jänner 1999 bis Dezember 2014 vor, insgesamt also 192 Daten. Der Graph der unbehandelten Zeitreihe $\hat{Y}$ sieht wie folgt aus:



Ganz offensichtlich ist diese Zeitreihe nicht stationär, weshalb nun zunächst die Trendkomponente T errechnet werden soll. Nachdem man schon mit bloßem Auge sehen kann, dass es sich im Wesentlichen um einen linearen Trend handelt, ist es sinnvoll diesen auch als solchen zu modellieren. Zu lösen ist also das lineare Gleichungssystem

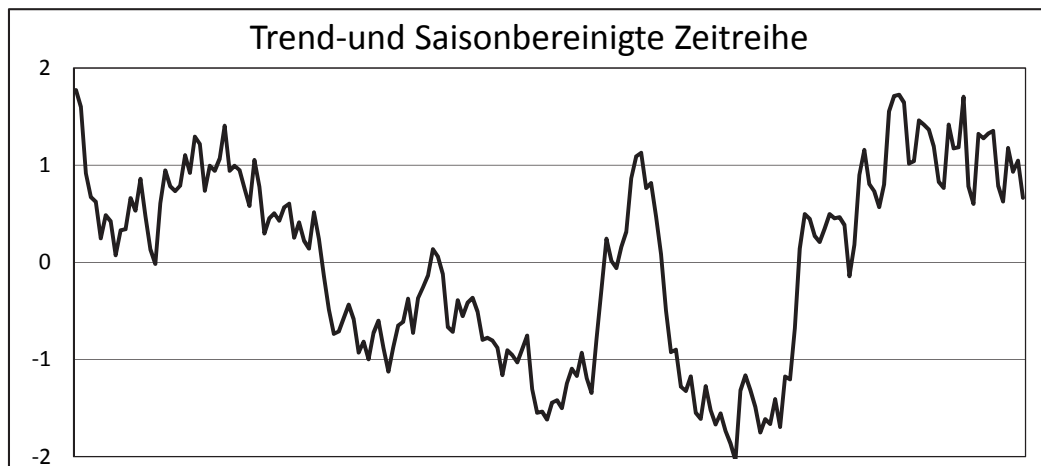
$$I: 192 \cdot a + a_1 \cdot \sum_{t=1}^{191} t = \sum_{t=1}^{191} \hat{Y}_t$$
$$II: a_0 \cdot \sum_{t=1}^{191} t + a_1 \cdot \sum_{t=1}^{191} t^2 = \sum_{t=1}^{191} t \cdot \hat{Y}_t$$

Man erhält so die lineare Funktion $f(t) = 101,0199 + 0,2033t \approx T_t$. Bereinigt man nun die ursprüngliche Zeitreihe $\hat{Y}_t$ von T , so erhält man die Zeitreihe $(\hat{Y}_t - T)$.

Nun muss noch die Saisonale Komponente S bestimmt werden. Nachdem die Aufzeichnungen monatlich erfolgen, wird die Periodendauer 12 gewählt. Man erhält so die Saisonalen Gewichte

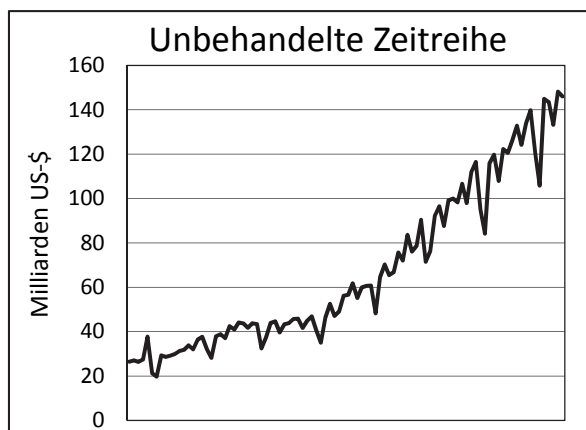
s_1	s_2	s_3	s_4	s_5	s_6	s_7	s_8	s_9	s_{10}	s_{11}	s_{12}
-0,39	-0,22	0,26	0,30	0,34	0,32	-0,13	-0,17	-0,02	-0,08	-0,19	-0,02

Bereinigt man nun die trendbereinigte Zeitreihe ($\hat{Y}_t - T$) noch von diesem Saisonalanteil, so erhält man eine Realisierung der Zeitreihe $X = Y - T - S$, die – wie es bei Anwendung der polynomialen Regression stets der Fall ist – zentriert ist und die man beispielsweise als *ARMA*-Zeitreihe modellieren kann.

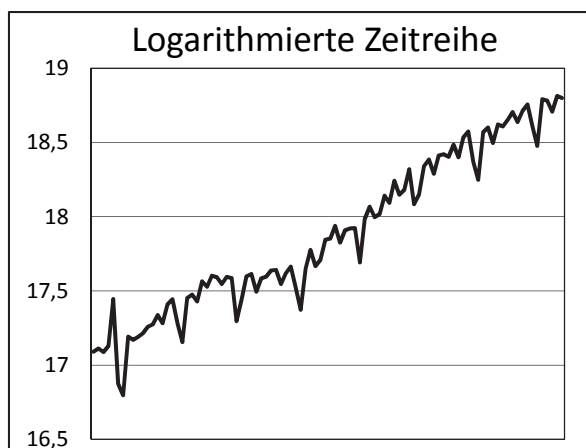


Beispiel 6.2 Außenhandelsvolumen der Volksrepublik China

Für dieses Beispiel wurde das monatlich zwischen Januar 1995 und Dezember 2012 ermittelte Gesamtvolumen der chinesischen Importe und Exporte betrachtet. Die Daten hierfür wurden entnommen von <http://data.stats.gov.cn/english>. Wie man leicht sehen kann ergibt sich bei diesem



Datensatz ein Problem: Die Varianz nimmt im Verlauf der beobachteten Zeitspanne offenbar zu. Deshalb empfiehlt es sich hier die gesamte Zeitreihe zu logarithmieren und in der Folge mit dem multiplikativen Modell zu arbeiten. Man erhält somit wiederum eine trend- und saisonbereinigte Zeitreihe, die man als stationären *ARMA*-Prozess modellieren kann.



7. Die empirische Autokorrelationsfunktion

Hat man nun eine stationäre und zentrierte Zeitreihe vorliegen, so kann man diese als Realisierung eines *ARMA*-Prozesses betrachten und dementsprechend verschiedene Kenngrößen dieses Prozesses abschätzen. Den hierzu benötigten Verfahren werden sich dieses und die kommenden Kapitel widmen.

Während es für die Behandlung von Zeitreihen allgemein sehr wichtig ist sich mit Schätzungen des Erwartungswertes zu beschäftigen, so kann man darauf in dieser Abhandlung über Modellierung mit kausalen *ARMA*-Zeitreihen durchaus verzichten, da diese ja ohnehin zentriert sind (siehe Korollar 4.1) und man daher vorliegende Realisierungen von Zeitreihen beispielweise mit den Methoden aus dem vorigen Kapitel soweit transformieren muss, dass die Annahme einer Zentriertheit der sich daraus ergebenden Zeitreihe sinnvoll ist.

Definition 7.1 Schätzfunktionen für *ARMA*-Zeitreihen

Sei $\widehat{X}_{[n]} := (\widehat{X}_t : t \in [1, n])$ eine Realisierung einer (unbekannten) stationären *ARMA*-Zeitreihe X , so bezeichnet man

- (i) die Schätzfunktion $\hat{\gamma}(h) := \frac{1}{n} \cdot \sum_{t=1}^{n-h} \widehat{X}_t \cdot \widehat{X}_{t+h}$ für $h \in [0, n-1]$ als empirische Autokovarianzfunktion
- (ii) die Schätzfunktion $\hat{\rho}(h) := \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}$ als empirische Autokorrelationsfunktion und
- (iii) die Schätzfunktion, die sich ergibt, wenn man sämtliche Autokorrelationen in der Berechnungsmethode der partiellen Autokorrelation (siehe Satz 2.1) durch deren empirische Gegenstücke ersetzt, als empirische partielle Autokorrelationsfunktion $\hat{\pi}(h)$.

Korollar 7.1

Die empirische Autokovarianzfunktion ist keine erwartungstreue Schätzfunktion für die Autokovarianzfunktion.

Beweis:

$$\mathbb{E}[\hat{\gamma}(h)] = \frac{1}{n} \cdot \sum_{t=1}^{n-h} \mathbb{E}(\widehat{X}_t \cdot \widehat{X}_{t+h}) = \frac{1}{n} \cdot \sum_{t=1}^{n-h} \mathbb{E}(X_t \cdot X_{t+h}) = \frac{1}{n} \cdot (n-h) \cdot \gamma(h) = \gamma(h) - \frac{h}{n} \cdot \gamma(h) \quad .$$

■

Anmerkung 7.1

Wie man sieht, ist lediglich der Schätzwert $\hat{\gamma}(0)$ erwartungstreu bzgl. $\gamma(0)$. Es stellt sich nun natürlich die Frage, weshalb man einen nicht erwartungstreuen Schätzer verwendet, obwohl man sofort sieht, dass die Schätzfunktion $\frac{1}{n-h} \cdot \sum_{t=1}^{n-h} \mathbb{E}(\widehat{X}_t \cdot \widehat{X}_{t+h})$ das Kriterium der Erwartungstreue erfüllen würde.

Wie in den folgenden Ausführungen ersichtlich werden wird, bringt die Verwendung der nicht erwartungstreuen empirischen Autokovarianzfunktion große Vorteile mit sich, die diesen „Makel“ mehr als wettmachen. Man sollte jedoch immer im Hinterkopf behalten, dass die Verwendung dieser Schätzfunktion nur dann sinnvoll ist, wenn der Umfang einer Realisierung ausreichend groß und die gesuchten Lags ausreichend klein sind, sodass der „Fehlerterm“ $\frac{h}{n} \cdot \gamma(h)$ weitgehend unbedeutend wird. Trivialerweise ist nämlich für jedes beliebige fest gewählte $k \geq 0$ die Schätzfunktion $\hat{\gamma}(h)$ asymptotisch erwartungstreu für $\gamma(0), \dots, \gamma(k)$.

In den folgenden Abhandlungen wird nun eine wichtige Eigenschaft der empirischen Autokorrelationsfunktion bewiesen, nämlich, dass die Wahl des Faktors $\frac{1}{n}$ anstelle von $\frac{1}{n-h}$ garantiert, dass die empirische Autokovarianzmatrix – und damit natürlich auch die empirische Autokorrelationsmatrix – positiv definit ist, woraus, wie man noch sehen wird, folgt, dass die empirischen partiellen Autokorrelationen immer eindeutig bestimmt werden können.

Lemma 7.1

Sei $\widehat{X}_{[n]} := (\widehat{X}_t : t \in [1, n])$ eine Realisierung einer (unbekannten) stationären *ARMA*-Zeitreihe X , so gilt

$$\hat{\gamma}(0) = 0 \Leftrightarrow \hat{X}_1 = \hat{X}_2 = \dots = \hat{X}_n = 0 \quad .$$

Beweis:

$$\hat{\gamma}(0) = 0 \Leftrightarrow \frac{1}{n} \cdot \sum_{t=1}^n \hat{X}_t^2 = 0 \Leftrightarrow \sum_{t=1}^n \underbrace{\hat{X}_t^2}_{\geq 0} = 0 \Leftrightarrow \hat{X}_t = 0 \quad \forall t \in [1, n] \quad .$$

■

Proposition 7.1

Sei $\widehat{X}_{[n]} := (\widehat{X}_t : t \in [1, n])$ eine Realisierung einer (unbekannten) stationären *ARMA*-Zeitreihe X , so gilt:

Die Matrix $\mathbf{A} := \begin{pmatrix} \hat{\gamma}(0) & \hat{\gamma}(1) & \dots & \hat{\gamma}(n-1) \\ \hat{\gamma}(1) & \hat{\gamma}(0) & & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\gamma}(n-1) & \dots & \dots & \hat{\gamma}(0) \end{pmatrix}$ ist positiv semidefinit.

Beweis:

$$\text{Es gilt: } \mathbf{A} = \frac{1}{n} \cdot \begin{pmatrix} \sum_{t=1}^n \hat{X}_t^2 & \sum_{t=1}^{n-1} X_t \cdot X_{t+1} & \dots & X_1 \cdot X_n \\ \sum_{t=1}^{n-1} X_t \cdot X_{t+1} & \sum_{t=1}^n \hat{X}_t^2 & & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ X_1 \cdot X_n & \dots & \dots & \sum_{t=1}^n \hat{X}_t^2 \end{pmatrix}$$

Dies kann man alternativ schreiben durch $\mathbf{A} = \frac{1}{n} \cdot \mathbf{B} \cdot \mathbf{B}^T$, wobei $\mathbf{B}$ eine $n \times 2n$ -Matrix ist,

$$\text{gegeben durch } \mathbf{B} := \begin{pmatrix} 0 & \dots & 0 & 0 & \hat{X}_1 & \hat{X}_2 & \dots & \hat{X}_n \\ 0 & \dots & 0 & \hat{X}_1 & \hat{X}_2 & \dots & \hat{X}_n & 0 \\ \vdots & & & & & & & \vdots \\ 0 & \hat{X}_1 & \hat{X}_2 & \dots & \hat{X}_n & 0 & 0 & 0 \end{pmatrix}.$$

Sei $\mathbf{v}$ ein beliebiger reellwertiger $1 \times n$ -Vektor, der nicht der Nullvektor ist, so gilt

$$\begin{aligned} \mathbf{v} \cdot \mathbf{A} \cdot \mathbf{v}^T &= \mathbf{v} \cdot \frac{1}{n} \cdot \mathbf{B} \cdot \mathbf{B}^T \cdot \mathbf{v}^T = \frac{1}{n} \cdot (\mathbf{v} \cdot \mathbf{B}) \cdot (\mathbf{B}^T \cdot \mathbf{v}^T) = \frac{1}{n} \cdot (\mathbf{v} \cdot \mathbf{B}) \cdot (\mathbf{v} \cdot \mathbf{B})^T = \frac{1}{n} \cdot \langle \mathbf{v} \cdot \mathbf{B}, \mathbf{v} \cdot \mathbf{B} \rangle = \\ &= \frac{1}{n} \cdot \|\mathbf{v} \cdot \mathbf{B}\|^2 \geq 0. \end{aligned}$$

■

Proposition 7.2

Sei $\widehat{X}_{[n]} := (\widehat{X}_t : t \in [1, n])$ eine Realisierung einer (unbekannten) stationären *ARMA*-Zeitreihe X , so ist für $\hat{\gamma}(0) \neq 0$ die Matrix $\mathbf{A}$ aus Prop.7.1 positiv definit.

Beweis:

Indirekt angenommen die Matrix $\mathbf{A}$ wäre nicht positiv definit (aber nach Prop.7.1 positiv semidefinit), dann müsste es (mindestens) einen $1 \times n$ -Vektor $\mathbf{g} = (g_1, \dots, g_n)$ ungleich dem Nullvektor geben, sodass $\frac{1}{n} \cdot \|\mathbf{g} \cdot \mathbf{B}\|^2 = 0$ gilt, wobei $\mathbf{B}$ gleich definiert ist wie in Prop.7.1. Bekanntermaßen kann dies nur der Fall sein, wenn $(\mathbf{g} \cdot \mathbf{B})$ der Nullvektor ist, das heißt es muss gelten, dass

$$(g_1, \dots, g_n) \cdot \begin{pmatrix} 0 & \dots & 0 & 0 & \hat{X}_1 & \hat{X}_2 & \dots & \hat{X}_n \\ 0 & \dots & 0 & \hat{X}_1 & \hat{X}_2 & \dots & \hat{X}_n & 0 \\ \vdots & & & & & & & \vdots \\ 0 & \hat{X}_1 & \hat{X}_2 & \dots & \hat{X}_n & 0 & \dots & 0 \end{pmatrix} = (0, 0, \dots, 0).$$

Man erhält so die Gleichungen $g_n \hat{X}_1 = 0$; $g_{n-1} \hat{X}_1 + g_n \hat{X}_2 = 0$; ...; $g_1 \hat{X}_1 + g_2 \hat{X}_2 + \dots + g_n \hat{X}_n = 0$; $g_1 \hat{X}_2 + \dots + g_{n-1} \hat{X}_n = 0$; ...; $g_1 \hat{X}_n = 0$ und dadurch das lineare Gleichungssystem $\mathbf{C} \cdot \hat{\mathbf{x}} = 0$,

$$\text{gegeben durch } \left(\begin{array}{ccccccc|c} g_n & 0 & \dots & \dots & \dots & \dots & 0 & 0 \\ g_{n-1} & g_n & 0 & \dots & \dots & \dots & 0 & 0 \\ \vdots & & \vdots & & \vdots & & \vdots & \vdots \\ g_1 & g_2 & \dots & \dots & \dots & \dots & g_n & \vdots \\ 0 & g_1 & \dots & \dots & \dots & \dots & g_{n-1} & \vdots \\ \vdots & & \vdots & & \vdots & & \vdots & \vdots \\ 0 & \dots & \dots & \dots & \dots & \dots & g_1 & 0 \end{array} \right) \text{ und } \hat{\mathbf{x}} = \begin{pmatrix} \hat{X}_1 \\ \vdots \\ \hat{X}_n \end{pmatrix}.$$

Bekanntermaßen ist ein lineares Gleichungssystem $\mathbf{D} \cdot \mathbf{x} = \mathbf{b}$ genau dann eindeutig lösbar, wenn der Rang der $m \times n$ -Matrix $\mathbf{D}$ und der Rang der um $\mathbf{b}$ erweiterten Matrix $(\mathbf{D}, \mathbf{b})$ gleich n ist. Nachdem $\mathbf{b}$ hier ein Nullvektor ist, durch dessen Hinzufügen sich der Rang einer Matrix nicht verändert, genügt es den Rang der Matrix $\mathbf{C}$ zu bestimmen.

Wie bekannt sein dürfte gilt, dass der Rang der Matrix $\mathbf{C}$ derselbe der Matrix $\mathbf{C}^T$ ist und sich dieser durch Zeilenvertauschungen nicht verändert, d.h. es gilt

$$\text{rang}(\mathbf{C}) = \text{rang} \begin{pmatrix} g_n & g_{n-1} & \cdots & g_1 & 0 & \cdots & \cdots & 0 \\ 0 & g_n & \cdots & g_2 & g_1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & g_n & g_{n-1} & \cdots & \cdots & g_1 \end{pmatrix}.$$

Man weiß nun, dass der Rang dieser $n \times 2n$ - Matrix höchstens n sein kann.

Indirekt angenommen der Rang der Matrix wäre kleiner als n . Für $g_n \neq 0$ sieht man sofort, dass sämtliche Zeilenvektoren linear unabhängig sein müssen, der Rang der Matrix also n wäre. g_n muss also gleich 0 sein. Dasselbe gilt nun für $g_{n-1}, g_{n-2}, \dots, g_1$ was bedeutet, dass der Vektor g der Nullvektor sein muss, was einen Widerspruch darstellt. Somit ist zunächst bewiesen, dass der Rang der Matrix $\mathbf{C}$ gleich n ist und das Gleichungssystem $\mathbf{C} \cdot \hat{\mathbf{x}} = \mathbf{0}$ eindeutig lösbar ist.

Offenbar ist nun der Vektor $\hat{\mathbf{x}} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}$ eine, und damit eben auch die einzige Lösung des

Gleichungssystems. Daraus folgt nach Lemma 7.1, dass $\hat{\gamma}(0)$ gleich 0 sein muss, was einen Widerspruch zur Voraussetzung darstellt.

■

Satz 7.1

Sei $\widehat{X}_{[n]} := (\widehat{X}_t : t \in [1, n])$ eine Realisierung einer (unbekannten) stationären $ARMA$ -Zeitreihe X und seien nicht alle Werte der Realisierung gleich 0, so lässt sich die empirische partielle Autokorrelation $\hat{\pi}(h)$ für alle Lags h mit $h < n$ eindeutig bestimmen.

Beweis:

Aus Lemma 7.1 folgt, dass $\hat{\gamma}(0)$ ungleich Null sein muss. Trivialerweise lassen sich für alle h mit $0 \leq h < n$ die Ergebnisse von Prop. 7.1 und Prop 7.2 auch auf die Matrizen

$$\mathbf{A}_h := \frac{1}{\hat{\gamma}(0)} \cdot \begin{pmatrix} \hat{\gamma}(0) & \hat{\gamma}(1) & \cdots & \hat{\gamma}(h-1) \\ \hat{\gamma}(1) & \hat{\gamma}(0) & & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\gamma}(h-1) & \cdots & \cdots & \hat{\gamma}(0) \end{pmatrix} \text{ übertragen.}$$

Aufgrund der positiven Definitheit von $\mathbf{A}_h$ muss für all ihre Eigenwerte λ_i gelten, dass $\lambda_1, \dots, \lambda_h > 0$.

Da es sich um eine reelle, symmetrische Matrix handelt, muss außerdem gelten, dass

$\det(\mathbf{A}_h) = \underbrace{\lambda_1}_{>0} \cdot \dots \cdot \underbrace{\lambda_h}_{>0} > 0$. Somit ist jedes lineare Gleichungssystem $\mathbf{A}_h \cdot \mathbf{x} = \mathbf{b}$ eindeutig lösbar.

■

Anmerkung 7.2

Die zweite, in der Praxis noch wesentlich wichtigere Eigenschaft der hier verwendeten Schätzfunktionen wird im folgenden Satz deutlich, der aussagt, dass unter bestimmten Bedingungen die empirische und auch die empirische partielle Autokorrelationsfunktion approximativ multivariat normalverteilt sind. Aufgrund des großen Umfangs der zu diesem Satz gehörigen Beweisführung wird diese hier ausnahmsweise nicht angeführt werden. Interessierte LeserInnen seien beispielsweise auf Peter J. Brockwell, Richard A. Davis, Time Series: Theory and Methods (Springer, 1991) S.221-236, 241, 262-264 verwiesen.

Satz 7.2

Sei $\widehat{X}_{[n]} := (\widehat{X}_t: t \in [1, n])$ eine Realisierung einer (unbekannten) kausalen *ARMA*-Zeitreihe X und seien die Zufallsvariablen des zugrundeliegenden Weißen Rauschens e identisch verteilt und paarweise stochastisch unabhängig, dann gilt für alle $h > 0$, dass der Zufallsvektor $(\hat{\rho}(1), \dots, \hat{\rho}(h))^T$ für großes n approximativ multivariat normalverteilt ist mit Erwartungswertvektor $(\rho(1), \dots, \rho(h))^T$ und Kovarianzmatrix $\frac{1}{n} \cdot \mathbf{M}$, wobei die Elemente $m_{i,k}$ von $\mathbf{M}$ bestimmt sind durch die Gleichung

$$m_{i,k} = \sum_{r=-\infty}^{\infty} [\rho(r+i) \cdot \rho(r+k) + \rho(r-i) \cdot \rho(r+k) + 2\rho(i)\rho(k)\rho^2(r) - 2\rho(i)\rho(r)\rho(r+k) - 2\rho(k)\rho(r)\rho(r+i)] .$$

Handelt es sich bei X außerdem um eine $AR(p)$ -Zeitreihe, so ist für $h > p$ der Wert der empirischen partiellen Autokorrelationsfunktion beim Lag h für große n approximativ normalverteilt mit Erwartungswert 0 und Varianz $\frac{1}{n}$.

