

4. Zeitreihenanalyse, ARCH & GARCH

Nach einführenden Bemerkungen wenden wir uns der Beschreibung von Zeitreihen zu. Die gängigen Modelle werden präsentiert. Dann werden wir uns mit der statistischen Analyse im Zeitbereich beschäftigen. Mit Prognosemethoden sowie ARCH- und GARCH-Modellen wenden wir uns dann den direkten Anwendungen in der Finanzmathematik zu.

4.1 Warum Zeitreihenanalyse in einer Vorlesung über Finanzmathematik

Wir haben in Kapiteln 1 bis 3 mehrere Modelle kennengelernt, welche man auf Probleme in der Praxis anwendet. Wir haben bereits bei der Diskussion der Formel von Black-Scholes darauf hingewiesen, dass beispielsweise die Volatilität σ in der Realität nicht bekannt ist. Sie muss demzufolge geschätzt werden. Ganz allgemein gilt:

Die Modelle werden einmal berechnet und "stehen dann". In diesen Modellen kommen aber Parameter vor, welche häufig nicht a priori bekannt sind. Sie müssen demzufolge aus historischen Daten geschätzt werden. Was nun aber die Wirtschaftswissenschaften und insbesondere die Finanzwelt - im Gegensatz zu Chemie, Biologie und Medizin - auszeichnet, ist, dass diese Daten meist in Form von Zeitreihen vorkommen.

Man will also die Modelle mit Hilfe von historischen Daten kalibrieren. Der Einsatz der Zeitreihenanalyse in den Wirtschaftswissenschaften ist vielfältig:

1. **Vorhersage von Preisen:** Mit vielfältigsten Methoden (auch aus der Zeitreihenanalyse) versucht man, Aktienkurse selber vorherzusagen. Neben naiven Versuchen und Problemstellungen gibt es aber auch sinnvollere Einsatzgebiete: wenn man verderbliche, nicht lagerbare Güter hat, welche nur zu einer gewissen Jahreszeit geerntet werden, so kann man auf Grund von früheren Daten unter Berücksichtigung von Witterung und saisonalen Einflüssen etc. versuchen, die künftigen Preise vorherzusagen. Damit ist nicht gesagt, dass man einen Informationsvorsprung gegenüber anderen Marktteilnehmern hat. Aber die heutigen Preise für Güter, welche erst in der Zukunft explizit ausgetauscht werden, können so eher begründet werden.
2. **Volkswirtschaftliche Untersuchungen:** Man beobachtet (u.a. wegen der Bauwirtschaft) im Winter in den entwickelten Ländern der nördlichen Hemisphäre eine höhere Arbeitslosigkeit als im Sommer. Wenn nun gegen Herbst hin in der Wirtschaft ein Aufschwung einsetzt, so kann die Arbeitslosigkeit zwar immer noch steigen, weil die Wintermonate kommen. Hingegen kann man mittels Berücksichtigung von saisonalen Einflüssen eventuell doch bereits schliessen, dass die *saisonbereinigte* Arbeitslosigkeit zurückgeht.
3. **Schätzung von Parametern für Modelle der Finanzmathematik:** Wir haben bereits das σ im Modell von Samuelson kennengelernt. Im Portfoliomanagement (siehe Kapitel 1) muss man die erwarteten Renditen μ und die Korrelationen Σ berechnen.
4. **Einsatz im Riskmanagement:** Bei der Besprechung der Formel von Black-Scholes haben wir bereits festgehalten, dass die Logreturns nicht i.i.d. und normalverteilt sind. Man kann realistischere Modelle für die Logreturns aufstellen (ARCH und GARCH), mit denen man die Wahrscheinlichkeiten für grosse Verluste (im Modell!) berechnen kann.
5. **Schätzung von Zinsstrukturen:** Wenn keine monetäre Extremsituation vorhanden ist, so sind die Zinsen von langfristig angelegtem Geld höher als von kurzfristig angelegtem Geld. Der Grund liegt in der längeren Ungewissheit. Der Zins als Funktion der Laufzeit wird als Zinsstruktur bezeichnet. Diesen Verlauf versucht man auch mit Zeitreihenanalyse anhand von Faktoren zu erklären.
6. **Länder- und Brancheneffekte:** Gerade mit der Einführung des Euro's hat man viel über Länder- und Brancheneffekte gelesen. Gab es in der EU früher noch eindeutige Ländereffekte (die Aktienkurse von Firmen im gleichen Land (mit gleicher Währung) reagierten bei gewissen relevanten Änderungen des wirtschaftlichen Umfeldes (Zinsentscheide der Notenbank) gleich), so fallen diese zumindest bei Währungsfragen jetzt weg. In den Vordergrund treten jetzt Brancheneffekte (die Aktienkurse von Firmen der gleichen Branche reagieren bei gewissen relevanten Änderungen des wirtschaftlichen Umfeldes (höhere Erdölpreise zum Beispiel) gleich). Früher waren diese beiden Effekte vermischt. Dies hat Folgen

für das Portfoliomanagement, wo man die Aktienanlagen gezielt diversifizieren möchte. Dazu muss man Zeitreihenanalysen machen.

7. **Lieferung von statistischen Grundlagen für ökonomische Theorien:** Ökonomische Theorien und Modelle werden nicht einfach im Elfenbeinturm geboren. Mit exploratorischer Statistik werden ökonomische Zeitreihen analysiert. Theorien über Wechselkursrelationen werden nach mehrdimensionaler Zeitreihenanalyse aufgestellt. Auch das Modell von Samuelson wird wohl nach Zeitreihenanalyse von realen Aktienkursen geboren worden sein.

In dieser Vorlesung werden aus Zeitgründen u.a. nicht behandelt: Spektralanalyse, mehrdimensionale Zeitreihenanalyse.