(kein Beweis)

■

8. Parameterschätzung in ARMA-Modellen

Hat man eine Realisierung einer Zeitreihe vorliegen, diese trend- und saisonbereinigt und hat man deren empirische Autokorrelationsfunktion errechnet, so kann man daran gehen für festgelegte Parameter p und q ein für diese Realisierung möglichst geeignetes $ARMA(p, q)$ -Modell zu erstellen. Wie man bei der Wahl von p und q vorgehen soll wird Thema des nächsten Kapitels sein – hier wird zunächst angenommen, man wüsste bereits die Ordnungen des der Realisierung optimal entsprechenden *ARMA*-Modells.

Obgleich für die Parameterschätzung für kausale *ARMA*-Zeitreihen mehrere Methoden existieren, wird hier lediglich die Methode der *Maximum-Likelihood-Schätzung* näher beleuchtet werden, da diese, obwohl sie durchaus als sehr rechenaufwändig bezeichnet werden kann, meist sehr gute Ergebnisse liefert und in vielerlei Hinsicht ohnehin äquivalent zu der qualitativ gleichwertigen Methode der Parameterschätzung mittels der *Methode der kleinsten Quadrate* ist.

Definition 8.1 Ein-Schritt-Vorhersage

Sei X eine beliebige zentrierte Zeitreihe und sei $H_n := \text{span}\{X_1, \dots, X_n\}$ für $n \geq 1$, so nennt man

$$\tilde{X}_{n+1} := P_{H_n}(X_{n+1}) = d_{n,1}X_n + \dots + d_{n,n}X_1$$

die *Ein-Schritt-Vorhersage* für X_{n+1} und definiert $\tilde{X}_1 := 0$.

Sei $\widehat{X}_{[n]}$ eine Realisierung von X vom Umfang n und sei für alle $1 \leq j \leq n+1$

$\bar{\bar{X}}_j$ jener Wert, den man durch Einsetzen von $\hat{X}_1 = X_1, \dots, \hat{X}_{j-1} = X_{j-1}$ in die Ein-Schritt-Vorhersage $\tilde{X}_j$ von X_j erhält, so nennt man $\bar{\bar{X}}_j$ die *empirische Ein-Schritt-Vorhersage* von $\hat{X}_j$ und setzt wiederum $\bar{\bar{X}}_1 := 0$.

Proposition 8.1

Sei X eine zentrierte, stationäre Zeitreihe, so ergeben sich für alle $n \geq 1$ die Koeffizienten $d_{n,1}, \dots, d_{n,n}$ der Ein-Schritt-Vorhersage von X_{n+1} als Lösung des linearen Gleichungssystems

$$\begin{pmatrix} \gamma(0) & \gamma(1) & \dots & \gamma(n-1) \\ \gamma(1) & \gamma(0) & & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \gamma(n-1) & \dots & \dots & \gamma(0) \end{pmatrix} \cdot \begin{pmatrix} d_{n,1} \\ d_{n,2} \\ \vdots \\ d_{n,n} \end{pmatrix} = \begin{pmatrix} \gamma(1) \\ \gamma(2) \\ \vdots \\ \gamma(n) \end{pmatrix},$$

sofern dieses eindeutig lösbar ist. Dabei entspricht der Koeffizient $d_{n,n}$ der partiellen Autokorrelation von X beim Lag n .

Beweis:

Der Beweis erfolgt weitgehend analog zu jenem von Satz 2.1. ■

Definition 8.2 Ein-Schritt-Vorhersagefehler

Sei X eine beliebige zentrierte Zeitreihe und seien $\tilde{X}_{n+1}$ für $n \geq 0$ die Ein-Schritt-Vorhersagen für X_{n+1} , so nennt man

$$v_n := \|X_{n+1} - \tilde{X}_{n+1}\|^2$$

den *Ein-Schritt-Vorhersagefehler* der Zeitreihe für X_{n+1} .

Satz 8.1

Sei X eine zentrierte, stationäre Zeitreihe und sei A invertierbar, wobei

$$A := \begin{pmatrix} \gamma(0) & \gamma(1) & \dots & \gamma(n-1) \\ \gamma(1) & \gamma(0) & & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \gamma(n-1) & \dots & \dots & \gamma(0) \end{pmatrix}, \mathbf{d}_n := \begin{pmatrix} d_{n,1} \\ d_{n,2} \\ \vdots \\ d_{n,n} \end{pmatrix} \text{ und } \mathbf{r}_n := \begin{pmatrix} \gamma(1) \\ \gamma(2) \\ \vdots \\ \gamma(n) \end{pmatrix},$$

so sind für $n \geq 0$ die Ein-Schritt-Vorhersagefehler v_n bestimmt durch

$$v_n = \gamma(0) - d_{n,1} \cdot \gamma(1) - \dots - d_{n,n} \cdot \gamma(n) ,$$

bzw. durch die äquivalente Gleichung

$$v_n = \gamma(0) - \boldsymbol{\gamma}_n^T \cdot \mathbf{A}^{-1} \cdot \boldsymbol{\gamma}_n .$$

Beweis:

Nach Prop.8.1 gilt: $\mathbf{A} \cdot \mathbf{d}_n = \boldsymbol{\gamma}_n \Leftrightarrow \mathbf{A}^{-1} \cdot \mathbf{A} \cdot \mathbf{d}_n = \mathbf{A}^{-1} \cdot \boldsymbol{\gamma}_n \Leftrightarrow \mathbf{d}_n = \mathbf{A}^{-1} \cdot \boldsymbol{\gamma}_n$ und somit, weil das Inverse symmetrischer Matrizen stets symmetrisch ist: $\mathbf{d}_n^T = \boldsymbol{\gamma}_n^T \cdot \mathbf{A}^{-1}$. Es gilt nun, dass

$$\begin{aligned} v_n &:= \|X_{n+1} - \tilde{X}_{n+1}\|^2 = \mathbb{E}[(X_{n+1} - \tilde{X}_{n+1})^2] = \mathbb{E}(X_{n+1}^2 - 2X_{n+1} \cdot \tilde{X}_{n+1} + \tilde{X}_{n+1}^2) = \\ &= \gamma(0) - 2\mathbb{E}(X_{n+1} \cdot \tilde{X}_{n+1}) + \mathbb{E}(\tilde{X}_{n+1}^2) = \gamma(0) - 2\mathbb{E}\left[\mathbf{d}_n^T \cdot \begin{pmatrix} X_n \\ \vdots \\ X_1 \end{pmatrix}\right] + \mathbb{E}\left(\left[\tilde{X}_{n+1} - \underbrace{\mathbb{E}(\tilde{X}_{n+1})}_{=0}\right]^2\right) = \\ &= \gamma(0) - 2 \cdot \mathbf{d}_n^T \cdot \boldsymbol{\gamma}_n + \text{Var}(\tilde{X}_{n+1}) = \gamma(0) - 2 \cdot \boldsymbol{\gamma}_n^T \cdot \mathbf{A}^{-1} \cdot \boldsymbol{\gamma}_n + \text{Var}\left[\mathbf{d}_n \cdot \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}\right] = \\ &= \gamma(0) - 2 \cdot \boldsymbol{\gamma}_n^T \cdot \mathbf{A}^{-1} \cdot \boldsymbol{\gamma}_n + \sum_{i,j=1}^n \text{Cov}(d_{n,i} \cdot X_{n-i+1}, d_{n,j} \cdot X_{n-j+1}) = \\ &= \gamma(0) - 2 \cdot \boldsymbol{\gamma}_n^T \cdot \mathbf{A}^{-1} \cdot \boldsymbol{\gamma}_n + \mathbf{d}_n \cdot \mathbf{A} \cdot \mathbf{d}_n^T = \gamma(0) - 2 \cdot \boldsymbol{\gamma}_n^T \cdot \mathbf{A}^{-1} \cdot \boldsymbol{\gamma}_n + \boldsymbol{\gamma}_n^T \cdot \mathbf{A}^{-1} \cdot \mathbf{A} \cdot \mathbf{A}^{-1} \cdot \boldsymbol{\gamma}_n = \\ &= \gamma(0) - 2 \cdot \boldsymbol{\gamma}_n^T \cdot \mathbf{A}^{-1} \cdot \boldsymbol{\gamma}_n + \boldsymbol{\gamma}_n^T \cdot \mathbf{A}^{-1} \cdot \boldsymbol{\gamma}_n = \gamma(0) - \boldsymbol{\gamma}_n^T \cdot \mathbf{A}^{-1} \cdot \boldsymbol{\gamma}_n . \end{aligned}$$

Außerdem gilt:

$$v_n = \gamma(0) - \boldsymbol{\gamma}_n^T \cdot \mathbf{A}^{-1} \cdot \boldsymbol{\gamma}_n = \gamma(0) - \mathbf{d}_n \cdot \boldsymbol{\gamma}_n = \gamma(0) - d_{n,1} \cdot \gamma(1) - \dots - d_{n,n} \cdot \gamma(n) .$$

■

Anmerkung 8.1

Obwohl es natürlich möglich ist – zumindest für stationäre Zeitreihen – mittels Prop. 8.1 die Koeffizienten $d_{n,1}, \dots, d_{n,n}$ und damit dann schlussendlich die Ein-Schritt-Vorhersage $\tilde{X}_{n+1}$ zu berechnen, ist es unumgänglich eine weitere, rekursive Berechnungsmethode vorzustellen, die des so genannten Innovationsalgorithmus. Neben dessen Wichtigkeit für die Beweisführung im Kontext der *Maximum-Likelihood*-Schätzer, wie man noch später sehen wird, hat dies vor allem praktische Gründe.

Gerade für große n ist nämlich die Berechnung der Koeffizienten $d_{n,1}, \dots, d_{n,n}$ sehr aufwändig, da hierfür ein sehr umfangreiches lineares Gleichungssystemen gelöst werden muss. In der praktischen Anwendung muss dies für die Bestimmung empirischer Ein-Schritt-Vorhersagen bei Beobachtung eines neuen Wertes außerdem stets wiederholt werden, was äußerst unpraktisch ist.

Der Innovationsalgorithmus, der nun beschrieben werden wird, verwendet hingegen in diesem Fall die bereits berechneten Koeffizienten $d_{n,1}, \dots, d_{n,n}$ direkt für die Bestimmung der Koeffizienten $d_{n+1,1}, \dots, d_{n+1,n+1}$, was ihn der „naiven“ Berechnungsmethode aus Prop. 8.1 deutlich überlegen macht.

Lemma 8.1

Sei X eine beliebige zentrierte Zeitreihe. Gelte für $n \geq 1$, dass $H_n := \text{span}\{X_1, \dots, X_n\}$ und seien die Ein-Schritt-Vorhersagefehler v_n stets ungleich Null, so gilt:

- (i) $H_n = \text{span}\{X_1 - \tilde{X}_1, \dots, X_n - \tilde{X}_n\}$
- (ii) $\forall m \in \mathbb{N}, 1 \leq m \leq n \exists f_{m,1}, \dots, f_{m,m} \in \mathbb{R}$, sodass gilt, dass

$$\tilde{X}_{m+1} = f_{m,1} \cdot (X_m - X_m) + \dots + f_{m,m} \cdot (X_1 - X_1) .$$

Beweis:

- (i) Nachdem die Ein-Schritt-Vorhersagefehler v_i für alle $i \in [0, n-1]$ ungleich Null sind, lässt sich keine der Zufallsvariablen $X_2, \dots, X_n$ vollständig als Linearkombination der vorangegangenen Zufallsvariablen darstellen, also sind $X_1, \dots, X_n$ linear unabhängig, woraus folgt, dass die Zufallsvariablen $X_1, \dots, X_n$ eine Basis des von ihnen erzeugten Raumes H_n bilden. Trivialerweise gilt nun zunächst, dass $X_1 - \tilde{X}_1 = X_1 - 0 = X_1 \in H_n$. Nachdem für alle $1 < i \leq n$ gilt, dass $\tilde{X}_i := P_{H_{i-1}}(X_i)$, liegen selbstverständlich diese $\tilde{X}_i$ und somit auch die einzelnen $X_i - \tilde{X}_i$ in H_n . Zusätzlich sind diese $X_i - \tilde{X}_i$ aufgrund des Projektionssatzes jeweils orthogonal auf den Raum H_{i-1} , woraus induktiv folgt, dass für $1 \leq j \leq n$ alle Zufallsvariablen $X_j - \tilde{X}_j$ orthogonal zueinander und damit trivialerweise auch linear unabhängig sind. Somit bilden also die Zufallsvariablen $X_j - \tilde{X}_j$ eine Basis des Raumes H_n .

- (ii) Trivial. ■

Satz 8.1 Der Innovations-Algorithmus

Sei X eine beliebige zentrierte Zeitreihe, deren Ein-Schritt-Vorhersagefehler stets ungleich Null sind. Sei $m \geq 1$ und sei die Kovarianzmatrix des Zufallsvektors $(X_1, \dots, X_m)^T$ invertierbar, so lassen sich für $0 \leq n \leq m$ die Ein-Schritt-Vorhersagen $\tilde{X}_{n+1}$ und ihre Vorhersagefehler v_n rekursiv berechnen mittels

- $$\tilde{X}_{n+1} = \sum_{j=1}^n f_{n,j} \cdot (X_{n-j+1} - \tilde{X}_{n-j+1})$$
- $$v_0 = \text{Cov}(X_1, X_1)$$
- $$f_{n,n-k} = \frac{1}{v_k} \cdot \left(\text{Cov}(X_{n+1}, X_{k+1}) - \sum_{j=0}^{k-1} f_{k,k-j} \cdot f_{n,n-j} \cdot v_j \right) \quad \forall k \in [0, n-1]$$
- $$v_n = \text{Cov}(X_{n+1}, X_{n+1}) - \sum_{j=0}^{n-1} f_{n,n-j}^2 \cdot v_j .$$

Beweis:

$\tilde{X}_{n+1} = \sum_{j=1}^n f_{n,j} \cdot (X_{n-j+1} - \tilde{X}_{n-j+1})$ gilt genau dann, wenn für alle k mit $0 \leq k < n$ gilt:

$$\begin{aligned} \langle \tilde{X}_{n+1}, X_{k+1} - \tilde{X}_{k+1} \rangle &= \left\langle \sum_{j=1}^n f_{n,j} \cdot (X_{n-j+1} - \tilde{X}_{n-j+1}), X_{k+1} - \tilde{X}_{k+1} \right\rangle = \\ &= \sum_{j=1}^n f_{n,j} \cdot \underbrace{\langle X_{n-j+1} - \tilde{X}_{n-j+1}, X_{k+1} - \tilde{X}_{k+1} \rangle}_{=0 \ \forall j \neq n-k} = f_{n,n-k} \cdot \langle X_{k+1} - \tilde{X}_{k+1}, X_{k+1} - \tilde{X}_{k+1} \rangle = \\ &= f_{n,n-k} \cdot v_k. \end{aligned}$$

Dies ist nun äquivalent zu Folgendem:

$$\begin{aligned} \langle X_{n+1}, X_{k+1} - \tilde{X}_{k+1} \rangle - f_{n,n-k} \cdot v_k &= \langle X_{n+1}, X_{k+1} - \tilde{X}_{k+1} \rangle - \langle \tilde{X}_{n+1}, X_{k+1} - \tilde{X}_{k+1} \rangle = \\ &= \langle X_{n+1} - \tilde{X}_{n+1}, X_{k+1} - \tilde{X}_{k+1} \rangle = 0. \end{aligned}$$

Somit ergibt sich

$$\begin{aligned} f_{n,n-k} &= \frac{1}{v_k} \cdot \langle X_{n+1}, X_{k+1} - \tilde{X}_{k+1} \rangle = \frac{1}{v_k} \cdot \langle X_{n+1}, X_{k+1} - \sum_{j=1}^k f_{k,j} \cdot (X_{k-j+1} - \tilde{X}_{k-j+1}) \rangle = \\ &= \frac{1}{v_k} \cdot \left(\langle X_{n+1}, X_{k+1} \rangle - \sum_{j=1}^k f_{k,j} \cdot \langle X_{n+1}, X_{k-j+1} - \tilde{X}_{k-j+1} \rangle \right) = \\ &= \frac{1}{v_k} \cdot \left(\text{Cov}(X_{n+1}, X_{k+1}) - \sum_{j=0}^{k-1} f_{k,k-j} \cdot \underbrace{\langle X_{n+1}, X_{j+1} - \tilde{X}_{j+1} \rangle}_{=f_{n,n-j} \cdot v_j} \right) = \\ &= \frac{1}{v_k} \cdot \left(\text{Cov}(X_{n+1}, X_{k+1}) - \sum_{j=0}^{k-1} f_{k,k-j} \cdot f_{n,n-j} \cdot v_j \right). \end{aligned}$$

Aus dem Projektionssatz folgt nun, dass $X_{n+1} - \tilde{X}_{n+1} \perp \tilde{X}_{n+1}$. Man erhält somit mittels des Satzes von Pythagoras $\|X_{n+1} - \tilde{X}_{n+1}\|^2 = \|X_{n+1}\|^2 - \|\tilde{X}_{n+1}\|^2$. Es gilt daher:

$$\begin{aligned} v_n &= \|X_{n+1} - \tilde{X}_{n+1}\|^2 = \|X_{n+1}\|^2 - \|\tilde{X}_{n+1}\|^2 = \text{Cov}(X_{n+1}, X_{n+1}) - \langle \tilde{X}_{n+1}, \tilde{X}_{n+1} \rangle = \\ &= \text{Cov}(X_{n+1}, X_{n+1}) - \left\langle \sum_{j=0}^{n-1} f_{n,n-j} \cdot (X_{j+1} - \tilde{X}_{j+1}), \sum_{j=0}^{n-1} f_{n,n-j} \cdot (X_{j+1} - \tilde{X}_{j+1}) \right\rangle = \\ &= \text{Cov}(X_{n+1}, X_{n+1}) - \sum_{j=0}^{n-1} f_{n,n-j} \cdot \langle X_{j+1} - \tilde{X}_{j+1}, X_{j+1} - \tilde{X}_{j+1} \rangle - \\ &\quad - \sum_{i \neq j} f_{n,n-i} \cdot f_{n,n-j} \cdot \underbrace{\langle X_{i+1} - \tilde{X}_{i+1}, X_{j+1} - \tilde{X}_{j+1} \rangle}_{=0} = \\ &= \text{Cov}(X_{n+1}, X_{n+1}) - \sum_{j=0}^{n-1} f_{n,n-j} \cdot \|X_{j+1} - \tilde{X}_{j+1}\|^2 = \text{Cov}(X_{n+1}, X_{n+1}) - \sum_{j=0}^{n-1} f_{n,n-j} \cdot v_j. \quad \blacksquare \end{aligned}$$

Beispiel 8.1

Sei X die stationäre $AR(1)$ -Zeitreihe mit $X_t = e_t + 0,5 \cdot X_{t-1}$ und $N(0,1^2)$ -verteiletem Weißen Rauschen e . Nach Satz 3.4 ist die Autokovarianzfunktion dieser Zeitreihe gegeben durch

$\gamma(h) = \sigma_e^2 \cdot \frac{0,5^h}{1-0,5^2} = \frac{0,5^h}{0,75}$. Möchte man nun mittels des Innovationsalgorithmus die Ein-Schritt-Vorhersagefehler v_n berechnen, so geht man folgendermaßen vor:

1. $v_0 = \gamma(0) = \frac{4}{3}$
2. $f_{1,1} = \frac{1}{v_0} \cdot \gamma(1) = \frac{1}{2}$
3. $v_1 = \gamma(0) - \sum_{j=0}^0 f_{n,n-j}^2 \cdot v_j = \gamma(0) - f_{1,1}^2 \cdot v_0 = 1$
4. $f_{2,2} = \frac{1}{v_0} \cdot \gamma(2) = \frac{1}{4}$
5. $f_{2,1} = \frac{1}{v_1} \cdot (\gamma(1) - \sum_{j=0}^0 f_{k,k-j} \cdot f_{n,n-j} \cdot v_j) = \frac{1}{v_1} \cdot (\gamma(1) - f_{1,1} \cdot f_{2,2} \cdot v_0) = \frac{1}{2}$
6. $v_2 = \gamma(0) - \sum_{j=0}^1 f_{n,n-j}^2 \cdot v_j = \gamma(0) - (f_{2,2}^2 \cdot v_0 + f_{2,1}^2 \cdot v_1) = 1$
- $\vdots$

Lemma 8.2

Sei X eine kausale $ARMA(p, q)$ -Zeitreihe mit $ARMA$ -Gleichung $\Phi(B) \cdot X_t = \Psi(B) \cdot e_t$ und Autokovarianzfunktion γ . Sei $m := \max(p, q)$ und sei Y jene Zeitreihe, die durch Transformation von X mittels

$$Y_t = \begin{cases} \frac{1}{\sigma_e} \cdot X_t, & t \in [1, m] \\ \frac{1}{\sigma_e} \cdot \Phi(B) \cdot X_t, & t > m \end{cases}$$

entsteht, so gilt:

- (i) Y ist zentriert
- (ii) Für $t > m$ gilt $Y_t = \frac{1}{\sigma_e} \cdot (e_t + b_1 e_{t-1} + \dots + b_q e_{t-q})$.
- (iii) $\forall n \geq 1 : H_n := \text{span}\{X_1, \dots, X_n\} = \text{span}\{Y_1, \dots, Y_n\}$.

Beweis:

- (i) Nachdem gemäß Korollar 4.1 kausale $ARMA$ -Zeitreihen immer zentriert sind, gilt:

1.Fall: $t \in [1, m]$

$$\mathbb{E}(Y_t) = \mathbb{E}\left(\frac{1}{\sigma_e} \cdot X_t\right) = \frac{1}{\sigma_e} \cdot \mathbb{E}(X_t) = 0$$

2.Fall: $t > m$

$$\mathbb{E}(Y_t) = \mathbb{E}\left(\frac{1}{\sigma_e} \cdot \Phi(B) \cdot X_t\right) = \frac{1}{\sigma_e} \cdot \left(-\sum_{j=0}^p a_j \cdot \mathbb{E}(X_t)\right) = 0$$

- (ii) $Y_t = \frac{1}{\sigma_e} \cdot (X_t - a_1 X_{t-1} - \dots - a_p X_{t-p}) =$

$$\begin{aligned} &= \frac{1}{\sigma_e} \cdot \left(\underbrace{a_1 X_{t-1} + \dots + a_p X_{t-p} + e_t + b_1 e_{t-1} + \dots + b_q e_{t-q}}_{=X_t} - a_1 X_{t-1} - \dots - a_p X_{t-p} \right) = \\ &= \frac{1}{\sigma_e} \cdot (e_t + b_1 e_{t-1} + \dots + b_q e_{t-q}) . \end{aligned}$$

- (iii) Es gilt zunächst trivialerweise, dass $Y_t \in H_n$, da man für alle $t \geq 1$ Y_t als Linearkombination von $X_1, \dots, X_n$ ausdrücken kann. Induktiv ist leicht zu sehen, dass jedes Y_k mit $k \geq 2$ nicht als Linearkombination von $Y_1, \dots, Y_{k-1}$ ausgedrückt werden kann, da e_k unkorreliert zu $e_1, \dots, e_{k-1}$ ist. Deshalb sind $Y_1, \dots, Y_n$ linear unabhängig und somit eine Basis von H_n . ■

Satz 8.2

Sei X eine kausale $ARMA(p, q)$ -Zeitreihe mit $ARMA$ -Gleichung $\Phi(B) \cdot X_t = \Psi(B) \cdot e_t$ und Autokovarianzfunktion γ . Sei $m := \max(p, q)$, sei $b_j := 0 \quad \forall j > q$ und sei Y jene Zeitreihe, die durch Transformation von X wie in Lemma 8.2 entsteht, so gilt:

- (i) Die Kovarianzen von Y sind gegeben durch

$$Cov(Y_i, Y_j) = \begin{cases} \frac{1}{\sigma_e^2} \cdot \gamma(i-j), & i, j \in [1, m] \\ \frac{1}{\sigma_e^2} \cdot \left[\gamma(i-j) - \sum_{k=1}^p a_k \cdot \gamma(k - |i-j|) \right], & \min(i, j) \leq m < \max(i, j) \leq 2m \\ \sum_{k=0}^q b_k \cdot b_{k+|i-j|}, & \min(i, j) > m \\ 0, & \text{sonst} \end{cases}$$

- (ii) Seien $n > j$, $n > m$ und $j < n - q$, so gilt $Cov(Y_n, Y_j) = 0$.

Beweis:

- (i) 1.Fall: $i, j \in [1, m]$

$$Cov(Y_i, Y_j) = Cov\left(\frac{1}{\sigma_e} \cdot X_i, \frac{1}{\sigma_e} \cdot X_j\right) = \frac{1}{\sigma_e^2} \cdot Cov(X_i, X_j) = \frac{1}{\sigma_e^2} \cdot \gamma(i-j) .$$

4. Fall: $\min(i, j) \leq m$ und $\max(i, j) > 2m$

Sei o.B.d.A. $i < j$.

Sei $p = m$. Dann gilt:

$$Cov(Y_i, Y_j) = \frac{1}{\sigma_e^2} Cov\left(a_1 X_{i-1} + \dots + a_p X_{i-p} + b_0 \underbrace{e_i}_{\leq p} + \dots + b_q \underbrace{e_{i-q}}_{\leq p}, b_1 \underbrace{e_{j-1}}_{> p} + \dots + b_q \underbrace{e_{j-q}}_{> p}\right)$$

Aufgrund der Unkorreliertheit des Weißen Rauschens gilt nun $Cov(Y_i, Y_j) = 0$.

Sei $q = m$. Dann gilt:

$$Cov(Y_i, Y_j) = \frac{1}{\sigma_e^2} Cov\left(a_1 X_{i-1} + \dots + a_p X_{i-p} + b_0 \underbrace{e_i}_{\leq q} + \dots + b_q \underbrace{e_{i-q}}_{\leq q}, b_1 \underbrace{e_{j-1}}_{> q} + \dots + b_q \underbrace{e_{j-q}}_{> q}\right)$$

Aufgrund der Unkorreliertheit des Weißen Rauschens gilt wiederum $Cov(Y_i, Y_j) = 0$.

2.Fall: $\min(i, j) \leq m < \max(i, j) \leq 2m$

Sei o.b.d.A. $i \leq m < j$, dann gilt:

$$\begin{aligned} \text{Cov}(Y_i, Y_j) &= \text{Cov}\left(\frac{1}{\sigma_e} \cdot X_i, \frac{1}{\sigma_e} \cdot \Phi(B) \cdot X_j\right) = \frac{1}{\sigma_e^2} \cdot \text{Cov}\left(X_i, -\sum_{k=0}^p a_k \cdot X_{j-k}\right) = \\ &= \frac{1}{\sigma_e^2} \cdot [\text{Cov}(X_i, -a_0 X_j) + \text{Cov}(X_i, -a_1 X_{j-1}) + \dots + \text{Cov}(X_i, -a_p X_{j-p})] = \\ &= \frac{1}{\sigma_e^2} \cdot [-a_0 \cdot \gamma(i-j) - a_1 \cdot \gamma(1-|i-j|) - \dots - a_p \cdot \gamma(p-|i-j|)] = \\ &= \frac{1}{\sigma_e^2} \cdot \left[\gamma(i-j) - \sum_{k=1}^p a_k \cdot \gamma(k-|i-j|) \right]. \end{aligned}$$

3.Fall: $\min(i, j) > m$

Sei $i < j$, dann gilt:

$$\begin{aligned} \text{Cov}(Y_i, Y_j) &= \frac{1}{\sigma_e} \cdot \text{Cov}(b_0 e_i + \dots + b_q e_{i-q}, Y_j) = \frac{1}{\sigma_e} \cdot \sum_{k=0}^q b_k \cdot \text{Cov}(e_{i-k}, Y_j) = \\ &= \frac{1}{\sigma_e^2} \cdot \sum_{k=0}^q b_k \cdot \sum_{r=0}^q b_r \cdot \underbrace{\text{Cov}(e_{i-k}, e_{j-r})}_{=0 \text{ für } i-k \neq j-r} = \frac{1}{\sigma_e^2} \cdot \sum_{k=0}^q b_k \cdot b_{k-(i-j)} \cdot \underbrace{\text{Cov}(e_{i-k}, e_{i-k})}_{=\sigma_e^2} \end{aligned}$$

Sei $i > j$, dann gilt:

$$\begin{aligned} \text{Cov}(Y_i, Y_j) &= \frac{1}{\sigma_e} \cdot \text{Cov}(Y_i, b_0 e_j + \dots + b_q e_{j-q}) = \frac{1}{\sigma_e} \cdot \sum_{k=0}^q b_k \cdot \text{Cov}(Y_i, e_{j-k}) = \\ &= \frac{1}{\sigma_e^2} \cdot \sum_{k=0}^q b_k \cdot \sum_{r=0}^q b_r \cdot \underbrace{\text{Cov}(e_{i-r}, e_{j-k})}_{=0 \text{ für } i-r \neq j-k} = \frac{1}{\sigma_e^2} \cdot \sum_{k=0}^q b_k \cdot b_{k+(i-j)} \cdot \underbrace{\text{Cov}(e_{j-k}, e_{j-k})}_{=\sigma_e^2} \\ &\Rightarrow \forall i, j : \min(i, j) > m : \text{Cov}(Y_i, Y_j) = \sum_{k=0}^q b_k \cdot b_{k+|i-j|} . \end{aligned}$$

(ii) 1.Fall: $j > m$

$$\text{Cov}(Y_n, Y_j) = \sum_{k=0}^q b_k \cdot \underbrace{b_{k+|n-j|}}_{\substack{> q \\ =0}} = 0 .$$

2.Fall: $j < m$

$$\text{Cov}(Y_n, Y_j) = \frac{1}{\sigma_e^2} \cdot \text{Cov}\left(b_0 e_n + \dots + b_q e_{n-q}, -\sum_{k=0}^p a_k \cdot X_{j-k} + \underbrace{b_0 e_j}_{j < n-q} + \dots + b_q e_{j-q}\right) = 0 .$$

■

Lemma 8.3

Sei Y eine Zeitreihe, die durch Transformation einer kausalen $ARMA(p, q)$ -Zeitreihe gemäß Lemma 8.2 entsteht und sei $m := \max(p, q)$. Verwendet man zur Berechnung der Ein-Schritt-Vorhersagen $\tilde{Y}_{i+1}$ den Innovationsalgorithmus, so gilt für die somit ermittelten Koeffizienten $f_{i,j}$

$$i \geq m \wedge j > q \Rightarrow f_{i,j} = 0 .$$

Beweis:

Sei $i \geq \max(m, q + 1)$ beliebig und sei $l := i - (q + 1)$. Trivialerweise gilt $l = i - q - 1 < i - q$ und somit selbstverständlich auch $l - 1 < i - q$; $l - 2 < i - q$; ...; $1 < i - q$.

Somit gilt aufgrund von Satz 8.2 (ii)

$$\begin{aligned} f_{i,i} &= \frac{1}{v_0} \cdot \underbrace{\text{Cov}(Y_{i+1}, Y_1)}_{=0} = 0 \\ f_{i,i-1} &= \frac{1}{v_1} \cdot \left[\underbrace{\text{Cov}(Y_{i+1}, Y_2)}_{=0} - f_{1,1} \cdot \underbrace{f_{i,i}}_{=0} \cdot v_0 \right] = 0 \\ &\vdots \\ f_{i,i-l} &= \frac{1}{v_l} \cdot \left[\underbrace{\text{Cov}(Y_{i+1}, Y_l)}_{=0} - f_{l,l} \cdot \underbrace{f_{i,i}}_{=0} \cdot v_0 + f_{l,l-1} \cdot \underbrace{f_{i,i-1}}_{=0} \cdot v_1 + \cdots + f_{l,1} \cdot \underbrace{f_{i,i-(l-1)}}_{=0} \cdot v_{l-1} \right] = 0 \quad . \end{aligned}$$

■

Satz 8.3 Berechnung der Ein-Schritt-Vorhersagen für kausale $ARMA$ -Zeitreihen

Sei X eine kausale $ARMA(p, q)$ -Zeitreihe mit $ARMA$ -Gleichung $\Phi(B) \cdot X_t = \Psi(B) \cdot e_t$ und Autokovarianzfunktion γ und sei Y die gemäß Lemma 8.2 transformierte Zeitreihe.

Seien $\tilde{Y}_{n+1}$ die Ein-Schritt-Vorhersagen und r_n die Ein-Schritt-Vorhersagefehler für Y_{n+1} und $f_{i,j}$ die Koeffizienten, die bei Berechnung von $\tilde{Y}_{n+1}$ mittels des Innovationsalgorithmus auftreten und sei $m := \max(p, q)$, so lassen sich die Ein-Schritt-Vorhersagen $\tilde{X}_{n+1}$ und die Vorhersagefehler v_n für X_{n+1} rekursiv berechnen mittels

$$\begin{aligned} \bullet \quad \tilde{X}_{n+1} &= \begin{cases} \sum_{j=1}^n f_{n,j} \cdot (X_{n+1-j} - \tilde{X}_{n+1-j}), & 1 \leq n < m \\ \sum_{j=1}^p a_j \cdot X_{n+1-j} + \sum_{j=1}^q f_{n,j} \cdot (X_{n+1-j} - \tilde{X}_{n+1-j}), & n \geq m \end{cases} \\ \bullet \quad v_n &= \sigma_e^2 \cdot r_n \quad . \end{aligned}$$

Beweis:

Zunächst gilt aufgrund von Lemma 8.3, dass

$$\tilde{Y}_{n+1} = \begin{cases} \sum_{j=1}^n f_{n,j} \cdot (Y_{n-j+1} - \tilde{Y}_{n-j+1}), & 1 \leq n < m \\ \sum_{j=1}^q f_{n,j} \cdot (Y_{n-j+1} - \tilde{Y}_{n-j+1}), & n \geq m \end{cases} \quad .$$

Projiziert man Y_t auf $H_{t-1} := \text{span}\{X_1, \dots, X_{t-1}\} = \text{span}\{Y_1, \dots, Y_{t-1}\}$, so erhält man

$$\tilde{Y}_t = \begin{cases} \frac{1}{\sigma_e} \cdot \tilde{X}_t, & t \in [1, m] \\ \frac{1}{\sigma_e} \cdot (\tilde{X}_t - a_1 X_{t-1} - \cdots - a_p X_{t-p}), & t > m \end{cases} \quad .$$

Außerdem gilt für alle $t \geq 1$

$$\sigma_e \cdot (Y_t - \tilde{Y}_t) = \begin{cases} X_t - \tilde{X}_t, & t \in [1, m] \\ -a_0 X_t - \dots - a_p X_{t-p} - (-a_0 \tilde{X}_t - \dots - a_p \tilde{X}_{t-p}), & t > m \end{cases} = X_t - \tilde{X}_t.$$

Durch einfaches Einsetzen erhält man somit

$$\tilde{X}_{n+1} = \begin{cases} \sum_{j=1}^n f_{n,j} \cdot (X_{n+1-j} - \tilde{X}_{n+1-j}), & 1 \leq n < m \\ -\left(\sum_{j=1}^p a_j \cdot X_{n+1-j}\right) + \sum_{j=1}^q f_{n,j} \cdot (X_{n+1-j} - \tilde{X}_{n+1-j}), & n \geq m \end{cases}.$$

$$\text{Außerdem gilt } v_n = \mathbb{E}(X_{n+1} - \tilde{X}_{n+1})^2 = \mathbb{E}[\sigma_e \cdot (Y_{n+1} - \tilde{Y}_{n+1})]^2 = \sigma_e^2 \cdot \mathbb{E}(Y_{n+1} - \tilde{Y}_{n+1})^2 = \sigma_e^2 \cdot r_n.$$

■

Korollar 8.1

Sei Y eine Zeitreihe, die durch Transformation einer kausalen $ARMA(p, q)$ -Zeitreihe X gemäß Lemma 8.2 entsteht und sei $m := \max(p, q)$, so gilt:

- (i) Für alle $i, j \geq 1$ ist $\text{Cov}(Y_i, Y_j)$ unabhängig von der Varianz des Weißen Rauschens σ_e .
- (ii) Verwendet man zur Berechnung der Ein-Schritt-Vorhersagen $\tilde{Y}_{n+1}$ den Innovationsalgorithmus, so sind alle Koeffizienten $f_{i,j}$ und alle Ein-Schritt-Vorhersagefehler r_n unabhängig von σ_e .
Außerdem sind die Ein-Schritt-Vorhersagen $\tilde{X}_{n+1}$ für alle $n \geq 1$ unabhängig von σ_e .

Beweis:

- (i) Gemäß Satz 5.1 gilt $\gamma(h) = \sigma_e^2 \cdot \sum_{j=0}^{\infty} \tilde{a}_j \tilde{a}_{j+|h|}$.

Für $i, j \in [1, m]$ ergibt sich also nach Satz 8.2(i):

$$\text{Cov}(Y_i, Y_j) = \frac{1}{\sigma_e^2} \cdot \gamma(i-j) = \frac{1}{\sigma_e^2} \cdot \sigma_e^2 \cdot \sum_{j=0}^{\infty} \tilde{a}_j \tilde{a}_{j+|i-j|} = \sum_{j=0}^{\infty} \tilde{a}_j \tilde{a}_{j+|i-j|},$$

für $\min(i, j) \leq m < \max(i, j) \leq 2m$ ergibt sich ebenfalls gemäß Satz 8.2(i)

$$\begin{aligned} \text{Cov}(Y_i, Y_j) &= \frac{1}{\sigma_e^2} \cdot \left[\gamma(i-j) - \sum_{k=1}^p a_k \cdot \gamma(k-|i-j|) \right] = \\ &= \frac{1}{\sigma_e^2} \cdot \left[\sigma_e^2 \cdot \sum_{j=0}^{\infty} \tilde{a}_j \tilde{a}_{j+|i-j|} - \sum_{k=1}^p a_k \cdot \sigma_e^2 \cdot \sum_{j=0}^{\infty} \tilde{a}_j \tilde{a}_{j+|k-|i-j||} \right] = \\ &= \sum_{j=0}^{\infty} \tilde{a}_j \tilde{a}_{j+|i-j|} - \sum_{k=1}^p a_k \cdot \sum_{j=0}^{\infty} \tilde{a}_j \tilde{a}_{j+|k-|i-j||}. \end{aligned}$$

Für die anderen beiden Fälle ist der Beweis der Aussage trivial.

- (ii) Die Aussage folgt direkt aus dem Innovationsalgorithmus, aus Satz 8.3 und aus (i).

■

Anmerkung 8.3

Um durch Verwendung des Innovationsalgorithmus die Ein-Schritt-Vorhersagen $\tilde{X}_{n+1}$ mittels $\tilde{X}_{n+1} = \sum_{j=1}^n f_{n,j} \cdot (X_{n-j+1} - \tilde{X}_{n-j+1})$ zu berechnen, müssen für jedes n stets sowohl alle vorhergegangenen Werte $X_{n-1}, \dots, X_1$ und ihre Projektionen als auch sämtliche Koeffizienten $f_{n,1}, \dots, f_{n,n}$ verwendet werden.

Handelt es sich bei X allerdings um eine kausale $ARMA(p, q)$ -Zeitreihe, so kann man durch Verwendung des Algorithmus aus Satz 8.3 Rechenzeit einsparen, da für $n \geq m$ maximal p Werte $X_{n+1-p}, \dots, X_n$ und maximal q Projektionen $\tilde{X}_{n+1-q}, \dots, \tilde{X}_n$ verwendet werden müssen. Wie man in Lemma 8.3 gesehen hat, wird auch die Anzahl der zu berechnenden Koeffizienten $f_{i,j}$ drastisch reduziert.

Während also der Innovationsalgorithmus den Vorteil hat, sogar für nicht-stationäre Zeitreihen einsetzbar zu sein, sollte für kausale $ARMA$ -Zeitreihen stets der Algorithmus aus Satz 8.3 verwendet werden.

Beispiel 8.2

Sei X die selbe $AR(1)$ -Zeitreihe wie in Beispiel 8.1. Für die Berechnung der Ein-Schritt-Vorhersagen $\tilde{X}_n$ mittels des Innovationsalgorithmus geht man folgendermaßen vor:

1. $\tilde{X}_1 := 0$
2. $\tilde{X}_2 = f_{1,1} \cdot (X_1 - \tilde{X}_1) = \frac{1}{2} \cdot X_1$
3. $\tilde{X}_3 = f_{2,1} \cdot (X_2 - \tilde{X}_2) + f_{2,2} \cdot (X_1 - \tilde{X}_1) = \frac{1}{9} \cdot X_1 + \frac{4}{9} X_2$
- $\vdots$

Die Berechnung für wachsende n gestaltet sich also, wie man sieht, immer aufwändiger. Verwendet man hingegen Satz 8.3, so reduziert sich der Arbeitsaufwand wegen $p = 1$, $q = 0$ und $m = 1$ auf die Berechnung von $\tilde{X}_{n+1} = a_1 X_n + \underbrace{\sum_{j=1}^q f_{n,j} \cdot (X_{n+1-j} - \tilde{X}_{n+1-j})}_{=0} = a_1 X_n = \frac{1}{2} \cdot X_n$.

Definition 8.3 Gaußsche Zeitreihe

Eine Zeitreihe $(X_t : t \in \mathbb{Z})$ heißt *Gaußsche Zeitreihe*, wenn für alle $k \in \mathbb{N}$ und alle $t_1, \dots, t_k \in \mathbb{Z}$ gilt, dass die gemeinsame Verteilung von $(X_{t_1}, \dots, X_{t_k})^T$ eine k -dimensionale Normalverteilung ist.

Sei $\mathbf{X}_n = (X_1, \dots, X_n)^T$ ein Zufallsvektor, bestehend aus Zufallsvariablen $X_1, \dots, X_n$ einer Gaußschen Zeitreihe mit Erwartungsvektor $\boldsymbol{\mu}_n$ und Kovarianzmatrix $\mathbf{A}_n$ und sei $\widehat{X}_{[n]} := (\hat{X}_1, \dots, \hat{X}_n)^T := \hat{\mathbf{X}}_n$ eine Realisierung von $\mathbf{X}_n$, so ist, sofern $\mathbf{A}_n$ invertierbar ist, die Dichtefunktion von $\mathbf{X}_n$ bei $\hat{\mathbf{X}}_n$ gegeben durch

$$f(\hat{\mathbf{X}}_n) = \frac{1}{\sqrt{(2\pi)^n \cdot \det(\mathbf{A}_n)}} \cdot \exp\left(-\frac{1}{2} \cdot (\hat{\mathbf{X}}_n - \boldsymbol{\mu}_n)^T \cdot \mathbf{A}_n^{-1} \cdot (\hat{\mathbf{X}}_n - \boldsymbol{\mu}_n)\right).$$

Satz 8.4

Sei X eine gaußsche, kausale $ARMA$ -Zeitreihe und sei für $n \geq 1$ $\hat{\mathbf{X}}_n := (\hat{X}_1, \dots, \hat{X}_n)^T$ eine Realisierung des Zufallsvektors $\mathbf{X}_n := (X_1, \dots, X_n)^T$, dessen Kovarianzmatrix $\mathbf{A}_n$ invertierbar sei. Seien außerdem für alle $1 \leq j \leq n+1$ $\bar{\bar{X}}_j$ die empirischen Ein-Schritt-Vorhersagen für $\hat{X}_j$ und seien v_j die Ein-Schritt-Vorhersagefehler für X_{j+1} , so lässt sich der Wert der Dichtefunktion von $\mathbf{X}_n$ bei $\hat{\mathbf{X}}_n$ berechnen mittels

$$f(\hat{\mathbf{X}}_n) = \frac{1}{\sqrt{(2\pi)^n \cdot v_0 \cdots v_{n-1}}} \cdot \exp\left(-\frac{1}{2} \cdot \sum_{j=1}^n \frac{(\hat{X}_j - \bar{\bar{X}}_j)^2}{v_{j-1}}\right).$$

Beweis:

Gemäß Korollar 4.1 sind kausale $ARMA$ -Zeitreihen stets zentriert, also ist der Erwartungsvektor von $\mathbf{X}_n$ der Nullvektor. Es gilt also zunächst

$$f(\hat{\mathbf{X}}_n) = \frac{1}{\sqrt{(2\pi)^n \cdot \det(\mathbf{A}_n)}} \cdot \exp\left(-\frac{1}{2} \cdot \hat{\mathbf{X}}_n^T \cdot \mathbf{A}_n^{-1} \cdot \hat{\mathbf{X}}_n\right).$$

Seien $f_{i,j}$ für $i \geq 1$ und $j \in [1, \dots, i]$ die Koeffizienten, die man mittels des Innovationsalgorithmus erhält, sei $\tilde{\mathbf{X}}_n := (\tilde{X}_1, \dots, \tilde{X}_n)^T$ und seien

$$\mathbf{B} := \begin{pmatrix} 1 & 0 & \cdots & \cdots & 0 & 0 \\ f_{1,1} & 1 & 0 & & & 0 \\ f_{2,2} & f_{2,1} & \ddots & \ddots & & \vdots \\ \vdots & & \ddots & 1 & 0 & \vdots \\ \vdots & & & \ddots & 1 & 0 \\ f_{n-1,n-1} & \cdots & \cdots & \cdots & f_{n-1,1} & 1 \end{pmatrix}, \quad \mathbf{C} := \begin{pmatrix} v_0 & 0 & 0 & \cdots & 0 \\ 0 & v_1 & 0 & & \vdots \\ 0 & 0 & \ddots & & \vdots \\ \vdots & & & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & v_{n-1} \end{pmatrix}.$$

Gemäß des Innovationsalgorithmus gilt

$$\tilde{\mathbf{X}}_n = \begin{pmatrix} \tilde{X}_1 \\ \tilde{X}_2 \\ \vdots \\ \tilde{X}_n \end{pmatrix} = \begin{pmatrix} 0 \\ \sum_{j=1}^1 f_{n,j} \cdot (X_{n-j+1} - \tilde{X}_{n-j+1}) \\ \vdots \\ \sum_{j=1}^{n-1} f_{n,j} \cdot (X_{n-j+1} - \tilde{X}_{n-j+1}) \end{pmatrix} = (\mathbf{B} - \mathbf{I}) \cdot (\mathbf{X}_n - \tilde{\mathbf{X}}_n).$$