4.2 Beschreibung von Zeitreihen

Definition 4.1 [Zeitreihe (ZR) (time series)] Eine Zeitreihe ist eine (zeitlich) geordnete Folge $(x_t)_{t \in T}$ von Daten (Beobachtungen). Zentral ist, dass die Reihenfolge der Messungen relevant ist und sukzessive Folgenglieder im Allgemeinen **nicht** als unabhängig betrachtet werden dürfen.

Nochmals: Die Folgenglieder der Messungen sind im Allgemeinen nicht unabhängig voneinander. Tägliche Aktienkurse sollten nicht als unabhängige Realisationen von positiven Zufallsgrößen modelliert werden - eher die Zuwächse davon! Wir werden uns auf Zeitreihen beschränken, bei denen die Zeit diskret ist (nicht aber die Werte, welche die Beobachtungen annehmen). T ist meist die Menge der natürlichen Zahlen oder die Menge der ganzen Zahlen. Damit ist nicht gesagt, dass die Werte in der Realität nur zu diskreten Zeitpunkten anfallen und diese Zeitpunkte zudem äquidistant sind.

Wenn man eine Realisation ω_1 einer Zeitreihe (z.B. simulierter Aktienkurs) betrachtet, so ist es wohl als MathematikerIn unvermeidlich, dass man Muster zu erkennen versucht und Hypothesen aufstellt. Darunter werden sich viele Hypothesen befinden, welche man bei einer nochmaligen Realisation ω_2 sofort fallenlassen wird. In der Realität hat man eigentlich nur *eine* Zeitreihe, eine Realisation. Wenn man diese betrachtet (als Plot), so wird man wohl auch viele Hypothesen aufstellen, welche "unsinnig" sind. Es gibt aber keine nochmalige Realisation, kein ω_2 (oder gar noch mehr Realisationen), mit deren Hilfe erste Versuche Versuche bleiben. Man könnte deshalb folgern, dass das Betrachten eines Plots (Diagramms), etwas gefährliches ist, weil es zu Fehlschlüssen verleiten könnte. Diese Gefahr besteht! Hingegen ist es trotzdem klar, dass der erste Schritt der Datenanalyse darin besteht, dass man einen Plot macht und diesen auf den Betrachter "einwirken lässt". Wir wollen jetzt einige Plots betrachten:

4.2.1 Beispiele von Zeitreihen

Es gibt klassische Standardbeispiele, welche praktisch in allen Büchern über Zeitreihenanalyse vorkommen.

4.2.1.1 Ökonomische Zeitreihen

Wir haben bereits über Aktienkurse gesprochen, welche an nachfolgenden Tagen angegeben werden können. Weitere Beispiele sind Exportzahlen der Volkswirtschaft in auf sich folgenden Monaten, durchschnittliche Einkommen der Haushalte, Gewinne von Firmen und weitere.

In Figur 4.1 ist die (klassische) Serie des Indexes der Preise von Weizen angegeben. Der Index umfasst jeweils den durchschnittlichen Preis aus etwa 50 Handelsorten diverser Länder. Dieser Preisindex (von 1500 bis 1869) wird vor allem von Wirtschaftshistorikern intensiv studiert.

4.2.1.2 Physikalische Zeitreihen

Viele Beispiele von Zeitreihen stammen aus der physikalischen Welt; aus der Meteorologie, marinen Studien und der Geophysik. Beispiele sind Regenfall auf nachfolgenden Tagen oder Monaten, Temperaturen auf nachfolgenden Stunden, Tagen, Monaten, Jahren. In Figur 4.2 sind die monatlichen mittleren Temperaturen

in Recife (Brasilien) abgebildet. Anhand dieses Beispiels wird ein methodisches Problem deutlich: es gibt Zeitreihen, welche **stetig aufgezeichnet** werden (Messungen der Temperatur durch eine Nadel auf einem Endlospapier, welches unter der Nadel durchgezogen wird). Solche Zeitreihen muss man **diskretisieren**, damit sie mit den gängigen Methoden analysiert werden können.

4.2.1.3 Zeitreihen aus dem Marketing

Im Marketing hat man Verkaufszahlen aus nachfolgenden Tagen, Wochen, Monaten und Jahren. Beispielsweise will man rechtzeitig feststellen können, ob ein Rückgang der Verkaufszahlen aus saisonalen Gründen erwartet werden kann, oder ob man dabei ist, Marktanteile zu verlieren. Andererseits will man auch Verkaufsprognosen haben, um die Produktion und die Lagerhaltung planen zu können. Daneben will man vielleicht auch die Verkaufszahlen und Werbeausgaben gegenüberstellen. Ein Beispiel solcher Verkaufszahlen finden wir in Figur 4.3 (Verkauf eines technischen Gerätes durch eine Firma).

4.2.1.4 Demographische Zeitreihen

Demographische Zeitreihen ermöglichen Planungen im Bereich von Schulen, Verkehr und Spitälern. Aber auch AHV-Prognosen sind von demographischen Zeitreihen abhängig. Die Neu-Rentner von 2065 werden gerade geboren!

4.2.1.5 Zeitreihen aus Produktionsprozessen

Die Zeitreihenanalyse wird auch in der Qualitätskontrolle in Produktionsprozessen eingesetzt. Beispielsweise darf eine Variable in einem Produktionsprozess (Konzentration einer chemischen Verbindung in der Pharmaindustrie) nicht zu stark um einen Zielwert variieren. Dies kann man im Zeitablauf aufzeichnen. Wenn der Wert der Variablen zu stark vom Zielwert abweicht, muss eingeschritten werden. Ein Beispiel davon ist in Figur 4.4 aufgezeichnet.

4.2.1.6 Binäre Prozesse

In der Nachrichtenübermittlung hat man ganz spezielle Zeitreihen, bei denen die interessante Grösse nur 2 Werte annehmen kann; diese werden dann meist mit 0 (aus) und 1 (ein) bezeichnet. Eine Realisation eines solchen sogenannten *binären Prozesses* ist in Figur 4.5 abgebildet.