Somit muss nun auch gelten, dass

$$\mathbf{X}_n = \mathbf{X}_n - \tilde{\mathbf{X}}_n + \tilde{\mathbf{X}}_n = \mathbf{I} \cdot (\mathbf{X}_n - \tilde{\mathbf{X}}_n) + (\mathbf{B} - \mathbf{I}) \cdot (\mathbf{X}_n - \tilde{\mathbf{X}}_n) = \mathbf{B} \cdot (\mathbf{X}_n - \tilde{\mathbf{X}}_n).$$

Bekanntermaßen gilt für alle Matrizen $\mathbf{M} \in \mathbb{R}^{m \times n}$ und alle n -dimensionalen Zufallsvektoren $\mathbf{Y}$, dass

$\text{Cov}(\mathbf{M} \cdot \mathbf{Y}) = \mathbf{M} \cdot \text{Cov}(\mathbf{Y}) \cdot \mathbf{M}^T$ und offensichtlich gilt $\mathbf{C} = \text{Cov}(\mathbf{X}_n - \tilde{\mathbf{X}}_n)$. Somit erhält man

$$\mathbf{A}_n = \text{Cov}(\mathbf{X}_n) = \text{Cov}(\mathbf{B} \cdot (\mathbf{X}_n - \tilde{\mathbf{X}}_n)) = \mathbf{B} \cdot \text{Cov}(\mathbf{X}_n - \tilde{\mathbf{X}}_n) \cdot \mathbf{B}^T = \mathbf{B} \cdot \mathbf{C} \cdot \mathbf{B}^T.$$

$$\begin{aligned}
\text{Somit folgt, dass } \mathbf{X}_n^T \cdot \mathbf{A}_n^{-1} \cdot \mathbf{X}_n &= [\mathbf{B} \cdot (\mathbf{X}_n - \tilde{\mathbf{X}}_n)]^T \cdot (\mathbf{B} \cdot \mathbf{C} \cdot \mathbf{B}^T)^{-1} \cdot \mathbf{B} \cdot (\mathbf{X}_n - \tilde{\mathbf{X}}_n) = \\
&= (\mathbf{X}_n - \tilde{\mathbf{X}}_n)^T \cdot \underbrace{\mathbf{B}^T \cdot (\mathbf{B}^T)^{-1}}_{=I} \cdot \mathbf{C}^{-1} \cdot \underbrace{\mathbf{B}^{-1} \cdot \mathbf{B}}_{=I} \cdot (\mathbf{X}_n - \tilde{\mathbf{X}}_n) = (\mathbf{X}_n - \tilde{\mathbf{X}}_n)^T \cdot \mathbf{C}^{-1} \cdot (\mathbf{X}_n - \tilde{\mathbf{X}}_n) = \\
&= \left(\frac{X_1 - \tilde{X}_1}{v_0}, \dots, \frac{X_n - \tilde{X}_n}{v_{n-1}} \right) \cdot (X_1 - \tilde{X}_1, \dots, X_n - \tilde{X}_n)^T = \sum_{j=1}^n \frac{(X_j - \tilde{X}_j)^2}{v_{j-1}} \\
\Rightarrow \hat{\mathbf{X}}_n^T \cdot \mathbf{A}_n^{-1} \cdot \hat{\mathbf{X}}_n &= \sum_{j=1}^n \frac{(\hat{X}_j - \bar{\bar{X}}_j)^2}{v_{j-1}} .
\end{aligned}$$

Nachdem die Determinante einer Dreiecksmatrix stets das Produkt ihrer Hauptdiagonalelemente ist, gilt außerdem

$$\det(\mathbf{A}_n) = \det(\mathbf{B} \cdot \mathbf{C} \cdot \mathbf{B}^T) = \underbrace{\det(\mathbf{B})}_{=1} \cdot \det(\mathbf{C}) \cdot \underbrace{\det(\mathbf{B}^T)}_{=1} = \det(\mathbf{C}) = v_0 \cdot v_1 \cdots v_{n-1} .$$

Gesamt ergibt sich durch Einsetzen für $\det(\mathbf{A}_n)$ und für $\hat{\mathbf{X}}_n^T \cdot \mathbf{A}_n^{-1} \cdot \hat{\mathbf{X}}_n$ also

$$f(\hat{\mathbf{X}}_n) = \frac{1}{\sqrt{(2\pi)^n \cdot v_0 \cdots v_{n-1}}} \cdot \exp\left(-\frac{1}{2} \cdot \sum_{j=1}^n \frac{(\hat{X}_j - \bar{\bar{X}}_j)^2}{v_{j-1}}\right) .$$

■

Definition 8.4 Maximum- Likelihood -Schätzer

Sei $\widehat{X}_{[n]} := (\widehat{X}_t : t \in [1, n])$ eine Realisierung einer $ARMA(p, q)$ -Zeitreihe, deren Parameter $a_1, \dots, a_p, b_1, \dots, b_q$ und σ_e^2 unbekannt sind, so nennt man $\hat{a}_1, \dots, \hat{a}_p, \hat{b}_1, \dots, \hat{b}_q, \hat{\sigma}_e^2$ die *Maximum-Likelihood-Schätzer* für die Parameter $a_1, \dots, a_p, b_1, \dots, b_q$ und σ_e^2 , wenn für $a_1 := \hat{a}_1, \dots, a_p := \hat{a}_p, b_1 := \hat{b}_1, \dots, b_q := \hat{b}_q, \sigma_e^2 := \hat{\sigma}_e^2$ der Funktionswert der Dichtefunktion der gemeinsamen Verteilung von $(X_1, \dots, X_n)^T$ bei $(\hat{X}_1, \dots, \hat{X}_n)^T$ unter allen möglichen Kombinationen von Parametern maximal ist.

Satz 8.5

Sei X definiert wie in Satz 8.4 und sei $v_n := \sigma_e^2 \cdot r_n$, so sind die *Maximum- Likelihood -Schätzer* $\hat{a}_1, \dots, \hat{a}_p, \hat{b}_1, \dots, \hat{b}_q$ gegeben als jene Kombination von Parametern $a_1, \dots, a_p, b_1, \dots, b_q$, für die folgender Ausdruck minimal wird:

$$M := \ln\left(\frac{1}{n} \cdot \sum_{j=1}^n \frac{(\hat{X}_j - \bar{\bar{X}}_j)^2}{r_{j-1}}\right) + \frac{1}{n} \cdot \sum_{j=0}^{n-1} \ln(r_j) .$$

Der *Maximum- Likelihood -Schätzer* $\hat{\sigma}_e^2$ ist dann gegeben durch

$$\hat{\sigma}_e^2 = \frac{1}{n} \cdot \sum_{j=1}^n \frac{(\hat{X}_j - \bar{\bar{X}}_j)^2}{r_{j-1}} .$$

Beweis:

Durch Einsetzen von $v_n := \sigma_e^2 \cdot r_n$ in die in Satz 8.4 ausgearbeitete Berechnungsmethode der Dichtefunktion erhält man

$$f(\hat{\mathbf{X}}_n) = \frac{1}{\sqrt{(2\pi\sigma_e^2)^n \cdot r_0 \cdots r_{n-1}}} \cdot \exp\left(-\frac{1}{2\sigma_e^2} \cdot \sum_{j=1}^n \frac{(\hat{X}_j - \bar{\bar{X}}_j)^2}{r_{j-1}}\right)$$

und mittels Logarithmierens beider Seiten

$$\ln[f(\hat{\mathbf{X}}_n)] = -\frac{n}{2} \ln(\sigma_e^2) - \frac{1}{2} \cdot \ln(2\pi \cdot r_0 \cdots r_{n-1}) - \frac{1}{2\sigma_e^2} \cdot \sum_{j=1}^n \frac{(\hat{X}_j - \bar{\bar{X}}_j)^2}{r_{j-1}}.$$

Nachdem aus Korollar 8.1 bekannt ist, dass $\tilde{X}_j$ (und somit selbstverständlich auch $\bar{\bar{X}}_j$) und r_j unabhängig von σ_e^2 sind, kann man zunächst partiell nach σ_e^2 ableiten und erhält

$$\frac{\partial \ln[f(\hat{\mathbf{X}}_n)]}{\partial \sigma_e^2} = -\frac{n}{2\sigma_e^2} + \frac{1}{2(\sigma_e^2)^2} \cdot \sum_{j=1}^n \frac{(\hat{X}_j - \bar{\bar{X}}_j)^2}{r_{j-1}} \stackrel{!}{=} 0 \Rightarrow \sigma_e^2 = \frac{1}{n} \cdot \sum_{j=1}^n \frac{(\hat{X}_j - \bar{\bar{X}}_j)^2}{r_{j-1}}.$$

Setzt man nun dieses die Wahrscheinlichkeit der beobachteten Realisierung maximierende σ_e^2 wiederum in $\ln[f(\hat{\mathbf{X}}_n)]$ ein, so erhält man

$$\ln[f(\hat{\mathbf{X}}_n)] = -\frac{n}{2} \cdot \ln\left(\frac{1}{n} \cdot \sum_{j=1}^n \frac{(\hat{X}_j - \bar{\bar{X}}_j)^2}{r_{j-1}}\right) - \frac{1}{2} \cdot \ln(2\pi \cdot r_0 \cdots r_{n-1}) - \frac{n}{2}.$$

Nun sucht man nach Parametern $\hat{a}_1, \dots, \hat{a}_p, \hat{b}_1, \dots, \hat{b}_q$, für die dieser Ausdruck maximal ist. Durch Vereinfachen erhält man

$$\begin{aligned} & -\frac{n}{2} \cdot \ln\left(\frac{1}{n} \cdot \sum_{j=1}^n \frac{(\hat{X}_j - \bar{\bar{X}}_j)^2}{r_{j-1}}\right) - \frac{1}{2} \cdot \ln(2\pi \cdot r_0 \cdots r_{n-1}) - \frac{n}{2} \rightarrow \text{maximal} \Leftrightarrow \\ & \Leftrightarrow \ln\left(\frac{1}{n} \cdot \sum_{j=1}^n \frac{(\hat{X}_j - \bar{\bar{X}}_j)^2}{r_{j-1}}\right) + \underbrace{\frac{1}{n} \cdot \ln(2\pi) + \frac{1}{n} \cdot \ln(r_0 \cdots r_{n-1})}_{>0} + 1 \rightarrow \text{minimal} \Leftrightarrow \\ & \Leftrightarrow \ln\left(\frac{1}{n} \cdot \sum_{j=1}^n \frac{(\hat{X}_j - \bar{\bar{X}}_j)^2}{r_{j-1}}\right) + \frac{1}{n} \cdot \sum_{j=0}^{n-1} \ln(r_j) \rightarrow \text{minimal}. \end{aligned}$$

■

Anmerkung 8.4

Obwohl im vorhergehenden Satz die *Maximum-Likelihood*-Schätzer nur für gaußsche, kausale *ARMA*-Zeitreihen beschrieben wurden ist es, wie in vielen Bereichen der Statistik, durchaus sinnvoll beliebigen kausalen *ARMA*-Zeitreihen eine multivariate Normalverteilung, also gaußsch zu sein, zu unterstellen, da hierdurch meist sehr gute Ergebnisse erzielt werden – selbst dann, wenn es sich nicht um eine gaußsche Zeitreihe handelt.

Anmerkung 8.5

Der in Satz 8.5 beschriebene Ausdruck $\ln \left(\frac{1}{n} \cdot \sum_{j=1}^n \frac{(\hat{x}_j - \bar{\bar{x}}_j)^2}{r_{j-1}} \right) + \frac{1}{n} \cdot \sum_{j=0}^{n-1} \ln(r_j)$, der in Abhängigkeit von Parametern $a_1, \dots, a_p, b_1, \dots, b_q$ für die gesuchte $ARMA(p, q)$ -Zeitreihe minimal werden soll, lässt sich nicht mehr weiter vereinfachen oder nach einem der Parameter partiell differenzieren. Deshalb sind die einzigen Möglichkeiten diese Minimalisierung durchzuführen iterative „Trial and Error – Methoden“, die, nachdem in Abhängigkeit jeder versuchten Parameterkombination sämtliche empirische Ein-Schritt-Vorhersagen $\bar{\bar{x}}_1, \dots, \bar{\bar{x}}_n$ und sämtliche Vorhersagefehler $r_0, \dots, r_{n-1}$ mittels Satz 8.3 berechnet werden müssen, sehr rechenaufwändig sind. Man kann die Rechenzeit signifikant verringern, indem man – ungenaue – Vorabschätzungen verwendet. Diese Schätzungen erhält man folgendermaßen:

Seien $\hat{f}_{i,j}$ die Koeffizienten und $\hat{v}_j$ die Ein-Schritt-Vorhersagefehler, die bei Durchführung des Innovationsalgorithmus für eine kausale $ARMA$ -Zeitreihe entstehen, wenn man bei dessen Durchführung die tatsächliche (unbekannte) Autokovarianzfunktion durch die empirische Autokovarianzfunktion ersetzt, so erhält man für ausreichend großes m mit $q + p \leq m < n$ eine erste grobe Abschätzung der Parameter $\hat{a}'_1, \dots, \hat{a}'_p, \hat{b}'_1, \dots, \hat{b}'_q, \hat{\sigma}_e^{2'}$ mittels Lösung des linearen Gleichungssystems

$$\begin{pmatrix} \hat{f}_{m,q} & \hat{f}_{m,q-1} & \cdots & \hat{f}_{m,q+1-p} \\ \hat{f}_{m,q+1} & \hat{f}_{m,q} & \cdots & \hat{f}_{m,q+2-p} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{f}_{m,q+p-1} & \hat{f}_{m,q+p-2} & \cdots & \hat{f}_{m,q} \end{pmatrix} \cdot \begin{pmatrix} \hat{a}'_1 \\ \hat{a}'_1 \\ \vdots \\ \hat{a}'_p \end{pmatrix} = \begin{pmatrix} \hat{f}_{m,q+1} \\ \hat{f}_{m,q+2} \\ \vdots \\ \hat{f}_{m,q+p} \end{pmatrix},$$

des rekursiven Anwendens von

$$\hat{b}'_j = \hat{f}_{m,j} - \sum_{i=1}^{\min(j,p)} \hat{a}'_i \cdot \hat{f}_{m,j-i} \quad \forall j \in [1, q]$$

und des einfachen Gleichsetzens von $\hat{\sigma}_e^{2'} = \hat{v}_m$.

Nachdem die hierzu benötigte Beweisführung recht umfangreich ist, und es sich hierbei nur um eine Vereinfachung der Berechnung der *Maximum-Likelihood*-Schätzer handelt, die auch ohne diese Vorabschätzung prinzipiell möglich ist, wird hierbei auf einen Beweis verzichtet. Dieser ist beispielsweise in Peter J. Brockwell, Richard A. Davis, Time Series: Theory and Methods (Springer, 1991) S.245-254 nachzulesen.

Anmerkung 8.6

Da die gesuchte Zeitreihe X im vorhergegangenen Satz nach Voraussetzung kausal sein muss, ist es unumgänglich beim Ausprobieren von Parameterkombinationen $a_1, \dots, a_p, b_1, \dots, b_q$ zur Findung der *Maximum-Likelihood*-Schätzer $\hat{a}_1, \dots, \hat{a}_p, \hat{b}_1, \dots, \hat{b}_q, \hat{\sigma}_e^2$ diese auf solche Parameter zu beschränken, die die Bedingungen aus Satz 4.1 erfüllen.

Ebenfalls sinnvoll (wenn auch nicht zwingend notwendig) ist es, von vornherein *Maximum-Likelihood*-Lösungen auszuschließen, die gemäß Satz 4.4 zu nicht-invertierbaren $ARMA$ -Zeitreihen gehören.

9. Modellierung mittels *ARMA*-Zeitreihen

Das vorherige Kapitel lieferte Methoden ein kausales (und evtl. auch invertierbares) *ARMA*(p, q)-Modell zu beobachteten Daten zu erstellen, wobei davon ausgegangen wurde, dass die Ordnungen p und q bereits bekannt seien. Offensichtlich ist dies aber in der Praxis üblicherweise nicht der Fall. Es stellt sich also die Frage danach, wie man jene Ordnungen bestimmen kann, die mit der Methodik aus Kapitel 8 das beste Modell liefern.

Dieses Kapitel wird nun zwei Methoden hierfür beschreiben: Zunächst wird beschrieben werden, wie man durch Betrachtung der empirischen (partiellen) Autokorrelationsfunktion bestimmen kann, ob es sich bei der beobachteten Zeitreihe um eine *AR*- oder *MA*-Zeitreihe oder gar um ein Weißes Rauschen handelt. Sollte dies nicht der Fall sein, bleibt schließlich nur die weniger elegante Methode verschiedenste Kombination von Ordnungen p und q mittels dem in Kapitel 8 beschriebenen *Maximum-Likelihood*-Verfahren zu berechnen und anschließend auf ihre Güte zu untersuchen. Hierfür stehen mehrere Kriterien zur Verfügung, wovon hier das wohl gebräuchlichste näher vorgestellt werden wird, nämlich das so genannte *AIC*(C).

Anschließend wird kurz ein Verfahren vorgestellt werden mit dem überprüft werden kann ob ein erstelltes Modell überhaupt die vorliegende Realisierung einer Zeitreihe adäquat beschreiben kann. Zum Abschluss wird schlussendlich näher auf die Thematik der Erstellung von Zukunftsprognosen eingegangen werden, die ja eine, wenn nicht sogar die wichtigste, Motivation für die Beschäftigung mit den Methoden der Zeitreihenanalyse darstellt.

9.1 Überprüfen auf einfache *MA*(q) oder *AR*(p)-Zeitreihen

9.1.1 Überprüfen auf Weißes Rauschen

Nachdem ein identisch verteiltes und paarweise stochastisch unabhängiges Weißes Rauschen nichts anderes ist als eine *ARMA*(0,0)-Zeitreihe, und diese offensichtlich kausal ist, lässt sich zur Behandlung der Frage, ob es sich bei einer beobachteten Realisierung $\widehat{X}_{[n]} := (\widehat{X}_t: t \in [1, n])$ einer Zeitreihe um solch ein Weißes Rauschen handelt, für großes n Satz 7.2 anwenden.

Es gilt $\hat{\rho}(h) \sim N\left(\rho(h), \frac{1}{n} \cdot m_{h,h}\right) \Leftrightarrow \frac{\hat{\rho}(h) - \rho(h)}{\sqrt{\frac{1}{n} \cdot m_{h,h}}} \sim N(0,1)$.

Bezeichne $z_{(1-\frac{a}{2})}$ das $(1 - \frac{a}{2})$ -Quantil der Standardnormalverteilung, so ergibt sich für $0 < a < 1$

$$\begin{aligned} \mathbb{P}\left(-z_{(1-\frac{a}{2})} \leq \frac{\hat{\rho}(h) - \rho(h)}{\sqrt{\frac{1}{n} \cdot m_{h,h}}} \leq z_{(1-\frac{a}{2})}\right) &= \\ &= \mathbb{P}\left(\rho(h) - \frac{z_{(1-\frac{a}{2})} \cdot \sqrt{m_{h,h}}}{\sqrt{n}} \leq \hat{\rho}(h) \leq \rho(h) + \frac{z_{(1-\frac{a}{2})} \cdot \sqrt{m_{h,h}}}{\sqrt{n}}\right) = 1 - a \end{aligned}$$

Aus Definition 3.1 ist klar ersichtlich, dass nun, sollte es sich bei X tatsächlich um ein Weißes Rauschen handeln, die Autokorrelationen von X ab Lag 1 gleich Null sein müssen.

Um dies zu überprüfen erstellt man nun zu jedem Lag $h \geq 1$ ein Konfidenzintervall, so wie es gerade beschrieben wurde, und setzt $\rho(h)$ gleich Null. Somit erhält man jeweils ein Intervall, in dem, sollte die Hypothese, es handle sich bei X um ein Weißes Rauschen, stimmen, in $(1 - \alpha) \cdot 100$ Prozent der Realisierungen dieser Zeitreihe der Wert $\hat{\rho}(h)$ liegen muss.

Sollte die empirische Autokorrelation nun nicht in diesem Intervall liegen, kann man die Hypothese mit einer Irrtumswahrscheinlichkeit von $\alpha \cdot 100$ Prozent verwerfen.

Um diese wichtigen Intervalle berechnen zu können ergibt sich nun noch das Problem der Berechnung von $m_{h,h}$. Unter der Annahme, dass es sich um Weißes Rauschen handelt, ist dies sehr einfach, es gilt nach Satz 7.2 für alle $h \geq 1$:

$$m_{h,h} = \sum_{r=-\infty}^{\infty} \underbrace{\rho^2(r+h)}_{=0 \forall (r+h) \neq 0} + \underbrace{\rho(r-h) \cdot \rho(r+h)}_{=0} + \underbrace{2 \cdot \rho^2(h) \cdot \rho^2(r)}_{=0} - \underbrace{4 \cdot \rho(h) \cdot \rho(r) \cdot \rho(r+h)}_{=0} = \\ = \rho^2(0) = 1 \quad .$$

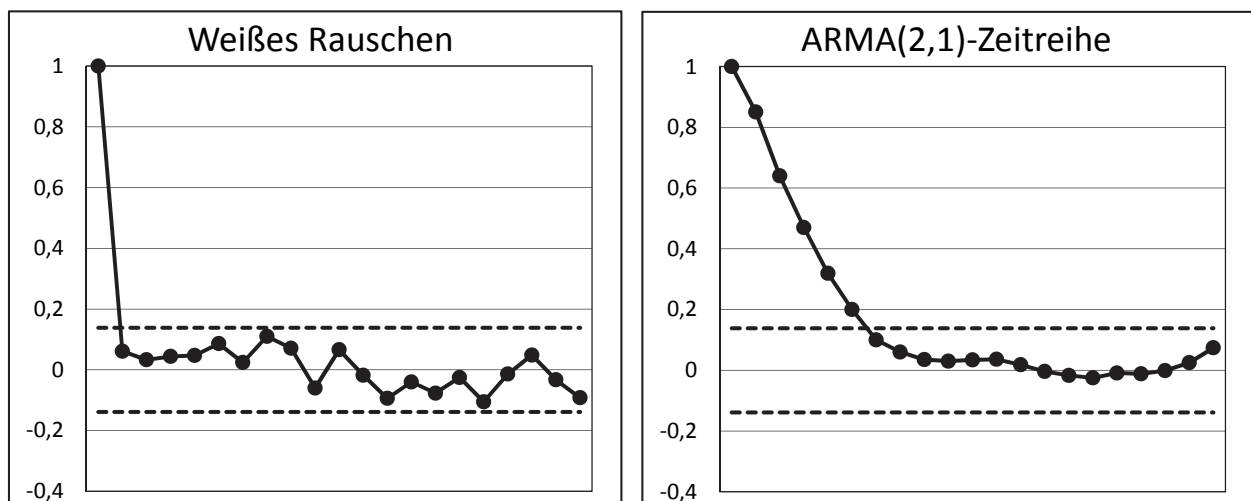
Somit ergibt sich also für alle Lags $h \geq 1$ für Irrtumswahrscheinlichkeit α das Konfidenzintervall

$$\left(-\frac{z_{(1-\frac{\alpha}{2})}}{\sqrt{n}} ; \frac{z_{(1-\frac{\alpha}{2})}}{\sqrt{n}} \right)$$

Beispiel 9.1

Die folgenden zwei Diagramme zeigen die empirischen Autokorrelationsfunktionen bis zum Lag 20 von Realisierungen vom Umfang 200 eines $N(0,2^2)$ -verteilten Weißen Rauschens und einer $ARMA(2,1)$ -Zeitreihe mit der $ARMA$ -Gleichung $X_t = 0,6X_{t-1} + 0,2X_{t-2} + 0,6e_{t-1} + e_t$.

Man sieht sofort, dass die Hypothese, es handle sich um ein Weißes Rauschen im Fall der $ARMA(2,1)$ -Zeitreihe eindeutig abgelehnt werden kann, im anderen Fall allerdings nicht, wobei die in der Praxis übliche Irrtumswahrscheinlichkeit $\alpha = 0,05$ verwendet wurde.



9.1.2 Überprüfen auf $MA(q)$ -Zeitreihen

Möchte man eine Realisierung $\widehat{X}_{[n]} := (\widehat{X}_t : t \in [1, n])$ darauf untersuchen, ob sie sich gut durch eine $MA(q)$ -Zeitreihe modellieren lässt, geht man genau so wie in 9.1.1 vor. Aus Satz 5.3 weiß man, dass die Autokorrelation von $MA(q)$ -Zeitreihen ab dem Lag $q + 1$ immer gleich Null ist.

Unter der Annahme, dass es sich um eine solche Zeitreihe handelt, erhält man also für alle $h > q$ wiederum die Konfidenzintervalle

$$\mathbb{P} \left(-\frac{z_{(1-\frac{a}{2})} \cdot \sqrt{m_{h,h}}}{\sqrt{n}} \leq \hat{\rho}(h) \leq \frac{z_{(1-\frac{a}{2})} \cdot \sqrt{m_{h,h}}}{\sqrt{n}} \right) = 1 - a ,$$

wobei gilt, dass

$$\begin{aligned} m_{h,h} &= \sum_{r=-\infty}^{\infty} \rho^2(r+h) + \underbrace{\rho(r-h) \cdot \rho(r+h)}_{=0} + \underbrace{2 \cdot \rho^2(h) \cdot \rho^2(r)}_{=0} - \underbrace{4 \cdot \rho(h) \cdot \rho(r) \cdot \rho(r+h)}_{=0} = \\ &= \rho^2(-q) + \rho^2(-q+1) + \dots + \rho^2(0) + \rho^2(1) + \dots + \rho^2(q) = 1 + 2 \cdot (\rho^2(1) + \dots + \rho^2(q)) . \end{aligned}$$

Der Faktor $m_{h,h}$ ist also vom wahren Wert der Autokorrelation der $MA(q)$ -Zeitreihe abhängig.

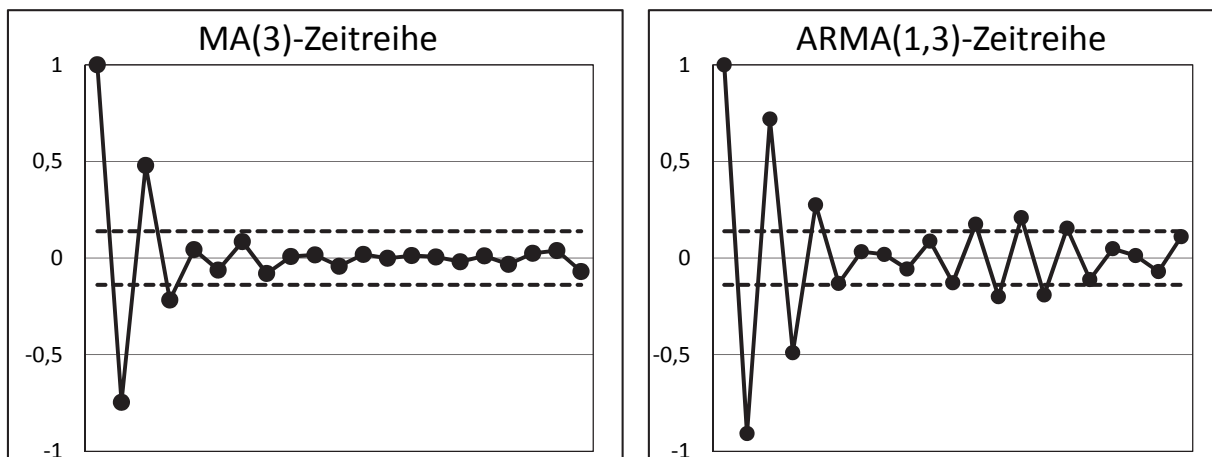
Nachdem man diese in der Praxis nicht kennt, werden für die Werte der Autokorrelationsfunktion oft einfach die Werte der empirischen Autokorrelationsfunktion eingesetzt – was aber ein ungenaues Intervall liefert, das sowohl zu groß als auch zu klein sein kann – oder aber der Faktor wird komplett weggelassen, was ein „strengeres“ Intervall liefert und zusätzlich den Vorteil hat, dass dieses nicht für jede untersuchte Ordnung q neu berechnet werden muss.

Beispiel 9.2

Die folgenden zwei Diagramme zeigen die empirische Autokorrelationsfunktion bis zum Lag 20 von Realisierungen vom Umfang 200 einer $MA(3)$ - und einer $ARMA(1,3)$ -Zeitreihe mit den folgenden $ARMA$ -Gleichungen:

- $X_t = e_t - 0,8 \cdot e_{t-1} + 0,75 \cdot e_{t-2} - 0,6 \cdot e_{t-3}$
- $Y_t = -0,7 \cdot Y_{t-1} - 0,8 \cdot e_{t-1} + 0,75 \cdot e_{t-2} - 0,6 \cdot e_{t-3} + e_t$

Bei beiden Zeitreihen wurde die Hypothese, es handle sich um eine $MA(3)$ -Zeitreihe überprüft (mit Irrtumswahrscheinlichkeit $\alpha = 0,05$ und unter Weglassen des Faktors $m_{h,h}$). Offensichtlich liefert das Verfahren die richtigen Ergebnisse, es verwirft die Hypothese für die $ARMA(1,3)$ -Zeitreihe, kann dies aber für $MA(q)$ -Prozesse mit Ordnung $q \geq 3$ nicht tun.



9.1.3 Überprüfen auf $AR(p)$ -Zeitreihen

Möchte man eine Realisierung $\widehat{X}_{[n]} := (\widehat{X}_t : t \in [1, n])$ darauf untersuchen, ob sie sich gut durch eine kausale $AR(p)$ -Zeitreihe modellieren lässt, so hilft das Betrachten der empirischen Autokorrelationsfunktion nicht wirklich weiter. Aus Satz 5.4 weiß man allerdings, dass für solche Zeitreihen deren partielle Autokorrelationsfunktion für Lags größer als p konstant Null ist und aus Satz 7.2, dass $\hat{\pi}(h)$ für diese Lags approximativ $N(0, \frac{1}{n})$ -verteilt sind.

Unter der Annahme, es handle sich bei der beobachteten Realisierung um eine $AR(p)$ -Zeitreihe kann man also für $h > p$ das Konfidenzintervall

$$\left(-\frac{z_{(1-\frac{\alpha}{2})}}{\sqrt{n}} ; \frac{z_{(1-\frac{\alpha}{2})}}{\sqrt{n}} \right)$$

erstellen, in dem bei $(1 - \alpha) \cdot 100$ Prozent der Realisierungen der Wert $\hat{\pi}(h)$ liegen muss.

Anmerkung 9.1

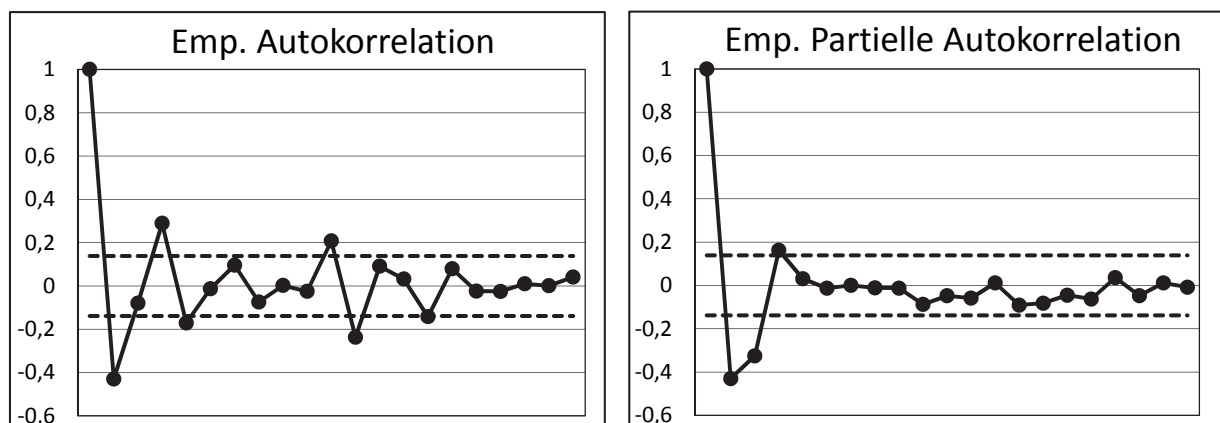
Die in 9.1.1 bis 9.1.3 vorgestellten Konfidenzintervalle für die diversen Schätzgrößen kann (und sollte) man selbstverständlich auch anders interpretieren. Wie man gesehen hat, wird in allen drei Fällen schlussendlich ein Intervall verwendet, das unabhängig vom jeweiligen Lag ist, was es, wie es auch in den Beispielen schon stillschweigend angewendet wurde, ermöglicht, anstatt einzelner Werte der Schätzfunktion gleich deren mehrere zu betrachten.

Nun ist es aber wichtig zu verstehen, dass nicht unbedingt jeder dieser Werte innerhalb der Intervallgrenzen liegen muss. Man kann im Gegenteil sagen, dass sich eben nur mindestens $(1 - \alpha) \cdot 100$ Prozent von ihnen innerhalb des Konfidenzintervalls befinden müssen, damit die jeweilige Hypothese nicht abgelehnt werden kann.

Beispiel 9.3

Im Folgenden sind die Diagramme der empirischen Autokorrelationsfunktion und der empirischen partiellen Autokorrelationsfunktion einer Realisierung der stationären $AR(3)$ -Zeitreihe

$X_t = -\frac{1}{2} \cdot X_{t-1} - \frac{1}{4} \cdot X_{t-2} + \frac{1}{8} \cdot X_{t-3} + e_t$ angeführt. Man sieht sehr deutlich anhand des linken Diagrammes, dass man die Zeitreihe nicht angemessen durch ein $MA(q)$ -Modell niedriger Ordnung modellieren kann. Anhand des rechten Diagrammes sieht man, dass man die Hypothese, es handle sich um eine $AR(3)$ -Zeitreihe (zurecht) mit einer Irrtumswahrscheinlichkeit $\alpha = 0,05$ nicht ablehnen kann.



Anmerkung 9.2

Obwohl es natürlich möglich ist, für eine Realisierung vom Umfang n die empirischen (partiellen) Autokorrelationen bis zum Lag $n - 1$ zu berechnen, ist die Untersuchung der Schätzfunktion für sehr große Lags nicht sonderlich sinnvoll, einerseits da für deren Berechnung schlicht zu wenige Werte vorliegen und somit sämtliche asymptotische Aussagen über die empirische (partielle) Autokorrelation in keiner Weise mehr gültig sind, andererseits da man, vorausgesetzt natürlich es handelt sich bei der untersuchten Zeitreihe tatsächlich um die Realisierung einer $ARMA$ -Zeitreihe, gemäß Satz 5.2 ohnehin davon ausgehen kann, dass die Werte für große Lags verschwinden werden.

9.2 Das Akaike-Informationskriterium

Wie man an den vorangegangenen Beispielen sehen konnte, ist es also möglich durch Betrachtung der empirischen (partiellen) Autokorrelationsfunktion oftmals gut zur Modellierung geeignete Ordnungen zu finden.

Dies gilt aber nur, wenn es sich bei der beobachteten Zeitreihe um Weißes Rauschen oder um entweder eine $AR(p)$ - oder eine $MA(q)$ -Zeitreihe handelt, für allgemeinere $ARMA(p, q)$ -Prozesse ist dies nur sehr bedingt möglich.

Aus Satz 5.2 weiß man, dass die Werte der Autokorrelationsfunktion einer kausalen $ARMA$ -Zeitreihe für wachsende Lags stets geometrisch abklingen, was aufgrund der asymptotischen Erwartungstreue der empirischen Autokorrelationsfunktion in gewisser Weise auch für diese gilt. Somit ist für ausreichend großen Umfang der Stichprobe mit Sicherheit gewährleistet einen Lag q finden zu können, ab dem sich die Werte der empirischen Autokorrelationsfunktion (etwaige statistische Ausreißer ausgenommen, siehe Anmerkung 9.1) innerhalb des in 9.1.2 ermittelten Intervalls befinden, wodurch man die Hypothese, es handle sich um eine $MA(q)$ -Zeitreihe, nicht mehr ablehnen kann.

Durch ein derartiges Vorgehen stößt man aber in der Regel sehr schnell auf das in der Statistik allgegenwärtige Problem des „Overfittings“ – also die Gefahr durch das Wählen zu vieler Parameter ein zu „starres“ Modell zu erhalten, das zwar sehr gut zur vorliegenden Realisierung passt, aber kaum noch zu den dieser zugrunde liegenden Strukturen.

Der geläufigste Weg mit diesem Problem bei der Ordnungswahl für $ARMA(p, q)$ -Zeitreihen umzugehen ist der Vergleich verschiedener Parameterkombinationen mittels des *Akaike Information Criteria*, kurz *AIC*. Dies funktioniert im Allgemeinen, nicht speziell auf das Themenfeld der Zeitreihenanalyse bezogen, folgendermaßen:

Stehen verschiedene Parameteranzahlen K_i zur Auswahl, so wird für diese jeweils mittels *Maximum-Likelihood*-Schätzung ein Modell erstellt und die Wahrscheinlichkeit der vorliegenden Realisierung (die man ja im Zuge des Verfahrens maximieren muss) unter der Annahme, das Modell wäre korrekt, notiert. Sei nun L_i diese Wahrscheinlichkeit, so berechnet sich für jedes i der Wert des *AIC* mittels

$$AIC(i) = -2 \cdot \log(L_i) + 2 \cdot K_i \quad .$$

Die am besten zur Modellierung geeignete Parameteranzahl K_i ist nun jene, für die $AIC(i)$ minimal ist. Während bei wachsender Parameteranzahl L_i üblicherweise immer größer wird und somit $-2 \cdot \log(L_i)$ der 0 „von oben“ immer näher kommt, wird der „Korrekturfaktor“ $2 \cdot K_i$ immer größer, was ein „Overfitting“ verhindern soll.

Die genaue Herleitung dieses Verfahrens wird hier schuldig geblieben werden, interessierte LeserInnen seien hierfür beispielsweise auf Sadanori *Konishi*, Genshiro *Kitagawa*, *Information Criteria and Statistical Modeling* (Springer, 2008) S.29-60 verwiesen.

Versucht man nun dieses allgemeine statistische Verfahren an die Erfordernisse der Zeitreihenanalyse mittels kausaler $ARMA(p, q)$ -Modelle anzupassen, so gestaltet sich dies denkbar einfach. Man erstellt für mögliche Kombinationen von p und q mittels der Methoden aus Kapitel 8 *Maximum-Likelihood*-Schätzungen und ermittelt deren AIC mit Hilfe der Formel

$$AIC(p, q) = -2 \cdot \ln(f(\hat{X}_n)) + 2 \cdot (p + q + 1),$$

wobei $f(\hat{X}_n)$ der Wert der Dichtefunktion des für gegebene Parameter p und q ermittelten *Maximum-Likelihood*-Modells bei der beobachteten Realisierung ist (siehe Satz 8.4). Offenbar ist die Anzahl der geschätzten Parameter gleich $p + q + 1$, wobei die Addition der "1" daher kommt, dass ja nicht nur die Koeffizienten des AR - und des MA -Polynoms, sondern auch die Varianz des Weißen Rauschens geschätzt werden muss.

Während die Verwendung des AIC -Kriteriums durchaus gängig ist, hat sich im Lauf der Zeit ergeben, dass dieses weiterhin zu „Overfitting“ neigt, weshalb oftmals statt des AIC das so genannte *Akaike Information Corrected Criterion* ($AICC$) verwendet wird. Dieses ist gerade für in der Praxis vorkommende, verhältnismäßig kleine, Stichprobenanzahlen aussagekräftiger und berechnet sich für kausale $ARMA(p, q)$ -Zeitreihen mittels

$$AICC(p, q) = -2 \cdot \ln(f(\hat{X}_n)) + 2n \cdot \frac{p + q + 1}{n - p - q - 2},$$

wobei, wie man sieht, lediglich der „Korrekturfaktor“ verändert wurde, sodass dieser die Wahl hoher Ordnungen im Vergleich zum klassischen AIC stärker „bestraft“.

Gesamt kann man also zum Vorgang der Ordnungswahl folgendes Vorgehen empfehlen: Zunächst sollte die beobachtete Realisierung stets mittels 9.1.1 -9.1.3 darauf überprüft werden, ob es sich dabei um ein Weißes Rauschen, eine $MA(q)$ - oder eine $AR(p)$ -Zeitreihe mit $p, q \leq 2$ handeln könnte. Sollte dies eindeutig der Fall sein, kann man auf das Berechnen des AIC oder des $AICC$ verzichten.

Sollte die Modellierung mittels solcher AR - oder AR -Modelle geringer Ordnung sich nicht anbieten, oder deren Ordnung nicht eindeutig sein, so kommt man nicht umhin für verschiedene mögliche Ordnungen p und q Modelle zu erstellen, deren AIC zu berechnen und wiederum jenes Modell zu wählen, für die das AIC minimal ist.

Hierbei empfiehlt es sich zunächst für alle anhand der Methoden aus 9.1.2 und 9.1.3 ermittelten möglichen Ordnungen p und q die zugehörigen $AR(p)$ - und $MA(q)$ -Modelle und zusätzlich $ARMA(r, s)$ -Modelle für alle $r, s \neq 0$ und $r + s < \max(p, q)$ zu errechnen und diese anschließend mittels des AIC zu vergleichen.

Sollte beispielsweise anhand der empirischen und der empirischen partiellen Autokorrelationsfunktion ein $MA(2)$ -, ein $MA(4)$ - oder ein $AR(3)$ -Modell vorstellbar sein, so sollte man neben diesen auch noch $ARMA(p, q)$ -Modelle mit $(p, q) \in \{(1,1), (2,1), (1,2)\}$ erstellen.

Dies ist sinnvoll, da die Verwendung eines „gemischten“ Modelles die Anzahl der zu schätzenden Parameter reduzieren kann. Diese besondere Stärke des $ARMA$ -Modelles wird später in Beispiel 10.3 klar zutage treten.

9.3 Überprüfen des Modells

Hat man nun ein nach den bisherigen angeführten Maßstäben optimales *ARMA*-Modell für eine Zeitreihe gefunden, muss es noch daraufhin überprüft werden, ob dieses wirklich zur Beschreibung der Zeitreihe und im speziellen für Vorhersagen überhaupt geeignet ist – bisher wurde ja lediglich untersucht, ob unter der Annahme ein Modell wäre richtig, die Wahrscheinlichkeit, dass bei Durchführen eines Zufallsexperiments die beobachtete Zeitreihe entsteht, auch ausreichend groß ist.

Geht man nun davon aus, dass das mittels *Maximum-Likelihood*-Methode ermittelte *ARMA*(p, q)-Modell mit Koeffizienten $a_1, \dots, a_p, b_1, \dots, b_q, \sigma_e^2$ tatsächlich der „wahre“ der Realisierung $\widehat{X}_{[n]} := (\hat{X}_t : t \in [1, n])$ zugrundeliegende stochastische Prozess ist und ermittelt man anhand dieses Modells die Ein-Schritt-Vorhersagen $\tilde{X}_1, \dots, \tilde{X}_n$ und die Ein-Schritt-Vorhersagefehler $r_0, \dots, r_{n-1}$, so lässt sich die Güte des Modelles durch Betrachtung der Zeitreihe

$$Y_t = \frac{X_t - \tilde{X}_t}{\sqrt{r_{t-1}}}$$

für $1 \leq t \leq n$ bestimmen, wobei man die einzelnen Y_t auch als *Residuen* bezeichnet.

Wenn das verwendete kausale *ARMA*-Modell tatsächlich das der Realisierung zugrundeliegende ausreichend gut beschreibt, so müssen die Ein-Schritt-Vorhersagen $\tilde{X}_t$ den Wert X_t bis auf einen rein zufälligen, nicht von vorherigen oder zukünftigen Zeitpunkten abhängigen, Teil $X_t - \tilde{X}_t$ vorhersagen können, woraus folgt, dass

$$\bullet \quad \text{Cov}(Y_t, Y_{t+h}) = \frac{1}{r_{t-1}} \cdot \text{Cov}(X_t - \tilde{X}_t, X_{t+h} - \tilde{X}_{t+h}) = 0 \quad \forall h \neq 0 \quad .$$

Außerdem müsste gelten, dass

$$\begin{aligned} \bullet \quad \mathbb{E}(Y_t) &= \frac{1}{\sqrt{r_{t-1}}} \cdot (\mathbb{E}(\tilde{X}_t) - \mathbb{E}(X_t)) = \frac{1}{\sqrt{r_{t-1}}} \cdot (\mathbb{E}(\sum_{j=1}^{t-1} d_{t-1,j} \cdot X_{t-j}) - \mathbb{E}(X_t)) = 0 \\ \bullet \quad \text{Var}(Y_t) &= \frac{1}{r_{t-1}} \cdot \text{Var}(X_t - \tilde{X}_t) = \frac{v_{n-1}}{r_{t-1}} = \frac{\sigma_e^2 \cdot r_{t-1}}{r_{t-1}} = \sigma_e^2 \quad . \end{aligned}$$

Man sieht also, dass alle Voraussetzungen aus Definition 3.1 erfüllt sind, es sich also bei Y um ein Weißes Rauschen mit Varianz σ_e^2 und damit natürlich bei $\hat{Y}$ um eine Realisierung eines solchen Weißen Rauschens handeln müsste.

Die einfachste Art und Weise dies zu überprüfen ist nun den Graphen der Residuen auf auffällige Abweichungen, die auf nicht-Stationarität hindeuten würden, zu untersuchen, also auf Trend, Saisonalität oder sich verändernde Varianzen, und anschließend gemäß 9.1.1 die Hypothese, es handle sich um ein Weißes Rauschen, zu überprüfen.