4.2.1.7 Punktprozesse

In Figur 4.6 sieht man eine Zeitreihe, bei der nur der Zeitpunkt selber die relevante Information ist. Beispielsweise kann man sich fragen, ob Bootsbrände am Zürichseeufer, welche von einem Brandstifter ausgelöst werden, mit gewisser Regelmässigkeit stattfinden oder nicht.

4.2.2 Worum geht es in der Zeitreihenanalyse

In der Zeitreihenanalyse hat man allgemein vier Ziele vor Augen: Beschreibung, Erklärung, Vorhersage und Kontrolle.

4.2.2.1 Beschreibung

Wie bereits erwähnt, ist der erste Schritt ein Plot der Zeitreihe. Dann wird man diverse einfache, deskriptive Methoden anwenden (mehr im verbleibenden Teil von 4.2): In Figur 4.3 sahen wir einen Saisoneffekt; zudem gab es auch einen leichten Aufwärtstrend (verbunden mit immer grösseren Ausschlägen). Die Variation (darunter verstehen wir die umgangssprachliche Variation, Varianz), welche man in gewissen Zeitreihen beobachtet, kann eventuell beinahe vollständig auf derart offensichtliche Ursachen zurückgeführt werden. Meist ist jedoch eine kompliziertere Zeitreihenanalyse von Nöten; zur Modellierung wird man dann Modelle

einsetzen, welche in Teil 4.3 vorgestellt werden. Im Teil "Beschreibung" fließt auch "gesunder Menschenverstand" ein. Mathematische Pharisäer dürften damit Schwierigkeiten haben. Die Abbildung {Serie von Daten} \rightarrow {Hypothesen, Vermutungen} "ist nicht aus C^∞ "; diese "Abbildung" ist mathematisch nicht formalisiert. Das ist eine Stärke (grosse Flexibilität, Innovation), birgt aber auch Gefahren (Fehlschlüsse, Problem der Trennung "exploratorische" und "konfirmatorische" Statistik).

4.2.2.2 Erklärung

Wenn man zwei Zeitreihen beobachtet, so kann man einen Teil der Variation in der einen Zeitreihe eventuell mit der anderen Zeitreihe erklären. Beispielsweise kann man Verkaufszahlen bei Erdöl mit den Preisen von Erdöl zu erklären versuchen (oder Exporte/Importe mit Wechselkursen). Wir werden in diesem Skript (aus Zeitgründen) dieses Gebiet nicht behandeln.

4.2.2.3 Vorhersage

Wenn man Werte einer Zeitreihe bis Heute hat, so will man vielleicht die zukünftigen Werte vorhersagen. Wie weiter oben bereits angeführt, ist dies auch bei Preisen *nicht* im naiven Sinne zu verstehen, dass man dann quasi eine Geldmaschine besitzt. In der Finanzwelt hochaktuell sind gute Schätzungen der Volatilität σ von Aktien. Wir haben gesehen, dass diese einen sehr direkten Einfluss auf die Preise eines Calls oder Puts hat (steigende Volatilität, steigende Preise für Calls und Puts).

4.2.2.4 Kontrolle

Wenn man in einem industriellen Prozess eine Variable (Konzentration) möglichst um einen Zielwert konzentrieren will, so wird man mit der Zeitreihenanalyse auch Eingreifkriterien finden wollen.

4.2.3 Überblick über die Ursachen von Variation in Zeitreihen

Häufig werden Zeitreihen analysiert, indem man die Variation in verschiedene Teile zerlegt: Trend, saisonale Einflüsse, andere zyklische Einflüsse und "irreguläre" Schwankungen. Dieser Ansatz muss nicht immer der Weisheit letzter Schluss sein; er ist angebracht, wenn Trend oder/und saisonale Schwankungen die Variation in der Zeitreihe klar dominieren. Ohne Annahmen ist diese Zerlegung sowieso nicht immer eindeutig. Damit ist man auch nicht mehr nur beim *Beschreiben* einer Zeitreihe, sondern kommt langsam in den Bereich der *Modellierung*. Trotzdem: Bevor man komplizierteste Methoden der Zeitreihenanalyse anwendet, sollte man die Zeitreihe auch anhand eines Plots von Auge (mit gesundem Menschenverstand) auf folgende Ursachen von Variation untersuchen:

4.2.3.1 Saisonale Einflüsse

Wir haben bereits bei Temperaturmessungen oder Arbeitslosenzahlen über saisonale Effekte gesprochen. Die Periode ist dort 1 Jahr. Diese Art Variation wird erstens schnell erkannt (von Auge und mit mathematischen Masszahlen), es gibt zweitens eine plausible Erklärung und drittens kann man sie auch leicht entfernen (wir wollen sogenannte *stationäre* Zeitreihen am Schluss).

4.2.3.2 Andere zyklische Schwankungen

Neben jährlichen Schwankungen können die Schwankungen auch täglich sein: Temperaturen zum Beispiel. Ob es auch "zyklische" Schwankungen gibt, welche "keine feste Periode" haben, ist strittig. Ein umstrittenen Beispiel ist der Konjunkturzyklus von etwa 5 bis 7 Jahren. Vor allem die Konjunktur seit den frühen 80er Jahren hat keinerlei derartige Muster gezeigt. Vielleicht gab es diese Zyklen auch deshalb, weil genug Leute daran geglaubt haben.

4.2.3.3 Trends

Umgangssprachlich formuliert wollen wir unter einem Trend eine "langfristige Änderung des Mittelwerts einer Zeitreihe" verstehen. Was ist unter *langfristig* zu verstehen? Diese Frage kann in der Zeitreihenanalyse oft nicht abschliessend beantwortet werden. Wir wissen aus der aktuellen Diskussion über den Treibhauseffekt, dass man wohl erst dann unumstösslich eine Trendwende der "durchschnittlichen" Temperaturen feststellen wird, wenn wir zu viel Schaden angerichtet haben. Dass es Effekte gibt, welche die Temperaturen auf der Erde erhöhen, ist wohl mittlerweile unbestritten. Nur ist fraglich, ob diese auch tatsächlich unter dem Strich zu einer dauerhaften Erwärmung führen. Weder Theorie noch empirische Daten der Temperaturen geben hier Klarheit. Hingegen wird man bei Aktienkursen einen allgemeinen, (exponentiell) ansteigenden Trend kaum in Frage stellen. Wenn man zyklische Schwankungen mit Perioden hat, welche länger sind als das Zeitintervall, welches die Zeitreihe abdeckt, kann es eventuell doch sinnvoll sein, solche ansteigenden Phasen als Trend zu betrachten.