Stellt es sich heraus, dass es sich bei den empirischen Residuen tatsächlich um die Realisierung eines Weißen Rauschens handeln könnte, so kann man nun das ermittelte *ARMA*(p, q)-Modell als passend für die vorliegende Zeitreihe bezeichnen. Sollte dies hingegen nicht der Fall sein, kann man das erstellte Modell getrost verwerfen und sollte entweder andere Ordnungen p, q in Betracht ziehen (idealerweise jene, für die das *AIC*(C) am zweitkleinsten ist) oder die ermittelten, nicht passenden Koeffizienten bei einem erneuten Durchlauf des *Maximum-Likelihood*-Verfahrens explizit ausschließen.

9.4 Erstellen von Zukunftsprognosen

Hat man gemäß der vorangegangenen Ausführungen für eine vorliegende Realisierung einer Zeitreihe ein möglichst gut passendes *ARMA*-Modell erstellt und ist dieses invertierbar – es sei nochmals angemerkt, dass man die Invertierbarkeit des mittels *Maximum-Likelihood*-Verfahrens gewonnenen Modelles stets garantieren kann, indem man Lösungen für die Parameter, die die Bedingungen von Satz 4.4 nicht erfüllen, einfach nicht zulässt – so lassen sich bekanntlich die einzelnen Werte $\hat{e}_t$ der vorliegenden Realisierung des zugrundeliegenden Weißen Rauschens anhand der Werte der Realisierung der Zeitreihe approximativ berechnen.

Nun kann man anhand dieser Daten die Hypothese, das Weiße Rauschen e sei normalverteilt, mittels diverser, hier nicht näher beschriebenen Testverfahren, wie beispielsweise mit dem *Shapiro-Wilk-Test* oder dem *Kolmogorow-Smirnow-Test* überprüfen. Wenn man diese nicht ablehnen kann, so ermöglicht dies die Erstellung von approximativen Konfidenzintervallen für den jeweils nächsten – noch unbekannten – Wert der Realisierung der Zeitreihe.

Für ausreichend viele Daten ist nämlich offensichtlich, dass für ebenfalls ausreichend großes t gilt, dass $X_t - \bar{X}_t \approx e_t$. Zusätzlich gilt trivialerweise, dass $\frac{e}{\sigma_e} \sim N(0,1)$ -verteilt ist. Somit erhält man für ausreichend großes t

$$\mathbb{P}\left(-z_{(1-\frac{a}{2})} \leq \frac{\hat{e}_t}{\sigma_e} \leq z_{(1-\frac{a}{2})}\right) = 1 - a \Rightarrow \mathbb{P}\left(-z_{(1-\frac{a}{2})} \leq \frac{\hat{X}_t - \bar{X}_t}{\sigma_e} \leq z_{(1-\frac{a}{2})}\right) \approx 1 - a .$$

Hieraus erhält man nun trivialerweise für den unbekannten Wert $\hat{X}_t$ das approximative $(1 - a)$ -Konfidenzintervall

$$I = \left[\bar{X}_t - \sigma_e \cdot z_{(1-\frac{a}{2})} ; \bar{X}_t + \sigma_e \cdot z_{(1-\frac{a}{2})} \right]$$

Man kann nun also etwas salopp sagen, dass die Wahrscheinlichkeit, dass der nächste Wert der beobachteten Zeitreihe innerhalb von I liegen wird, etwa $(1 - a) \cdot 100$ Prozent beträgt.

Obwohl es selbstverständlich möglich ist Vorhersagen nicht nur für den jeweils nächsten unbekannten Wert einer Zeitreihe zu treffen, sondern auch wesentlich weiter „in die Zukunft zu blicken“, geht dies trivialerweise mit einer starken Verschlechterung der Verlässlichkeit dieser Prognosen einher, weshalb diesbezügliche Methoden hier nicht näher behandelt werden.

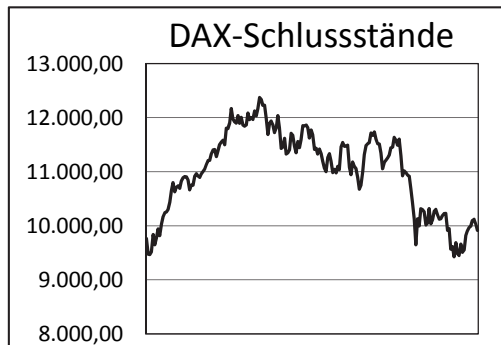
Es sei hier beispielsweise auf Peter J. Brockwell, Richard A. Davis, *Time Series: Theory and Methods* (Springer, 1991) S. 174f, 179-182 verwiesen.

Mit der soeben vorgestellten Methode zur Erstellung von Ein-Schritt-Zukunftsprognosen sind die theoretischen Ausführungen dieser Diplomarbeit an ihrem Ende angekommen, der/die LeserIn sollte nun in der Lage sein, anhand vorliegender Rohdaten einer Realisierung einer Zeitreihe ein möglichst gut dazu passendes *ARMA*-Modell zu erstellen und anhand dessen Zukunftsprognosen zu errechnen. Zum besseren Verständnis der behandelten Themen und zur Demonstration der Wirksamkeit der beschriebenen Verfahren sollen nun im folgenden Kapitel abschließend drei Anwendungsbeispiele ausführlich behandelt werden.

10. Anwendungsbeispiele

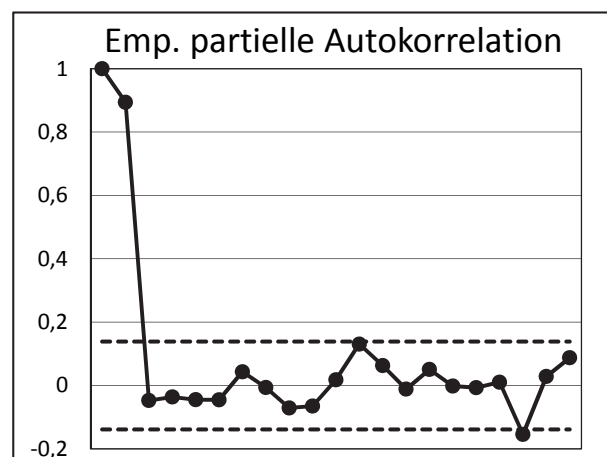
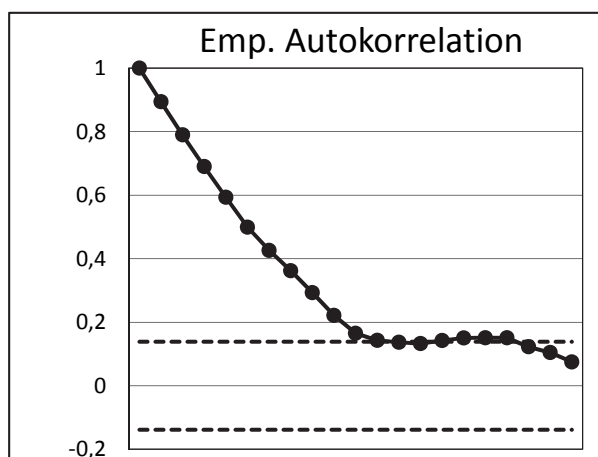
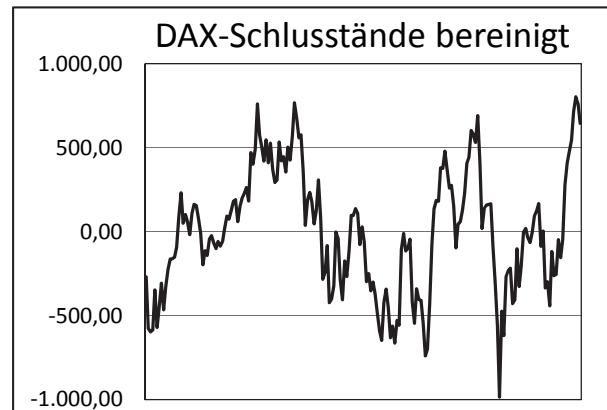
Beispiel 10.1

Gegenstand der Betrachtung ist hier der Verlauf der täglichen Schlussstände des Deutschen Aktienindex zwischen 14.11.2014 und 02.09.2015. Die Daten hierfür wurden entnommen von <http://www.finanzen.net/index/DAX/Historisch>.



Wie man links sehen kann, ist ein klarer Trend erkennbar, der durch eine quadratische Funktion modelliert wurde. Auch die „saisonalen“ Anteile wurden berechnet, wobei sich ergab, dass die Kursstände tendenziell Montags eher durchschnittlich sind, Dienstag und Mittwoch abfallen, um dann Donnerstag und Freitag wieder zu steigen (an Samstagen und Sonntagen ist die Börse geschlossen, weshalb für diese Tage auch keine Daten vorhanden sind).

Nach dieser Trend- und Saisonbereinigung ergab sich schließlich nebenstehende Zeitreihe, mit der von nun an weitergearbeitet wurde. Anschließend wurden die empirischen Autokorrelationen und die empirischen partiellen Autokorrelationen berechnet und die Zeitreihe daraufhin untersucht, ob sich für deren Modellierung ein einfaches AR - oder MA -Modell anbietet. Wie man sehen kann nimmt die empirische Autokorrelation sehr langsam ab, weshalb an Modellierung mittels einer $MA(q)$ Zeitreihe nicht zu denken ist.



Betrachtet man hingegen die empirische partielle Autokorrelationsfunktion, so sieht man recht deutlich, dass es sinnvoll wäre diese Zeitreihe als $AR(1)$ -Prozess zu modellieren (obwohl die empirische partielle Autokorrelation zum Lag 18 außerhalb des Konfidenzintervalls liegt – besagt dieses Intervall doch lediglich, dass mindestens 95% der Werte innerhalb derer Schranken liegen müssen, was hier trotz des „Ausreißers“ immer noch der Fall ist).

Anschließend wurde nun nach einem passenden Parameter a für die Zeitreihe $X_t = e_t + a \cdot X_{t-1}$ gesucht. Hierfür wären zwei Methoden möglich: Einerseits weiß man aus Satz 3.4, dass folgende Beziehung existiert: $\rho(1) = a$.

Man könnte nun einfach für den Wert von a die empirische Autokorrelation zum Lag 1 verwenden, wodurch man $a = 0,894$ erhalten würde.

Verwendet man hingegen das *Maximum-Likelihood*-Verfahren, so erhält man für *AR*(1)-Zeitreihen (siehe auch Beispiel 8.2) für die Berechnung der empirischen Ein-Schritt-Vorhersagen die einfache Formel $\bar{X}_{t+1} = a \cdot \hat{X}_t$. Für die Ein-Schritt-Vorhersagefehler der gemäß Lemma 8.2 transformierten Zeitreihe Y erhält man nun nach Satz 8.2

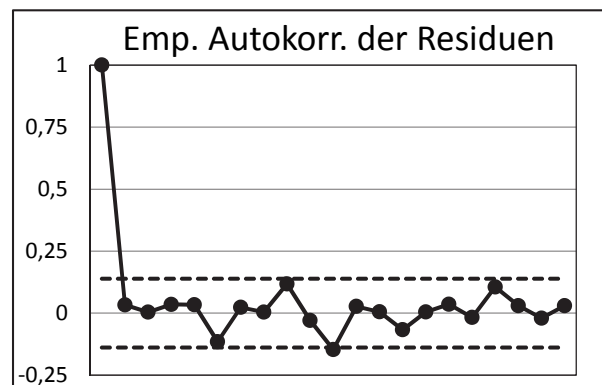
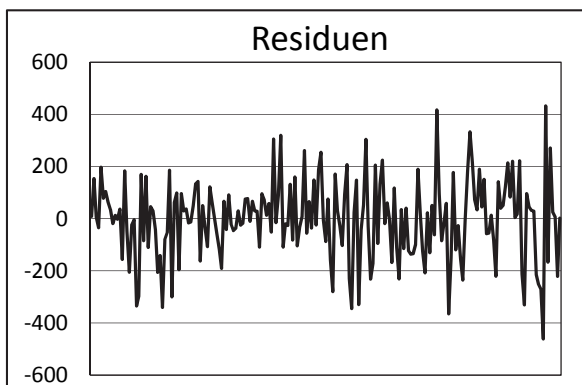
$$\begin{aligned} \bullet \quad r_0 &= \text{Cov}(Y_1, Y_1) = \frac{1}{\sigma_e^2} \cdot \gamma(0) \stackrel{\text{Satz 3.4}}{=} \frac{1}{1-a^2} \\ \bullet \quad r_n &= \text{Cov}(Y_{n+1}, Y_{n+1}) - \underbrace{\sum_{j=0}^{n-1} f_{n,n-j}^2 \cdot r_j}_{\text{Lemma 8.3} \Rightarrow 0} = \sum_{k=0}^n b_k^2 = 1^2 = 1 \quad \forall n \geq 1 \end{aligned}$$

Es wurde nun der bereits angeführte Wert $a = 0,894$ als erster Schätzwert verwendet und für ähnliche Werte die Größe M aus Satz 8.5 berechnet. Man erhält hierfür Folgendes:

a	0,875	0,88	0,885	0,89	0,895	0,9	0,905	0,91	0,915
M	10,0519	10,0498	10,0481	10,0466	10,0455	10,0446	10,0441	10,0439	10,0440

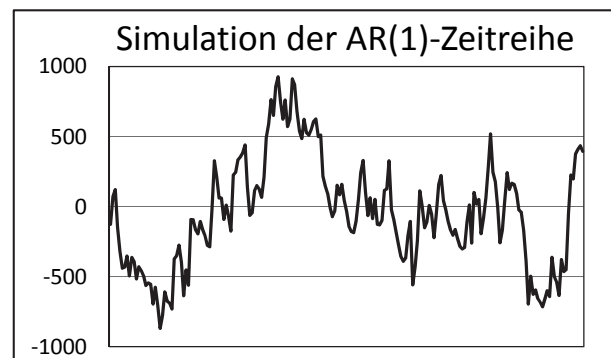
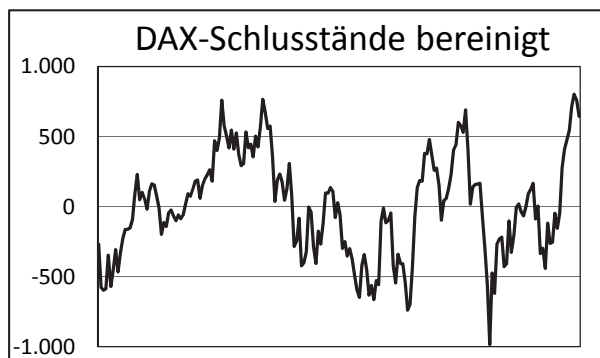
Das Minimum nimmt M also zwischen 0,905 und 0,915 an. Nochmaliges Durchprobieren der dazwischen liegenden Werte liefert $a = 0,911$ als *Maximum-Likelihood*-Schätzer. Ebenfalls gemäß Satz 8.5 erhält man nun $\sigma_e^2 = 22811,68$ und somit $\sigma_e = 151,04$.

Nun soll noch überprüft werden, ob das Modell überhaupt passend ist. Hierzu werden die Residuen gemäß Kapitel 9.3 betrachtet. Untersucht man die empirische Autokorrelationsfunktion, so erkennt man auch keinen Widerspruch in der Hypothese, es handle sich um Weißes Rauschen.



Gesamt kann man nun die beobachtete (trend- und saisonbereinigte) Zeitreihe der DAX-Kurse durch die stationäre *AR*(1)-Zeitreihe $X_t = 0,911 \cdot X_{t-1} + e_t$ mit $\sigma_e^2 = 22811,68$ beschreiben.

Stellt man nun eine Realisierung dieses Modelles der untersuchten Zeitreihe gegenüber, so kann man schon mit bloßem Auge gewisse Übereinstimmungen im Verhalten der beiden Prozesse erkennen, auch wenn deren gesamter Verlauf, dem Zufall geschuldet, natürlich völlig unterschiedlich ist.



Nun kann man anhand des ermittelten Modelles leicht Ein-Schritt-Vorhersagen erstellen.

Für den – nicht beobachteten und somit als zukünftig anzunehmenden - Wert der Realisierung bei $t = 201$ erhält man die empirische Ein-Schritt-Vorhersage

$$\bar{X}_{201} = 0,911 \cdot \hat{X}_{200} = 0,911 \cdot (-319,48) = -291,05 .$$

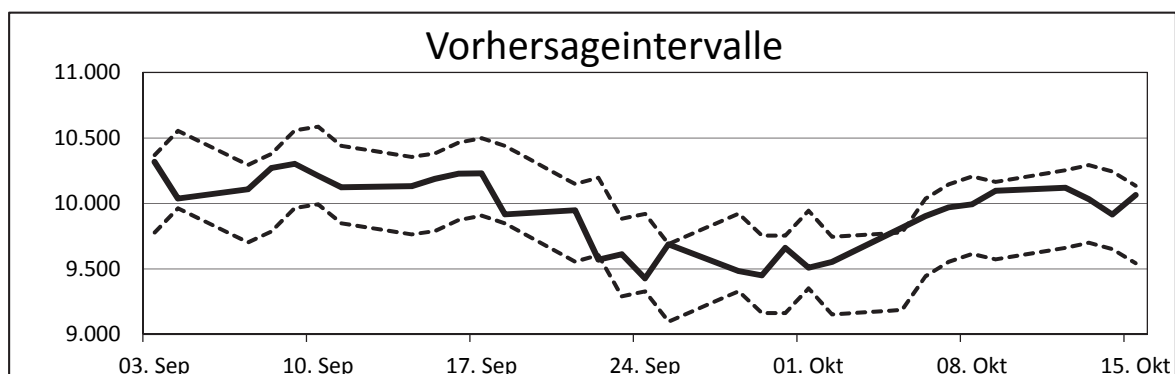
Nachdem es sich beim ermittelten Modell um eine $AR(1)$ -Zeitreihe handelt (und diese ja trivialerweise invertierbar sein muss), lassen sich die einzelnen Realisierungen des Weißen Rauschens für $t \geq 2$ mittels $\hat{\epsilon}_t = \hat{X}_t - 0,911 \cdot \hat{X}_{t-1}$ aus den vorliegenden Daten berechnen. Nun kann man die so erhaltenen Realisierungen des Weißen Rauschens $\hat{\epsilon}_2, \dots, \hat{\epsilon}_{200}$ mittels verschiedenen Testverfahren auf Normalität überprüfen. In diesem Fall wurde der Shapiro-Wilk-Test angewendet, der die Hypothese, es handle sich um normalverteilte Zufallsvariablen, mit Irrtumswahrscheinlichkeit 5% nicht verwerfen konnte. Somit kann man nun, vorausgesetzt das verwendete Modell entspricht der Realität, sagen, dass der (unbekannte) Wert der trend- und saisonbereinigten Zeitreihe der DAX-Schlusskurse am 03.09.2015 mit 95%-iger Wahrscheinlichkeit in folgendem Intervall liegen sollte:

$$[-291,05 - 151,04 \cdot 1,96 ; -291,05 + 151,04 \cdot 1,96] = [-587,09 ; 4,99] .$$

Dies allein sagt aber nun noch nicht allzu viel über den tatsächlich erwarteten Wert des DAX-Schlussstandes vom 03.09.2015 aus. Um diesen zu ermitteln, muss man nun zu dem gerade ermittelten Intervall die zuvor bereinigten Trend- und Saisonanteile hinzurechnen und man erhält für den Börsenkurs an diesem Tag das Intervall

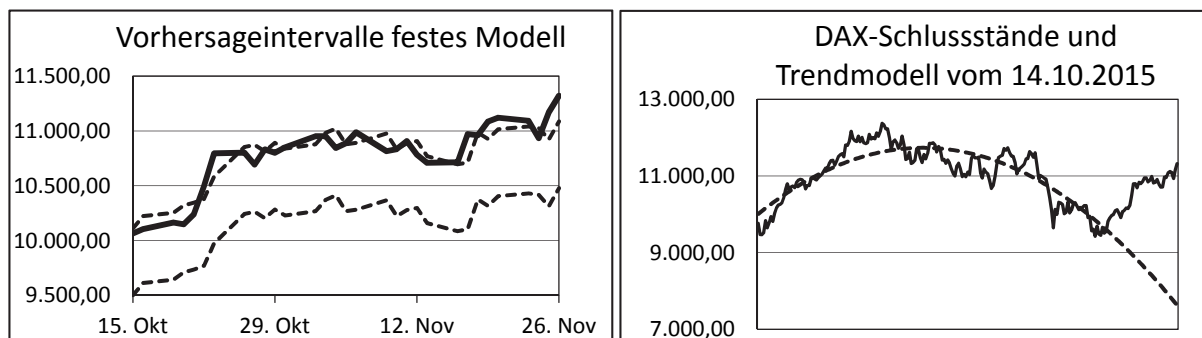
$$[-587,09 + T_{201} + S_{201} ; 4,99 + T_{201} + S_{201}] = [9742,52 ; 10334,59] .$$

Der tatsächliche Wert des DAX an diesem Tag betrug nun 10317,84, liegt also tatsächlich (knapp) innerhalb dieses Intervalls. Verwendet man nun das einmal erstellte Modell um Tag für Tag Zukunftsprognosen für die nächsten 30 Arbeitstage zu erstellen, so erhält man, wie man deutlich sehen kann, durchaus gute Ergebnisse.



Bei gesamt 30 Ein-Schritt-Prognosen liegen lediglich 2 tatsächlich aufgetretene Börsenkurse außerhalb der ermittelten Intervalle und selbst diese nur äußerst knapp – das erstellte Modell darf also als durchaus gut geeignet angesehen werden. Berechnet man an jedem Arbeitstag nun mittels des neu aufgetretenen tatsächlichen Wertes ein komplett neues Modell, so liefert dieses in diesem Fall kein signifikant anderes Ergebnis – das muss aber nicht immer der Fall sein.

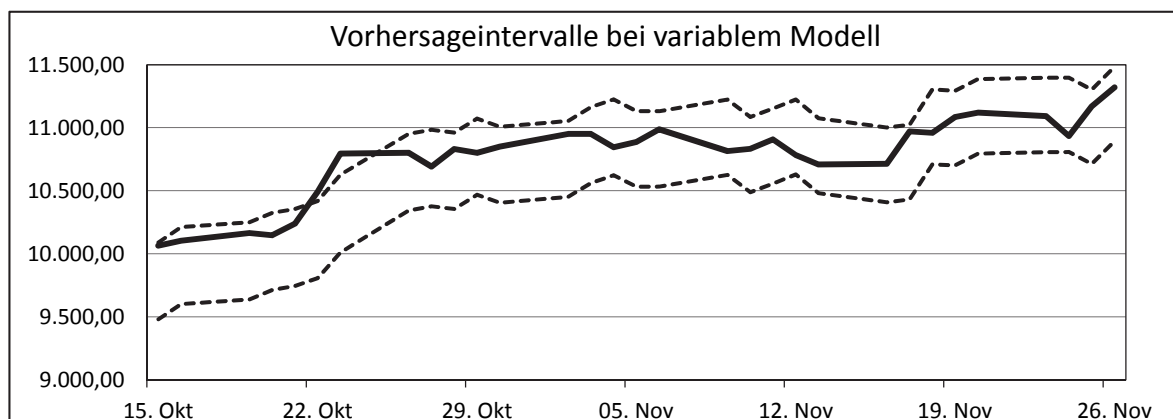
Erstellt man beispielsweise das oben ausgearbeitete Modell nicht für die DAX-Schlussstände zwischen 14.11.2014 und 02.09.2015, sondern stattdessen für jene zwischen 02.01.2015 und 14.10.2015 und verwendet man für die Vorhersage der nächsten 30 Werte durchgehend ein am 14.10. erstelltes Modell, so sieht man an rechtsstehendem Diagramm, dass dies plötzlich zu höchst unbefriedigenden Ergebnissen führt – nur knapp über die Hälfte der tatsächlichen Börsenkurse liegt in den jeweiligen Vorhersageintervallen.