4.2.3.4 Irreguläre Schwankungen

Nachdem man "offensichtliche" Trends oder zyklische Effekte aus den Zeitreihen entfernt hat, wird man immer noch Residuen haben, welche als zufällig betrachtet werden können oder auch nicht. Man wird dann noch versuchen, spezielle Modelle aus 4.3 anzupassen: Moving Averages (MA) oder Autorregressive (AR) Prozesse.

4.2.4 Stationäre Zeitreihen

Wir werden in 4.3 eine strenge mathematische Definition des Begriffes "Stationarität" geben. Intuitiv nennen wir eine Zeitreihe stationär, wenn

1. sich der Mittelwert nicht mehr ändert (kein Trend),
2. sich die Varianz nicht mehr ändert,
3. periodische Variationen nicht mehr vorkommen.

Die Zeitreihenanalyse im engeren Sinne befasst sich mit solchen stationären Zeitreihen. Damit muss eine nicht-stationäre Zeitreihe also zuerst stationär gemacht werden. Dann kann man die Modelle aus 4.3 auf die Residuen anzuwenden versuchen. Es ist aber auch gut denkbar, dass man sich in einem praktischen Problem gar nicht für diese Modellierung der Residuen interessiert, sondern lediglich den Trend oder die zyklische Komponente kennen möchte.

4.2.5 Der Plot der Reihe gegen die Zeit

Es ist bereits erwähnt worden, dass die Betrachtung des Plots am Anfang jeder Zeitreihenanalyse stehen sollte. Vorher wird man aber sicher Hintergrundwissen aus der Substanzwissenschaft einholen und die Fragestellungen der interessierten Kreise zur Kenntnis nehmen. Der Plot sollte Trend, saisonale Muster, Ausreisser und Diskontinuitäten zum Vorschein bringen. Beim plotten müssen folgende Punkte beachtet werden:

1. Wie wählt man die Skalen?
2. Wie wählt man den Intercept ("y-Achsenabschnitt")?
3. Wie werden die Punkte dargestellt? Verbunden als Linie - Einzel?

Diese Punkte können die Hypothesenbildung beeinflussen; es ist also Vorsicht geboten. Hier sieht man wieder, wie philosophisch unsicher die Methoden der Zeitreihenanalyse sind. Dass man mit dieser Disziplin trotzdem weitergefahren ist, liegt an den Erfolgen, welche man damit erzielt hat. Weitere selbstverständliche Anforderungen an einen Plot sind: Titel, Anschrift der Achsen inklusive Einheiten.

4.2.6 Transformationen

Wenn man die Daten geplottet hat, wird man eventuell sofort sehen, dass eine Transformation angebracht ist. Dies kann eine Logarithmierung der Daten sein oder man zieht von allen Daten die Quadratwurzel. Es gibt drei zentrale Gründe, warum man die Daten transformieren möchte:

4.2.6.1 Man möchte die Varianz stabilisieren

Wenn wir in der Zeitreihe einen Trend haben und die Varianz mit dem Mittelwert der Daten zunimmt, so ist eine Transformation eventuell angebracht. Wenn die Standardabweichung proportional zum Mittelwert ist, so ist eine logarithmische Transformation angesagt: $x_i = m_i \epsilon_i$ wird dann zu $\log(x_i) = \log(m_i) + \log(\epsilon_i)$. Man kommt so von einem *multiplikativen* Modell zu einem *additiven* Modell. Die Fehler ϵ_i werden nicht mehr mit den momentanen Mittelwerten multipliziert.

4.2.6.2 Man möchte den saisonalen Effekt additiv machen

Wenn wir wieder einen Trend in der Zeitreihe haben und der Saisoneffekt eine mit dem Mittelwert der Zeitreihe zunehmende Amplitude hat, so ist wieder eine Transformation angebracht, damit der Saisoneffekt eine etwa konstante Amplitude aufweist. Den Saisoneffekt nennt man dann additiv. Falls der Saisoneffekt direkt proportional zum Mittelwert der Zeitreihe ist, dann nennt man den Saisoneffekt multiplikativ - man wird wieder eine logarithmische Transformation machen. Die Logarithmierung ist nur dann sinnvoll, wenn auch der Fehlereffekt multiplikativ eingeht: aus $x_i = m_i s_i \epsilon_i$ wird dann $\log(x_i) = \log(m_i) + \log(s_i) + \log(\epsilon_i)$.

4.2.6.3 Man möchte die Daten normalverteilt machen

Modellbildung und Vorhersage bauen häufig auf der Annahme auf, dass die Daten normalverteilt sind. Der zentrale Grenzwertsatz mag von Fall zu Fall solche Annahmen als gerechtfertigt erscheinen lassen. In der Tat sieht man aber häufig, dass die Daten deutlich *nicht* normalverteilt sind. Wir haben auch bei den Logreturns festgehalten, dass wir Schiefe und Langschwänzigkeit in den Daten haben. Obschon diverse Ansätze (Transformationen) existieren, um die Daten etwa normalverteilt zu machen, ist dies ein schwieriges Unterfangen. Eventuell muss man eine andere Fehlerverteilung wählen.

Mehr zu Transformationen bei Zeitreihen findet man in Schlittgen/Streitberg (v.a. Box-Cox-Transformation). Es gibt Untersuchungen, bei denen man festgestellt hat, dass Vorhersagen mit transformierten Daten kaum zu besseren Resultaten führten als ohne Transformationen. Es gibt auch Probleme derart, dass eine Transformation eventuell den Saisoneffekt additiv macht aber die Varianz nicht stabilisiert; generell: alle Anforderungen können selten genügend berücksichtigt werden. Komplizierteste Modelle für transformierte Daten sind sowieso fragwürdig. Die Daten müssen meist wieder rücktransformiert werden; die rücktransformierten *Schlussfolgerungen* sind eventuell von fragwürdigem Wert. Auch sollte man für transformierte Daten eine praktische Interpretation haben. In dieser Vorlesung werden Probleme behandelt, bei denen die Transformationen "kanonisch" sich beinahe aufdrängen: wir sind nicht unbedingt an den aktuellen Aktienkurse interessiert, sondern an der Rendite von einer Periode zur nächsten. Wir haben gesehen, dass die Logreturns sehr nahe bei Renditen sind *und* im Modell von Samuelson eine Normalverteilung besitzen (siehe 3.12.3). Damit ist klar und weitgehend unbestritten, dass die Logreturns von Aktienkursen wichtige Studienobjekte sein müssen (aber noch lange nicht die einzigen!).

4.2.7 Wie behandelt man Zeitreihen mit Trends

Die einfachste Form eines Trends ist sicher die Familie "linearer Trend + Fehler". Dabei geht man davon aus, dass der Wert x_t zur Zeit t mit Konstanten α, β als Realisation einer Zufallsgrösse von der Form

$$X_t = \alpha + \beta t + \epsilon_t \tag{4.1}$$

dargestellt werden kann. Dabei fordert man sinnvollerweise $E[\epsilon_t] = 0$. Dann sind die Mittelwerte gegeben durch die Gerade $m_t = (\alpha + \beta t)$, dies ist der Trendterm. Der Trend in (4.1) ist hier deterministisch. Auch sind hier α und β konstant im Zeitablauf. Man kann sie auch zeitabhängig modellieren. Der Trend kann auch

nichtlinear sein: quadratisch oder exponentiell; wir haben beim Modell von Samuelson einen exponentiellen Trend beobachtet. Wenn man eine Zeitreihe mit Trend analysieren möchte, muss man sich zuerst entscheiden, ob man a) in erster Linie den Trend messen möchte oder/und b) den Trend entfernen will, um weitergehende Analysen durchführen zu können. Wenn die Daten neben dem Trend auch saisonale Einflüsse aufweisen, so kann man sukzessive Jahresdurchschnitte berechnen, um eine Idee vom Trend zu erhalten. Manchmal kann man mit derart simplen Methoden Trends in Zeitreihen entfernen. Nicht selten muss man aber eine der nachfolgenden Methoden anwenden:

4.2.7.1 Anpassung von Kurven

Eine traditionelle Methode, um bei Zeitreihen ohne Saisonkomponenten den Trend zu entfernen, ist die Anpassung relativ einfacher Funktionen an die Daten. Solche einfache Funktionen sind: polynomiale Kurven (linear, quadratisch, etc.), Gompertz- oder logistische Kurven. Die Gompertz-Kurven sind Kurven der Art

$$\log(x_t) = a + br^t,$$

wobei a, b, r Parameter sind mit $0 < r < 1$; die logistischen Kurven sind Kurven der Art

$$x_t = a/(1 + be^{-ct}).$$

Die logistische Kurve hat eine S-Form und sowohl die logistische Kurve wie auch die Gompertz-Kurven streben gegen einen asymptotischen Wert, wenn $t \rightarrow \infty$. Diese Funktionen kann man beispielsweise mit der Methode der kleinsten Quadrate anpassen. Die Berechnungen dazu sind eventuell sehr umfangreich und werden mit Rechnern vorgenommen. Wenn man dann eine Trendfunktion angepasst hat, so kann man diese von den ursprünglichen Daten subtrahieren und erhält damit die Residualzeitreihe, welche man mit speziellen Methoden weiter analysieren möchte.

4.2.7.2 Filter

Eine weitere Möglichkeit, um Trends in Zeitreihen beizukommen, sind lineare Filter. Sie wandeln eine Zeitreihe $\{x_t\}$ in eine neue Zeitreihe $\{y_t\}$ um. Diese Umwandlung geschieht nach folgendem Gesetz:

$$y_t = \sum_{r=-q}^{+s} a_r x_{t+r}.$$

Dabei nennt man $\{a_r\}$ die Gewichte. Wenn man lokale grössere Ausschläge glätten will und den Mittelwert lokal schätzen möchte, so bieten sich als Filter die sogenannten Moving Averages MA (Gleitende Durchschnitte) an. Man wählt die a_r 's derart, dass $\sum a_r = 1$. Oft wählt man auch $q = s$ und $a_{-j} = a_j$ (Symmetrie). Wenn zudem die Gewichte gleich gewählt werden ($a_r = 1/(2q + 1)$ für $r = -q, \dots, q$), so spricht man vom *einfachen gleitenden Durchschnitt*:

$$y_t = \frac{1}{2q + 1} \sum_{r=-q}^q x_{t+r}.$$

Der einfache gleitende Durchschnitt selber wird selten zur Trendbestimmung eingesetzt. Ein Beispiel einer Glättung mit unterschiedlichem q ist in Figur 4.7 zu sehen. Ein weiteres Beispiel sind a_r 's, welche als Folgenglieder in der ausmultiplizierten Summe von $(\frac{1}{2} + \frac{1}{2})^{2q}$ vorkommen. Mit $q = 1$ hat man $a_{-1} = 1/4, a_0 = 1/2, a_1 = 1/4$. Mit grossem q erhalten wir approximativ eine Normalverteilung. Wenn symmetrische Filter eingesetzt werden, hat man ein Problem am Anfang und am Schluss der Zeitreihe. Die Methoden dies unter Kontrolle zu kriegen reichen von "von Auge weiterfahren" zu Prognosewerten und vielen anderen Versuchen. Ein schönes Beispiel eines nicht-symmetrischen Filters ist die exponentielle Glättung. Sie hat den Vorteil,

dass sie nur mit vergangenen Werten arbeitet; das Randproblem entfällt (ausser am Anfang). Mit einer Konstanten α , $0 < \alpha < 1$ ist dieser Filter

$$y_t = \sum_{j=0}^{\infty} \alpha(1-\alpha)^j x_{t-j}.$$

Die Gewichte fallen geometrisch ab und summieren auf 1.