Der Grund hierfür ist im Wesentlichen eine nicht mit den Methoden der Zeitreihenanalyse vorhersagbare Trendumkehr in der Kursentwicklung des DAX – man sieht dies sehr deutlich an der rechts oben abgebildeten Gegenüberstellung der unbehandelten Zeitreihe bis 26.11. mit dem am 14.10. modellierten Trend.

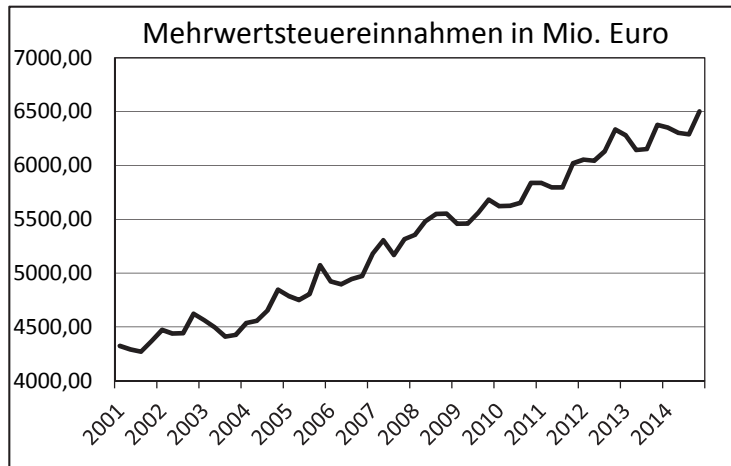
Es ist also im Allgemeinen unabdingbar nach jedem neu beobachteten Wert einer Zeitreihe ein von Grund auf neues Modell für diese zu erstellen. Während sich anfänglich das Versagen der Modellierung des Trends mittels einer quadratischen Funktion noch durch Zufall erklären lässt, sollte spätestens am 22.10. klar werden, dass man diesen wohl besser mit einem Polynom dritten Grades modellieren sollte.

Erstellt man nun Tag für Tag ein komplett neues *ARMA*-Modell, so führt dies in diesem Fall, wie schon auf den ersten Blick auf das unten abgebildete Diagramm ersichtlich ist, zu signifikant besseren Vorhersagen – nur an 2 Tagen liegt der tatsächliche DAX-Kurs außerhalb des ermittelten Intervalls.



Beispiel 10.2

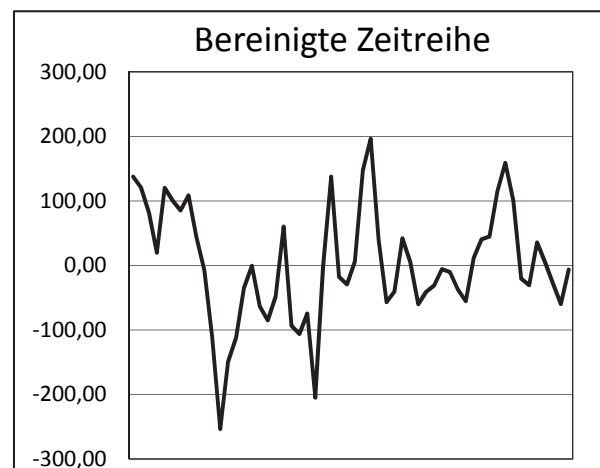
In diesem Beispiel soll die Entwicklung der Einnahmen der Republik Österreich durch die Mehrwertsteuer näher betrachtet werden. Die Daten hierfür liegen seit Anfang des Jahres 2001 quartalsweise vor und wurden von http://statistik.at/web_de/statistiken/wirtschaft/oeffentliche_finanzen_und_steuern/oeffentliche_finanzen/einnahmen_und_ausgaben_des_staates/020005.html entnommen und bis Ende des Jahres 2014 zur Erstellung eines Modelles verwendet.



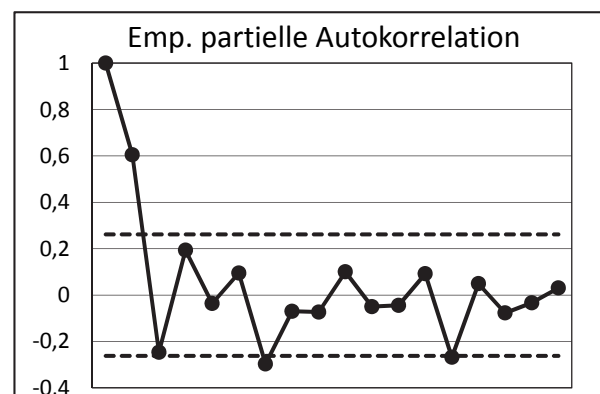
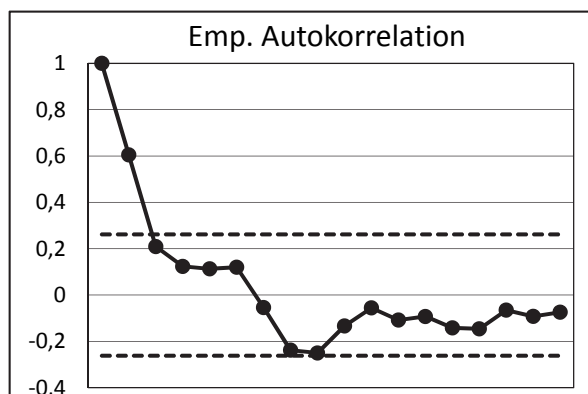
Wie man nebenstehendem Diagramm entnehmen kann, weist diese Zeitreihe einen einfachen Trend auf, der mittels einer linearen Funktion modelliert wurde. Bei der Betrachtung saisonaler Komponenten konnte festgestellt werden, dass die Steuereinnahmen im ersten und vierten Quartal eines Jahres tendenziell höher sind als in der Jahresmitte.

Bereinigt man die gegebene Zeitreihe nun um ihren Trend und ihre saisonale Komponenten, so erhält man die rechts abgebildete Zeitreihe, der man Stationarität unterstellen kann.

Betrachtet man nun die empirische Autokorrelationsfunktion und die empirische partielle Autokorrelationsfunktion dieser Zeitreihe, so kann man erkennen, dass sich ihre Modellierung mittels einer $MA(1)$ -Zeitreihe anbietet, da ihre empirische Autokorrelationsfunktion für Lags größer als 1 verschwindet.



Betrachtet man hingegen die empirische partielle Autokorrelationsfunktion, so sieht man, dass sich die Zeitreihe wohl eher nicht durch einen $AR(p)$ -Prozess mit kleinem p darstellen lässt, da hier insgesamt 3 Werte außerhalb des eingezeichneten Intervalls liegen.



Im Vergleich zu vorigem Beispiel benötigt nun die Modellbildung mittels einer $MA(1)$ -Zeitreihe

$$X_t = e_t + b_1 \cdot e_{t-1}$$

einiges mehr an theoretischer Arbeit.

Zunächst weiß man aus dem Beweis von Satz 3.1 Folgendes über die Autokovarianzfunktion eines solchen Prozesses, es gilt nämlich

- $\gamma(0) = \sigma_e^2 \cdot (1 + b_1^2)$
- $\gamma(1) = \sigma_e^2 \cdot b_1$
- $\gamma(h) = 0 \quad \forall h > 1$.

Mittels Satz 8.2 kann man nun die Kovarianzen der gemäß Lemma 8.2 transformierten Zeitreihe berechnen:

- $Cov(Y_1, Y_1) = \frac{1}{\sigma_e^2} \cdot \gamma(0) = \frac{1}{\sigma_e^2} \cdot \sigma_e^2 \cdot (1 + b_1^2) = 1 + b_1^2$
- $Cov(Y_1, Y_2) = Cov(Y_2, Y_1) = \frac{1}{\sigma_e^2} \cdot \gamma(1) = \frac{1}{\sigma_e^2} \cdot \sigma_e^2 \cdot b_1 = b_1$
- $Cov(Y_t, Y_{t+1}) = b_0 \cdot b_1 + b_1 \cdot b_2 = b_1 \quad \forall t > 1$
- $Cov(Y_t, Y_t) = b_0^2 + b_1^2 = 1 + b_1^2 \quad \forall t > 1$

Anhand dieser Informationen kann man nun mittels des Innovationsalgorithmus die Ein-Schritt-Vorhersagefehler r_n und die Koeffizienten $f_{i,j}$ der Zeitreihe Y berechnen. Es gilt

- $r_0 = Cov(Y_1, Y_1) = 1 + b_1^2$
- $f_{1,1} = \frac{1}{r_0} \cdot Cov(Y_2, Y_1) = \frac{b_1}{1+b_1^2}$
- $r_1 = Cov(Y_2, Y_2) - f_{1,1}^2 \cdot r_0 = 1 + b_1^2 - \frac{b_1^2}{(1+b_1^2)^2} \cdot (1 + b_1^2) = \frac{b_1^4 + b_1^2 + 1}{1+b_1^2}$
- Lemma 8.3 $\Rightarrow f_{2,2} = 0$
- $f_{2,1} = \frac{1}{r_1} \cdot \left(Cov(Y_3, Y_2) - \underbrace{f_{1,1} \cdot f_{2,2}}_{=0} \cdot r_0 \right) = b_1 \cdot \frac{b_1^2 + 1}{b_1^4 + b_1^2 + 1}$
- $r_2 = Cov(Y_3, Y_3) - f_{2,1}^2 \cdot r_1 = b_1^2 + 1 - b_1^2 \cdot \frac{b_1^2 + 1}{b_1^4 + b_1^2 + 1}$
- Lemma 8.3 $\Rightarrow f_{3,3} = f_{3,2} = 0$.

Man erhält schlussendlich folgende rekursive Berechnungsmethode:

- $r_0 = 1 + b_1^2$
- $f_{n,1} = \frac{b_1}{r_{n-1}}$
- $r_n = b_1^2 + 1 - b_1 \cdot f_{n,1}$.

Gemäß Satz 8.3 lassen sich nun die Ein-Schritt-Vorhersagen der nicht transformierten Zeitreihe X folgendermaßen bestimmen:

$$\tilde{X}_{n+1} = f_{n,1} \cdot (X_n - \tilde{X}_n)$$

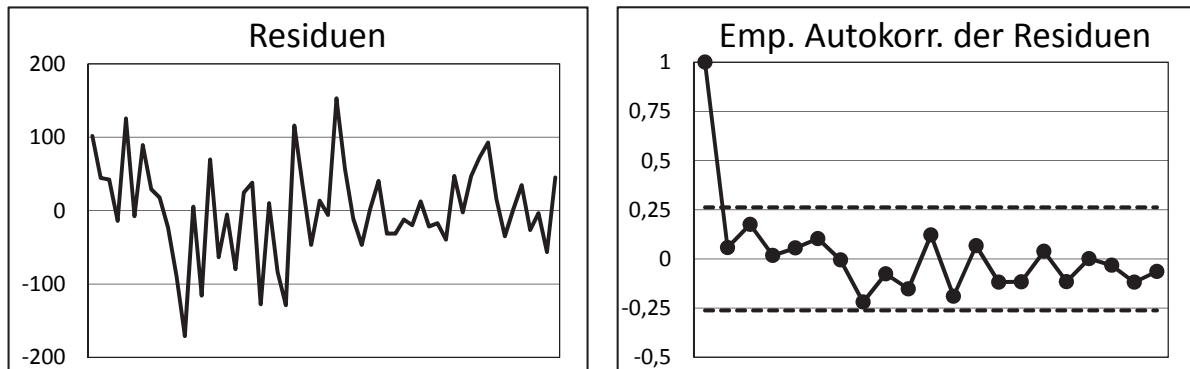
Bevor man sich nun daran macht eine *Maximum-Likelihood*-Lösung zu erstellen, ist es sinnvoll sich Gedanken zu machen, welche Lösungen man überhaupt zulässt. Aus Satz 4.5 weiß man, dass $MA(q)$ -Zeitreihen immer kausal sind, weshalb man hierfür keine Einschränkungen möglicher Werte für den Parameter b_1 treffen muss.

Nachdem es aber zur Erstellung von Konfidenzintervallen für die Ein-Schritt-Vorhersagen unumgänglich ist ein invertierbares Modell zu verwenden, darf gemäß Satz 4.4 das *MA*-Polynom $1 + b_1x$ keine Nullstellen x mit $|x| \leq 1$ besitzen, man sollte also nur Parameter b_1 mit $|b_1| < 1$ zulassen.

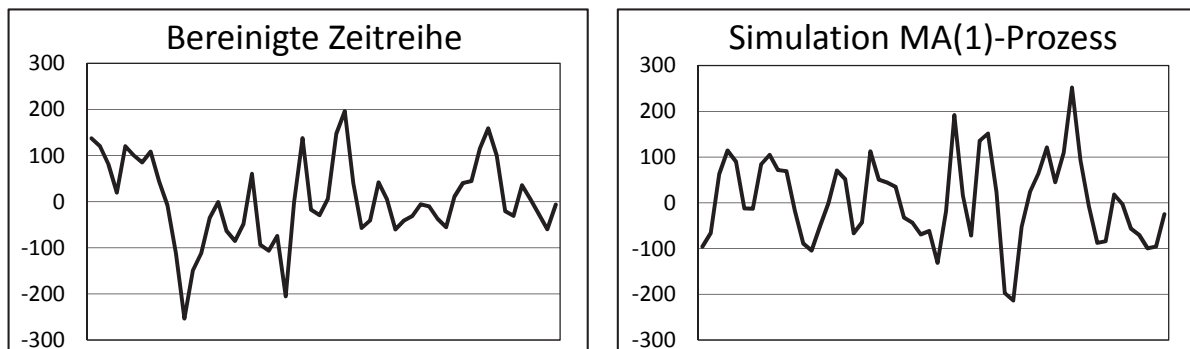
Mittels des *Maximum- Likelihood*-Verfahrens erhält man nun das Modell

$$X_t = e_t + 0,913e_{t-1} ; \sigma_e \approx 63,65 \text{ .}$$

Betrachtet man nun die Residuen, so zeigt sich, wie man an den untenstehenden Diagrammen sehen kann, dass die Hypothese es handle sich um Weißes Rauschen, nicht verworfen werden kann. Dementsprechend kann man davon ausgehen, dass das ausgearbeitete Modell gut zur Modellierung der vorliegenden Zeitreihe geeignet ist.



Auch hier kann man mit bloßem Auge feststellen, dass eine simulierte Zeitreihe gemäß des erarbeiteten Modelles ein durchaus ähnliches Verhalten wie die trend- und saisonbereinigte ursprüngliche Zeitreihe an den Tag legt.



Nun, nachdem also nichts gegen die Verwendung des Modells $X_t = e_t + 0,913e_{t-1}$ spricht, soll noch das Weiße Rauschen e auf Normalverteilung überprüft werden. Nachdem der Parameter $b_1 = 0,9103$ die Invertierbarkeit der Zeitreihe garantiert, die einzelnen Werte der Realisierung des Weißen Rauschens also berechenbar sind, müssen nun die Koeffizienten $\tilde{b}_j$ der Darstellung des Weißen Rauschens mittels $e_t = \sum_{j=0}^{\infty} \tilde{b}_j X_{t-j}$ berechnet werden.

Es gilt nach Satz 4.7, dass

$$\tilde{b}_j = -a_j - \sum_{k=1}^j \tilde{b}_{j-k} b_k = -a_j - \tilde{b}_{j-1} b_1 \quad \forall j \in \mathbb{N}_0 \text{ .}$$

Somit erhält man:

- $\tilde{b}_0 = - \underbrace{a_0}_{=-1} - \underbrace{\tilde{b}_{-1}}_{=0} b_1 = 1$
- $\tilde{b}_1 = - \underbrace{a_1}_{=0} - \tilde{b}_0 b_1 = -0,913$
- $\tilde{b}_2 = - \underbrace{a_2}_{=0} - \tilde{b}_1 b_1 = 0,913^2$
- $\vdots$

Man sieht sofort, dass sich die Koeffizienten $\tilde{b}_j$ also mittels der einfachen Formel $\tilde{b}_j = (-0,913)^j$ berechnen lassen. Obwohl man, da diese Koeffizienten auch für sehr große j nie ganz verschwinden, gerade bei der sehr begrenzten Anzahl der hier vorliegenden Daten die Werte der Realisierung des Weißen Rauschens nicht genau berechnen kann, lassen sich doch für ausreichend großes t (bspw. für $t > 10$) gute Näherungen an $\hat{e}_t$ errechnen (man beachte Satz 4.6). Untersucht man diese Näherungen wieder mittels diverser Testverfahren auf Normalität, so kann die Hypothese, es handle sich um normalverteilte Daten, mit einer Irrtumswahrscheinlichkeit von 5% nicht verworfen werden – es können also Konfidenzintervalle für zukünftige Werte der Zeitreihe erstellt werden.

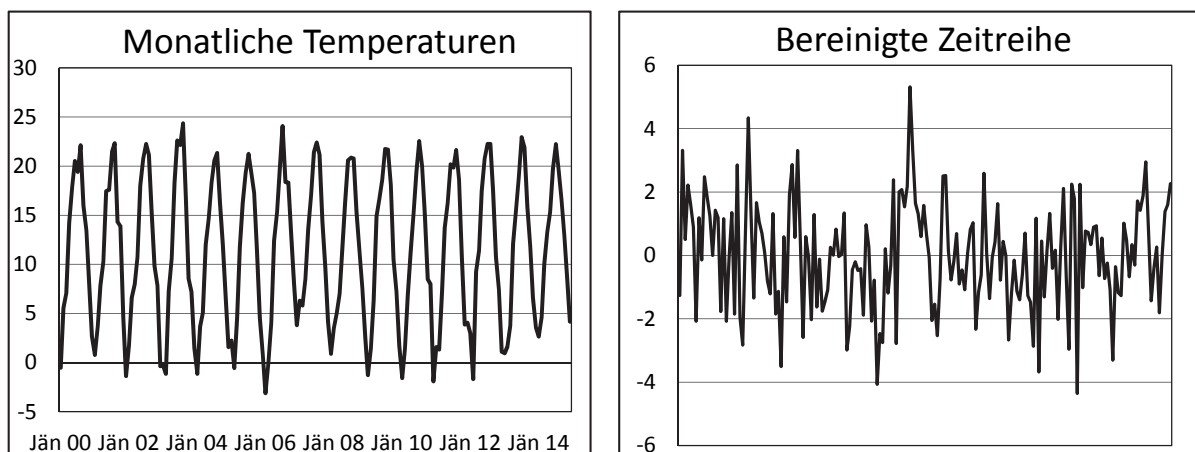
Mittels der bereits ausführlich beschriebenen Verfahren lassen sich also für die Mehrwertsteuereinnahmen der Republik Österreich folgende Aussagen treffen:

Vorhersage für:	Vorhersageintervall $\alpha = 0,05$	Tatsächlicher Wert
1. Quartal 2015	[6428,52 ; 6677,97]	6469
2. Quartal 2015	[6294,87 ; 6544,32]	6457
3. Quartal 2015	[6420,68 ; 6670,13]	6505

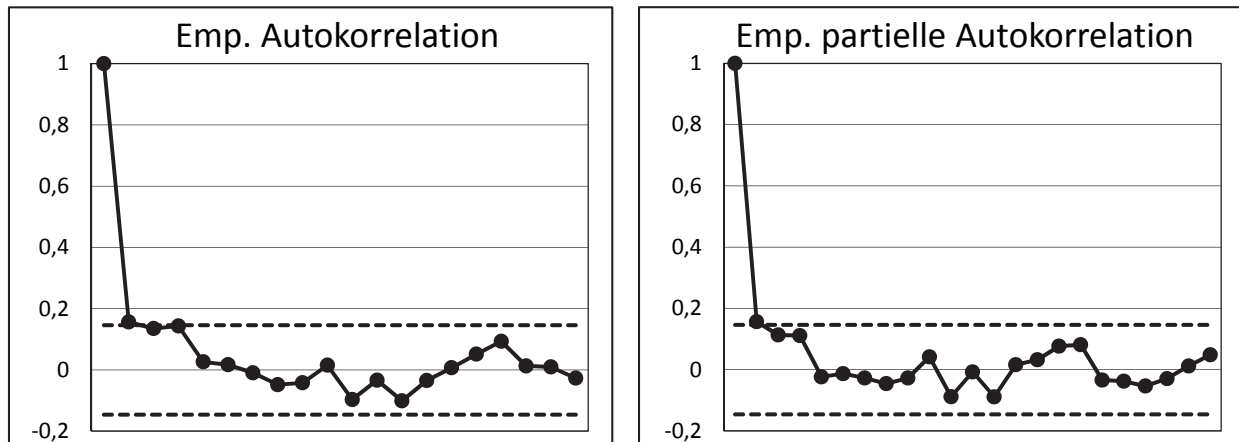
Man sieht also, dass die tatsächlich eingetroffenen Werte stets in den Grenzen der vorhergesagten Intervalle liegen – das Modell ist also tatsächlich gut geeignet um Zukunftsprognosen zu erstellen.

Beispiel 10.3

Zum Abschluss sollen hier die monatlichen Durchschnittstemperaturen in Wien analysiert werden. Zur Erstellung eines Modelles wurden Werte von Jänner 2000 bis Dezember 2014 verwendet, die Daten hierfür wurden von der Website <http://rimfrost.no> entnommen und um einen linearen Trend und die saisonalen Anteile bereinigt.



Wie gewohnt wurden anschließend die empirische Autokorrelationsfunktion und die empirische partielle Autokorrelationsfunktion betrachtet. Im Gegensatz zu den vorherigen zwei Beispielen stößt man hier aber auf ein Problem: Es ist anhand dieser Hilfsmittel nicht klar erkennbar, durch welche Art von $ARMA(p, q)$ -Modell sich die Zeitreihe optimal modellieren lässt.



Anhand der empirischen Autokorrelation lässt sich die Verwendung eines $MA(1)$ -Modelles argumentieren, wenn man das äußerst knappe Überschreiten des Konfidenzintervalls beim Lag 3 als zufälligen Ausreißer ansieht. Andererseits ist es aber natürlich auch möglich, dass eine $MA(3)$ -Zeitreihe besser geeignet ist und es sich eben nicht um solch einen Ausreißer handelt. Zusätzlich ist die empirische Autokorrelation beim Lag 2 zwar innerhalb der Intervallgrenzen, kommt diesen aber so nahe, dass man auch ein $MA(2)$ -Modell in Betracht ziehen sollte. Erschwerend kommt nun außerdem hinzu, dass auch der Wert beim Lag 1 nur sehr knapp über der gezogenen Grenze liegt – es könnte sich bei dieser Zeitreihe also durchaus auch um ein einfaches Weißes Rauschen handeln.