Wie wählt man "den richtigen Filter"? Die Kriterien, welche sich Spezialisten dazu bereitlegen, benötigen ein vertieftes Studium der Zeitreihenanalyse im Frequenzbereich (wir behandeln dieses Gebiet hier nicht).

Was geschieht, wenn man mehrere Filter in Serie hintereinander schaltet? Der Input ist die Reihe (x_t) , der erste Filter mit Gewichten (a_r) produziert den (Zwischen-)Output (y_t) ; dieser Output ist der neue Input für den zweiten Filter mit Gewichten (b_r) . Den Output des zweiten Filters bezeichnen wir mit (z_t) . Wir wollen jetzt untersuchen, durch welche Gewichte (c_r) der Input (x_t) in den Output (z_t) umgewandelt wird. Dies ist eine einfache Rechnung:

$$z_t = \sum_v b_v y_{t+v} = \sum_v b_v \left(\sum_u a_u x_{t+v+u} \right) = \sum_u \sum_v b_v a_u x_{t+v+u} = \sum_r \left(\sum_v b_v a_{r-v} \right) x_{t+r} = \sum_r c_r x_{t+r}.$$

Die Gewichte des kombinierten Filters sind also $c_r = \sum_v b_v a_{r-v}$. Die Summation erstreckt sich über alle möglichen Indexkombinationen. Es handelt sich also um die *Faltung* der beiden Folgen. Die Faltung ist kommutativ und assoziativ. Auch eine einfache Filtration kann als Faltung einer Zeitreihe mit dem Filter aufgefasst werden.

4.2.7.3 Differenzenbildung

Eine einfache Methode, Trends aus den Daten zu entfernen, ist die Bildung der Differenzen. Dies macht man eventuell mehrmals, bis der Trend wirklich nicht mehr vorhanden ist; bei nicht-saisonalen Daten reicht meist das einmalige Bilden der Differenzen. Die Bildung der Differenzen erfolgt nach folgendem Gesetz:

$$y_t = x_{t+1} - x_t =: \Delta x_{t+1}.$$

Diese Differenzenbildung kann man auch als Filtration auffassen; wähle $a_1 = 1, a_0 = -1$. Wenn man die Differenzenbildung zweimal durchführt, so entspricht dies folgender Filterung:

$$z_t = \Delta^2 x_{t+2} = \Delta x_{t+2} - \Delta x_{t+1} = x_{t+2} - 2x_{t+1} + x_t.$$

Dann sind die Gewichte: $c_2 = 1, c_1 = -2, c_0 = 1$. Einfache Beispiele sind: für die Folge der natürlichen Zahlen erhält man nach einmaliger Differenzenbildung die konstante Folge mit Gliedern gleich 1; für die Folge der Quadratzahlen erhält man nach zweifacher Differenzenbildung die konstante Folge mit Gliedern gleich 2. Folglich ist für lineare Trends die einmalige Differenzenbildung ausreichend; für quadratische Trends ist eine zweifache Differenzenbildung notwendig. Ein Beispiel dazu findet man in Figur 4.8.

4.2.8 Wie behandelt man Zeitreihen mit saisonalen Schwankungen

Wir haben bereits die saisonalen Schwankungen kennengelernt; sie haben im Allgemeinen die Periode 1. Bei der Betrachtung von Transformationen haben wir additive- und multiplikative Saisonmuster gestreift. Es gibt drei Haupttypen von Modellen für saisonale Schwankungen:

$$A : X_t = m_t + s_t + \epsilon_t$$

$$B : X_t = m_t s_t + \epsilon_t$$

$$C : X_t = m_t s_t \epsilon_t.$$

Dabei ist m_t der Mittelwert, s_t die Saisonkomponente und ϵ_t der Fehler. Modell A ist der additive Fall; während Modelle B und C beide multiplikativ sind. Modell C kann (im Gegensatz zu B) einfach durch Logarithmierung in ein additives Modell vom Typ A umgewandelt werden. Die Modelle vom Typ A sind einfacher zu behandeln. Die einfachste Methode, herauszufinden, ob ein Modell vom Typ A, B oder C angebracht ist, ist die Betrachtung des Plots der Daten gegen die Zeit. Ein Saisoneffekt s_t wird sich langsam mit der Zeit bewegen, es sollte etwa gelten dass $s_t \doteq s_{t-i}$, wenn in einem Jahr i Messungen vorgenommen werden. Man normiert in den additiven Modellen den Saisoneffekt derart dass

$$\sum_{j=1}^i s_j = 0$$

respektive

$$\prod_{j=1}^i s_j = 1$$

in den multiplikativen Modellen. In der Praxis hat man Schwierigkeiten, wenn der Saisoneffekt zunimmt - aber nicht wie in einem multiplikativen Modell sondern schwächer: zwischen multiplikativen- und additiven Modellen.

Analog zur Situation mit den Trends muss man auch hier unterscheiden, ob man bei Daten mit Saisoneffekten a) die Saisonkomponente einfach messen will oder/und b) die Saisonkomponente entfernen will.

In Daten, welche praktisch keinen Trend aufweisen, ist folgendes einfaches Verfahren angebracht: Um beispielsweise in einem additiven Modell die Saisonkomponente für den Januar zu messen, nimmt man den Durchschnitt aller Werte vom Januar über alle Jahre und subtrahiert davon den totalen Durchschnitt über das ganze Jahr. Bei multiplikativen Modellen muss man den Januarwert nehmen und durch den Jahresdurchschnitt dividieren.

Hat man hingegen eine Zeitreihe, welche einen klaren Trend aufweist, muss man speziellere Methoden benutzen: Wenn man monatlich erhobene Daten hat, so kann man den Saisoneffekt mit folgendem Ansatz eliminieren:

$$m_t = \frac{\frac{1}{2}x_{t-6} + x_{t-5} + \dots + x_{t+5} + \frac{1}{2}x_{t+6}}{12}.$$

Die Summe der Koeffizienten ist hier 1. Man hat damit also den Jahresdurchschnitt genommen, wenn man die Mitte des Jahres bei t ansetzt und ein halbes Jahr zurück und nach vorne schaut. Man kann nicht 13 Monate nehmen, weil dann 1 Monat doppelt gezählt würde. Falls wir nur Quartalsdaten zur Verfügung haben, so kann man die Saison folgendermassen eliminieren:

$$m_t = \frac{\frac{1}{2}x_{t-2} + x_{t-1} + x_t + x_{t+1} + \frac{1}{2}x_{t+2}}{4}.$$

Der Saisonbeitrag selber wird dann mit $x_t - m_t$ (additiv) bzw. x_t/m_t (multiplikativ) geschätzt. Eine andere Methode, um den Saisonbeitrag zum verschwinden zu bringen ist die geschickte Bildung von Differenzen:

$$y_t = \Delta_{12}x_t = x_t - x_{t-12}.$$

Es gibt auch Kalendereffekte (29. Februar, Anzahl Arbeitstage, Ostern im März und so weiter). Diese können die Zeitreihen ebenfalls beeinflussen. Ob saisonale Muster in Finanzdaten vorkommen (und wie) ist umstritten. Es gibt aber zahlreiche "Daumenregeln" von Händlern wie "sell in October and buy in January". Folglich sollte der Saisoneffekt mindestens als Hypothese formuliert werden. Über den wahren Gehalt dieser Regeln streitet man sich.

4.2.9 Autokorrelation

Für die Untersuchung von Zeitreihen haben sich die sogenannten "empirischen Autokorrelationen" als zentral wichtig herausgestellt. Sie geben die Korrelation zwischen Daten an, welche unterschiedlich weit voneinander

entfernt sind. Die empirischen Autokorrelationen helfen uns später auch, ein geeignetes theoretisches Modell an die Zeitreihe anzupassen. Wenn wir N Paare von Realisationen von Zufallsgrößen (X, Y) haben, dann ist der "normale" Korrelationskoeffizient (im Gegensatz zu Autokorrelationskoeffizient) gegeben durch

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}. \quad (4.2)$$

Mit dieser Definition im Hintergrund werden wir nun analog eine Autokorrelationen in Zeitreihen definieren. Das "Auto" weist darauf hin, dass wir die *selben* Daten, mehrmals verschoben, auf Korrelationen hin untersuchen. Sei x_1, x_2, \dots, x_N eine Realisationen einer Zeitreihe der Länge N . Wir können jetzt $N - 1$ Paare von aufeinander folgenden Daten bilden: $(x_1, x_2), (x_2, x_3), (x_3, x_4), \dots, (x_{N-1}, x_N)$. Jetzt betrachtet man in dieser Folge von $N - 1$ Paaren jeweils die erste Komponente als Realisationen einer Zufallsgröße und die zweiten Komponenten als Realisationen einer zweiten Zufallsgrößen. Die Korrelation zwischen x_t und x_{t+1} ist dann

$$r_1 = \frac{\sum_{t=1}^{N-1} (x_t - \bar{x}_{(1)})(x_{t+1} - \bar{x}_{(2)})}{\sqrt{\sum_{t=1}^{N-1} (x_t - \bar{x}_{(1)})^2 \sum_{t=1}^{N-1} (x_{t+1} - \bar{x}_{(2)})^2}}. \quad (4.3)$$

Dabei bezeichne

$$\bar{x}_{(1)} := \sum_{t=1}^{N-1} x_t / (N - 1)$$

den Mittelwert der ersten $N - 1$ Beobachtungen und

$$\bar{x}_{(2)} := \sum_{t=2}^N x_t / (N - 1)$$

den Mittelwert der letzten $N - 1$ Beobachtungen. Gleichung (4.3) ist nicht ganz einfach zu berechnen. Da jedoch, vor allem bei grossem N (wir haben stationäre Zeitreihen), die beiden Mittelwerte $\bar{x}_{(1)}$ und $\bar{x}_{(2)}$ etwa gleich sind, ersetzt man (4.3) häufig durch

$$r_1 = \frac{\sum_{t=1}^{N-1} (x_t - \bar{x})(x_{t+1} - \bar{x})}{(N - 1) \sum_{t=1}^N (x_t - \bar{x})^2 / N}. \quad (4.4)$$

Hier bezeichnet \bar{x} den gewöhnlichen Mittelwert $\bar{x} = \sum_{t=1}^N x_t / N$. Häufig wird auch der Faktor $N / (N - 1)$ gekippt, da er ja bei grossen N fast 1 ist. Dann kommt man zu der Form, welche wir von jetzt an wirklich benutzen werden:

$$r_1 = \frac{\sum_{t=1}^{N-1} (x_t - \bar{x})(x_{t+1} - \bar{x})}{\sum_{t=1}^N (x_t - \bar{x})^2}. \quad (4.5)$$

Analog werden wir die Autokorrelationen über grössere Distanzen folgendermassen definieren:

$$r_k = \frac{\sum_{t=1}^{N-k} (x_t - \bar{x})(x_{t+k} - \bar{x})}{\sum_{t=1}^N (x_t - \bar{x})^2}. \quad (4.6)$$

k wird englisch als "Lag" bezeichnet. Praktisch berechnet man die Autokorrelationen am Besten über die Autokovarianzen c_k :

$$c_k = \frac{1}{N} \sum_{t=1}^{N-k} (x_t - \bar{x})(x_{t+k} - \bar{x}). \quad (4.7)$$

Dies ist die Autokovarianz mit Lag k . Die Autokorrelationen erhält man dann mit $r_k = c_k / c_0$; insbesondere gilt $r_0 = 1$. In gewissen Büchern wird in (4.7) durch $N - k$ geteilt anstelle von N . Beide Schätzer der Kovarianz sind jedoch nur asymptotisch erwartungstreu.