Anhand der empirischen partiellen Autokorrelation sieht man zusätzlich, dass sich eventuell auch ein $AR(1)$ -Modell anbieten würde. Zieht man zusätzlich auch noch die Möglichkeit in Betracht, dass ein „gemischtes“ Modell optimal sein könnte, erhält man somit folgende Möglichkeiten für passende Modelle: $MA(1)$, $MA(2)$, $MA(3)$, $AR(1)$ und $ARMA(1,1)$.

Das Vorgehen bei der Modellbildung bei $AR(1)$ und $MA(1)$ -Zeitreihen wurde in den vorigen Beispielen ausführlich beschrieben, es bleibt den Innovationsalgorithmus für $MA(2)$ -, $MA(3)$ - und $ARMA(1,1)$ -Zeitreihen exemplarisch näher zu behandeln.

Der Innovationsalgorithmus für $ARMA(1,1)$ -Zeitreihen

Gegeben sei zunächst die $ARMA(1,1)$ -Zeitreihe $-a_0 X_t = a_1 X_{t-1} + b_1 e_{t-1} + b_0 e_0$ mit $a_0 := -1$, $b_0 := 1$ und Weißem Rauschen e mit Varianz σ_e^2 .

Gemäß Satz 4.7 gilt für alle $j \geq 0$, dass $\tilde{a}_j = b_j + \sum_{k=1}^p \tilde{a}_{j-k} \cdot a_k$.

Angewendet auf diese Zeitreihe ergibt sich somit $\tilde{a}_0 = 1$ und $\tilde{a}_j = (a_1 + b_1) \cdot a_1^{j-1}$ für alle $j \geq 1$.

Nach Satz 5.1 lassen sich hieraus nun die Autokovarianzen von X bestimmen mittels

$$\gamma(h) = \sigma_e^2 \cdot \sum_{j=0}^{\infty} \tilde{a}_j \cdot a_{j+h}.$$

Die Kovarianzen der durch Transformation von X gemäß Lemma 8.2 entstehenden Zeitreihe Y lassen sich nun nach Satz 8.2 einfach berechnen, es gilt

- $Cov(Y_1, Y_1) = \frac{1}{\sigma_e^2} \cdot \gamma(0) = 1 + (a_1 + b_1)^2 \cdot \sum_{j=0}^{\infty} a_1^{2j} = 1 + \frac{(a_1 + b_1)^2}{1 - a_1^2} = \frac{1 + 2a_1b_1 + b_1^2}{1 - a_1^2}$
- $Cov(Y_1, Y_2) = \frac{1}{\sigma_e^2} \cdot (\gamma(1) - a_1 \cdot \gamma(0)) = \sum_{j=0}^{\infty} a_j a_{j+1} - a_1 \cdot \sum_{j=0}^{\infty} a_j^2 =$
 $= a_1 + b_1 + (a_1 + b_1)^2 \cdot (\sum_{j=0}^{\infty} a_1^{2j+1}) - a_1 - a_1 \cdot (a_1 + b_1)^2 \cdot \sum_{j=0}^{\infty} a_1^{2j} = b_1$
- $Cov(Y_t, Y_t) = \underbrace{b_0}_{=1} + b_1 \cdot b_1 = 1 + b_1^2 \quad \forall t \geq 2$
- $Cov(Y_t, Y_{t+1}) = b_1 + b_1 \cdot \underbrace{b_2}_{=0} = b_1 \quad \forall t \geq 2$
- $Cov(Y_t, Y_{t+h}) = b_h + b_1 \cdot b_{1+n} = 0 \quad \forall t, h \geq 2$

Man kann hier und auch im Folgenden sehr schön das Resultat von Korollar 8.1 sehen: Die Kovarianzen sind allesamt unabhängig von der Varianz des Weißen Rauschens, was für die spätere Modellbildung unumgänglich ist, da ja σ_e^2 unbekannt ist.

Wie leicht nachzurechnen ist, lassen sich nun die Ein-Schritt-Vorhersagefehler r_n von Y und die Koeffizienten $f_{i,j}$ mittels des Innovationsalgorithmus folgendermaßen rekursiv bestimmen:

- $r_0 = \frac{1 + 2a_1b_1 + b_1^2}{1 - a_1^2}$
- $f_{n,1} = \frac{b_1}{r_{n-1}}$
- $r_n = 1 + b_1^2 - b_1 \cdot f_{n,1} \quad \forall n \geq 1$

Die Ein-Schritt-Vorhersagen ergeben sich nun für alle $n \geq 0$ gemäß Satz 8.3 mittels

$$\tilde{X}_{n+1} = a_1 \cdot X_n + f_{n,1} \cdot (X_n - \tilde{X}_n)$$

Der Innovationsalgorithmus für $MA(2)$ - und $MA(3)$ -Zeitreihen

Gegeben sei die $MA(3)$ -Zeitreihe $-a_0X_t = b_1e_{t-1} + b_2e_{t-2} + b_3e_{t-3} + b_0e_0$ mit $a_0 := -1$, $b_0 := 1$ und Weißem Rauschen e mit Varianz σ_e^2 .

Mit Verwendung von Satz 4.7 erhält man:

$\tilde{a}_0$	$\tilde{a}_1$	$\tilde{a}_2$	$\tilde{a}_3$	$\tilde{a}_j \quad \forall j > 3$
1	b_1	b_2	b_3	0

Nach Satz 5.1 lassen sich die Autokovarianzen von X nun einfach berechnen:

$\gamma(0)$	$\gamma(1)$	$\gamma(2)$	$\gamma(3)$	$\gamma(h)$
$\sigma_e^2 \cdot (1 + b_1^2 + b_2^2 + b_3^2)$	$\sigma_e^2 \cdot (b_1 + b_1b_2 + b_2b_3)$	$\sigma_e^2 \cdot (b_2 + b_1b_3)$	$\sigma_e^2 \cdot b_3$	0 $\forall h > 3$

Die Kovarianzen der durch Transformation von X gemäß Lemma 8.2 entstehenden Zeitreihe Y sind nun nach Satz 8.2 folgendermaßen bestimmt:

- $Cov(Y_t, Y_t) = \frac{1}{\sigma_e^2} \cdot \gamma(0) = 1 + b_1^2 + b_2^2 + b_3^2 \quad \forall t \in [1, 3]$
- $Cov(Y_t, Y_{t+1}) = \frac{1}{\sigma_e^2} \cdot \gamma(1) = b_1 + b_1 b_2 + b_2 b_3 \quad \forall t \in [1, 2]$
- $Cov(Y_1, Y_3) = \frac{1}{\sigma_e^2} \cdot \gamma(2) = b_2 + b_1 b_3$
- $Cov(Y_t, Y_{t+3}) = \frac{1}{\sigma_e^2} \cdot \gamma(3) = b_3 \quad \forall t \in [1, 3]$
- $Cov(Y_t, Y_{t+2}) = b_2 + b_1 b_3 \quad \forall t \in [2, 3]$
- $Cov(Y_3, Y_4) = b_1 + b_1 b_2 + b_2 b_3$
- 0 sonst.

Seien nun $f_{n,j} := 0$ für alle $j \notin [1, n]$, so lassen sich die Ein-Schritt-Vorhersagefehler r_n von Y und die Koeffizienten $f_{i,j}$ mittels des Innovationsalgorithmus folgendermaßen für alle $n > 0$ rekursiv berechnen:

- $r_n = 1 + b_1^2 + b_2^2 + b_3^2 - f_{n,3}^2 \cdot r_{n-3} - f_{n,2}^2 \cdot r_{n-2} - f_{n,1}^2 \cdot r_{n-1}$
- $f_{n,3} = \frac{b_3}{r_{n-3}}$
- $f_{n,2} = \frac{1}{r_{n-2}} \cdot (b_2 + b_1 b_3 - f_{n-2,1} \cdot f_{n,3} \cdot r_{n-3})$
- $f_{n,1} = \frac{1}{r_{n-1}} \cdot (b_1 + b_1 b_2 + b_2 b_3 - f_{n-1,2} \cdot f_{n,3} \cdot r_{n-3} - f_{n-1} \cdot f_{n,2} \cdot r_{n-2})$

Die Ein-Schritt-Vorhersagen ergeben sich nun für alle $n \geq 0$ gemäß Satz 8.3 folgendermaßen:

$$\tilde{X}_{n+1} = f_{n,1} \cdot (X_n - \tilde{X}_n) + f_{n,2} \cdot (X_{n-1} - \tilde{X}_{n-1})$$

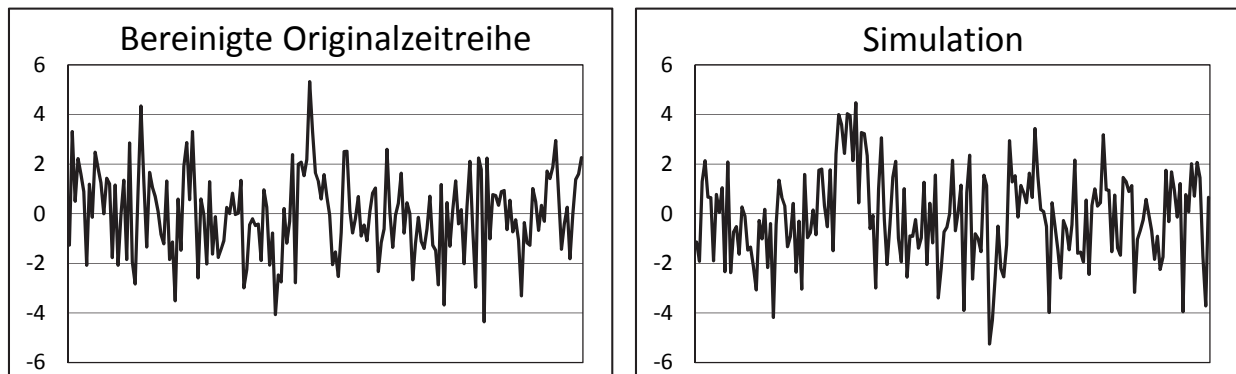
Die Berechnung der Ein-Schritt-Vorhersagen einer $MA(2)$ -Zeitreihe ergibt sich hieraus sehr einfach durch das Setzen von $b_3 := 0$ und die Berechnung des Innovationsalgorithmus für ein Weißes Rauschen darf als trivial bezeichnet werden.

Nun wurden für alle möglichen Ordnungen p und q mittels des *Maximum-Likelihood*-Verfahrens möglichst gut zur ursprünglichen (trend- und saisonbereinigten) Zeitreihe passende $ARMA(p, q)$ -Modelle erstellt, gemäß Satz 8.4 der Wert der Dichtefunktion von X_n bei der vorliegenden Realisierung $\hat{X}_n$ berechnet und anschließend die zu den jeweiligen Modellen gehörenden Werte des AIC und des $AICC$ verglichen.

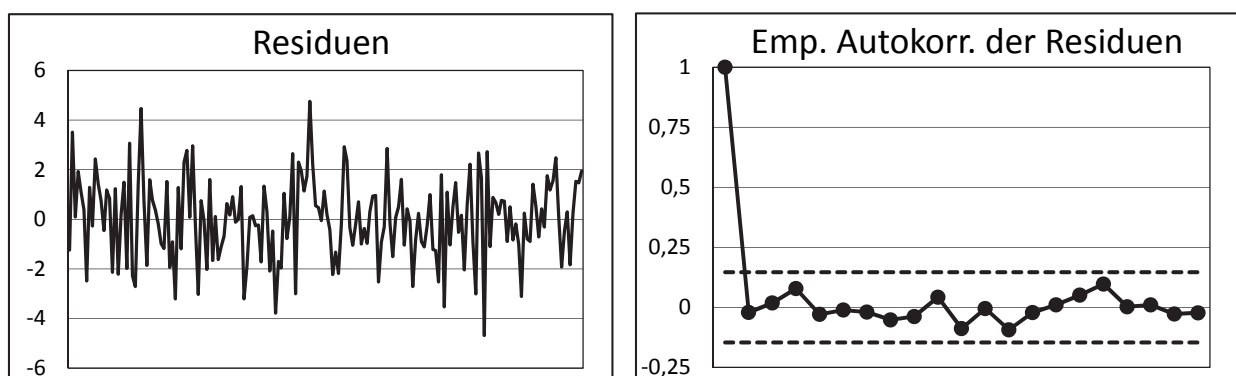
Modell	ARMA-Gleichung	σ_e
MA(1)	$X_t = 0,132e_{t-1} + e_t$	1,65778049
MA(2)	$X_t = 0,117e_{t-1} + 0,109e_{t-2} + e_t$	1,6492031
MA(3)	$X_t = 0,125e_{t-1} + 0,109e_{t-2} + 0,142e_{t-3} + e_t$	1,63285083
AR(1)	$X_t = 0,158X_{t-1} + e_t$	1,65407423
ARMA(1,1)	$X_t = 0,686X_{t-1} - 0,534e_{t-1} + e_t$	1,639592
Weißes Rauschen	$X_t = e_t$	1,67499012

Modell	Wert der Dichtefunktion bei $\hat{\mathbf{X}}_n$	AIC	$AICC$
$MA(1)$	$3,619 \cdot 10^{-151}$	696,808	696,876
$MA(2)$	$9,138 \cdot 10^{-151}$	696,956	697,092
$MA(3)$	$5,343 \cdot 10^{-150}$	695,424	695,652
$AR(1)$	$5,394 \cdot 10^{-151}$	696,010	696,078
$ARMA(1,1)$	$2,822 \cdot 10^{-150}$	694,701	694,837
Weißes Rauschen	$5,69 \cdot 10^{-152}$	698,508	698,531

Man sieht, dass sowohl das AIC als auch das $AICC$ beim $ARMA(1,1)$ -Modell minimal sind (obwohl die Realisierungswahrscheinlichkeit des $MA(3)$ -Modelles fast doppelt so hoch ist), weshalb man sich nun für die Verwendung dieses Modelles entscheidet. Zunächst muss man sich aber versichern, dass das mittels *Maximum-Likelihood*-Verfahrens ermittelte $ARMA(1,1)$ -Modell überhaupt kausal und invertierbar ist. Das AR - und das MA -Polynom besitzen jeweils nur eine einzige Nullstelle, nämlich 1,458 bzw. 1,873, weshalb diese Zeitreihe gemäß Satz 4.3 bzw. Satz 4.4 sowohl kausal als auch invertierbar ist.



Auch wenn eine mittels des ermittelten Modelles simulierte $ARMA(1,1)$ -Zeitreihe der vorliegenden Realisierung durchaus ähnlich sieht, muss nun erstmals wie gewohnt die Güte des Modelles mittels Betrachtung der Residuen verifiziert werden.

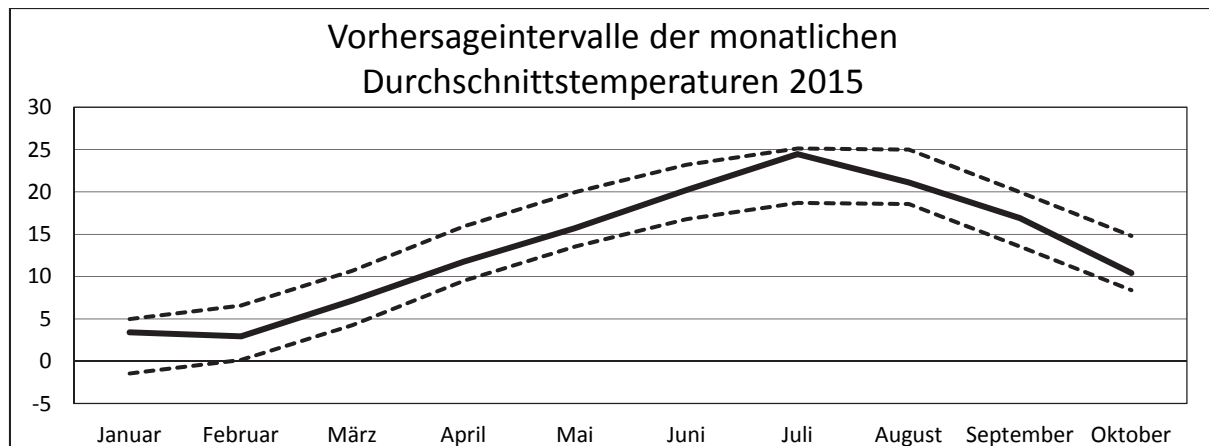


Man sieht deutlich, dass nichts gegen die Verwendung des $ARMA(1,1)$ -Modells spricht.

Nun soll wieder die vorliegende Realisierung des Weißen Rauschens $\hat{e}_t$ auf Normalverteilung überprüft werden. Hierfür müssen zunächst die Koeffizienten $\tilde{b}_j$ berechnet werden. Wie man leicht nachrechnen kann, sind diese bestimmt mittels

$$\begin{aligned} \bullet \tilde{b}_0 &= 1 & \bullet \tilde{b}_1 &= -a_1 - \tilde{b}_0 b_1 & \bullet \tilde{b}_n &= -\tilde{b}_{n-1} b_1 . \end{aligned}$$

Für ausreichend großes t lassen sich somit die Werte der Realisierung des Weißen Rauschens bestimmen und auf Normalität überprüfen. Es stellt sich heraus, dass die Hypothese, die $\hat{\epsilon}_t$ wären normalverteilt, mit einer Irrtumswahrscheinlichkeit von 5% nicht abgelehnt werden kann. Man erhält somit für die Zeit zwischen Januar und Oktober 2015 folgende Ein-Schritt-Vorhersageintervalle:



Sämtliche tatsächlich eingetroffenen Werte liegen in diesen Intervallen, was bedeutet, dass das erarbeitete Modell durchaus erfolgreich in seinen Vorhersagen ist. Allerdings zeigt sich hier auch sehr deutlich die Beschränkung der Wirksamkeit der Methodik der Zeitreihenanalyse: In vielen Fällen, wie eben beispielsweise hier, lässt sich zwar ein gutes Modell für eine Zeitreihe erstellen, es muss aber nicht immer auch der Fall sein, dass die so erhaltenen Ein-Schritt-Prognosen tatsächlich aufschlussreich für eine praktische Anwendung sind.

Mittels des soeben ermittelten Modells könnten nun Zukunftsprognosen wie folgende getroffen werden: „Die Durchschnittstemperatur im Juni 2015 wird mit 95%-iger Wahrscheinlichkeit zwischen 16,77°C und 23,20°C liegen.“

Diese Vorhersage stellte sich zwar im Nachhinein als völlig zutreffend heraus, ist aber viel zu unpräzise um tatsächlich anwendbare Erkenntnisse daraus ziehen zu können – ein Problem, das in der Zeitreihenanalyse immer wieder auftritt. Existieren zusätzlich zu den Daten einer Zeitreihe zum nächsten Wert beitragende bekannte Faktoren, wie hier zum Beispiel das Wissen um Hochdruck- und Tiefdruckgebiete, so ist es meist vernünftiger anhand dieser Modelle zu erstellen – oder aber auf multivariate Modelle der Zeitreihenanalyse zurückzugreifen.

Diese Beschränkung der Wirksamkeit von Prognoseverfahren erkannte bereits der große Physiker Niels Bohr, dessen treffendes Zitat diese Arbeit abschließen soll:

„Prognosen sind immer schwierig, besonders, wenn sie die Zukunft betreffen.“¹

¹ Jürgen Vogel, Prognose von Zeitreihen : Eine Einführung für Wirtschaftswissenschaftler (Springer, 2015) S.9

Literaturverzeichnis

Karl *Bosch*, Elementare Einführung in die Wahrscheinlichkeitsrechnung (Vieweg + Teubner, 2011)

Karl *Bosch*, Elementare Einführung in die angewandte Statistik (Vieweg + Teubner, 2010)

Günter *Bamberg*, Franz *Baur*, Michael *Krapp*, Statistik (Oldenbourg, 2012)

Peter J. *Brockwell*, Richard A. *Davis*, Introduction to Time Series and Forecasting (Springer, 2002)

Peter J. *Brockwell*, Richard A. *Davis*, Time Series: Theory and Methods (Springer, 1991)

Dmitry *Efrosinin*, Spezialvorlesung Zeitreihenanalyse – Mit Beispielen in Mathematica (Vorlesungsskriptum, 2006)

Norbert *Henze*, Stochastik für Einsteiger, Eine Einführung in die faszinierende Welt des Zufalls (Springer, 2013)

Harro *Heuser*, Lehrbuch der Analysis Teil 1 (Vieweg+Teubner, 2009)

Hajo *Holzmann*, Zeitreihenanalyse (Vorlesungsskriptum, 2009)

Jens-Peter *Kreiß*, Georg *Neuhaus*, Einführung in die Zeitreihenanalyse (Springer, 2006)

Sadanori *Konishi*, Genshiro *Kitagawa*, Information Criteria and Statistical Modeling (Springer, 2008)

David *Meintrup*, Stefan *Schäffler*, Stochastik – Theorie und Anwendungen (Springer, 2005)

Klaus *Neusser*, Zeitreihenanalyse in den Wirtschaftswissenschaften (Vieweg + Teubner, 2011)

Matti *Schneider*, Sebastian *Mentemeier*, Zeitreihenanalyse mit R (Vorlesungsskriptum, 2010)

Jürgen *Vogel*, Prognose von Zeitreihen : Eine Einführung für Wirtschaftswissenschaftler (Springer, 2015)

Anhang

Zusammenfassung

Diese Diplomarbeit befasst sich mit der Analyse von Zeitreihen, also mit dem Suchen von Modellen zur Beschreibung von in festen zeitlichen Abständen aufgezeichneten Daten.

Dazu werden zunächst allgemein der Begriff Zeitreihe und damit eng verbundene Konzepte wie Stationarität und Autokorrelation definiert und näher betrachtet um anschließend die elementaren Zeitreihenmodelle des Weißen Rauschens, der Autoregressiven- wie auch der Moving-Average-Prozesse einzuführen. Diese werden schließlich zu der dieser Arbeit titelgebenden Klasse der Autoregressiven-Moving-Average-Zeitreihen oder kurz der *ARMA*-Zeitreihen verallgemeinert, wobei besonderes Augenmerk auf die praktisch relevanten Eigenschaften dieser Modellklasse, wie beispielweise Kausalität, gelegt wird.

Anschließend wird das Verfahren der Maximum-Likelihood-Schätzung für die Parameter einer *ARMA*-Zeitreihe, die zu einer zu untersuchenden Zeitreihe möglichst gut beschreiben soll, näher behandelt und eine Methode zur Prognose zukünftiger Werte beschrieben.

Den Abschluss dieser Diplomarbeit bilden drei Anwendungsbeispiele.

Abstract

The subject of this diploma thesis is the analysis of Time Series, that means the search for suitable models to describe data, which are observed at regular intervals.

At first the term Time Series and associated concepts like Stationarity or Autocorrelation are defined and examined. Subsequently the most basic classes of Time Series, namely White Noise, Autoregressive- and Moving-Average-Processes, are introduced to the reader. These are then generalized to the so-called Autoregressive-Moving-Average-Processes, or *ARMA*-Processes, which are eponymous to this thesis. Following this, practically relevant properties of such processes like Causality are closer examined.

Afterwards the method of Maximum-Likelihood-Estimation for the unknown parameters of an *ARMA*-Process, which should describe a given Time Series as good as possible, are discussed as well as a way to forecast its future values. Finally three Time Series are exemplarily modelled as *ARMA*-Processes.