Wenn man die Folge der r_k berechnet hat, so stellt man sie am Besten im sogenannten *Korrelogramm* graphisch dar. Das Korrelogramm ist ein Plot der r_k gegen den Lag (die Zeitdifferenzen). Dieser Plot ist der zweite Plot den man bei einer Zeitreihenanalyse anfertigen sollte (nach der Zeitreihe selber). In Figuren 4.9, 4.10 und 4.11 (alle aus Chatfield) sind jeweils spezielle Zeitreihen selber und danach die Korrelogramme dargestellt. Wir werden uns nun der Frage zuwenden, wie man diese Korrelogramme interpretieren darf. Diese Interpretation ist nämlich mit gewissen Gefahren verbunden.

a) Völlig zufällige Zeitreihe: Wenn eine Zeitreihe eine Realisation einer Folge von *unabhängigen* Zufallsgrößen ist, so sollten wir für grosse N für r_k Werte um Null erhalten, wenn $k \neq 0$. Wir werden später sehen, dass unter schwachen Bedingungen (Gaussprozess, absolut summierbare Autokovarianzfunktion) gilt: r_k hat eine $\mathcal{N}(-1/N, 1/N)$ -Verteilung. Wenn wir also eine völlig zufällige Zeitreihe haben, so sollten 19 aus 20 der r_k 's innerhalb eines Bandes $\pm 2/\sqrt{N}$ liegen (5% Signifikanzniveau). Also sollte auch 1 aus 20 der r_k 's ausserhalb dieses Bandes liegen. Damit ist also Vorsicht geboten. Nicht jedes signifikant von 0 verschiedene r_k bedeutet ein statistisch relevantes Resultat.

b) Korrelation über kurze Lags: In vielen stationären Zeitreihen beobachtet man eine sogenannte "short-term correlation": r_1 ist relativ gross, auch die weiteren r_i 's sind noch signifikant von 0 verschieden, fallen aber langsam gegen null ab. Bei grösseren Lags sind die r_i 's praktisch null. Ein Beispiel davon findet man in Figur 4.9. Die Zeitreihe, welche ein derartiges Korrelogramm aufweist, hat die Eigenschaft, dass wenn ein Wert der Zeitreihe grösser als der Mittelwert ist, so wird der nächste (und auch weiter nachfolgende) auch eher grösser als der Mittelwert sein (analog mit Werten kleiner als der Mittelwert).

c) Alternierende Zeitreihe: Wenn in einer Zeitreihe jeweils ein Wert über dem Mittel ist und dann der nächste unter dem Mittel, so nennen wir die Zeitreihe alternierend. Das Korrelogramm hat dann auch ein alternierendes Muster: r_1 ist negativ, r_2 wieder positiv, r_3 wieder negativ und so weiter. Häufig gehen die absoluten Werte doch gegen Null. In Figur 4.10 haben wir eine derartige Zeitreihe mit dazugehörigem Korrelogramm. In Aufgabe 24 auf Blatt 10 wird ein solches Korrelogramm berechnet.

d) Nichtstationäre Zeitreihe: In Figur 4.11 sehen wir eine nichtstationäre Zeitreihe zusammen mit dem dazugehörigen Korrelogramm. Der Grund, weshalb die r_k 's so lange bei eins bleiben und erst für sehr grosse Werte von k gegen null gehen wird an der Tafel graphisch entwickelt. Aus einem solchen Korrelogramm kann man nichts herauslesen. Der Trend dominiert die ganze Entwicklung. Das Korrelogramm macht nur Sinn für stationäre Zeitreihen. Deshalb sollte man den Trend zuerst entfernen. In Aufgabe 25 auf Blatt 10 wird ein solches Korrelogramm berechnet.

e) Saisonale Fluktuationen: Wenn eine Zeitreihe saisonale Schwankungen aufweist, so wird auch das Korrelogramm dieser Zeitreihe saisonale Schwankungen aufweisen (und zwar mit derselben Periode). Wenn man monatliche Werte gemessen hat, so wird r_6 negativ sein (mit grossem Absolutbetrag) und r_{12} gross und positiv. Auch hier gilt, dass die Betrachtung des Korrelogramms mit saisonalen Schwankungen keinen Sinn macht. Man muss zuerst die Saisonkomponente entfernen.

f) "Ausreisser": Ausreisser können einen gewaltigen Einfluss auf die r_k 's haben. Es kann deshalb sinnvoll sein, diese Ausreisser zu korrigieren. Man bedenke, dass es mit einem Ausreisser x_t immer zwei Paare (x_t, x_{t+k}) und (x_{t-k}, x_t) gibt, deren Differenz grosse Werte annimmt, womit die Varianz sehr gross wird. Damit werden bei der Berechnung der r_k 's alle Summanden klein, welche den Ausreisser nicht beinhalten.

Offenbar ist die Interpretation des Korrelogramms kein so einfaches Unterfangen, wie man meinen könnte. Wir werden jetzt einerseits theoretische Modelle in 4.3 entwickeln und deren Korrelogramme berechnen. Andererseits werden wir in 4.4 auch die statistischen Eigenschaften der Korrelationskoeffizienten untersuchen. Erst dann werden wir die Korrelogramme gut interpretieren können.

4.3 Modelle für Zeitreihen

4.3.1 Stochastische Prozesse

Bei den einführenden Bemerkungen haben wir darauf hingewiesen, dass wir mit einer Zeitreihe zwar durchaus eventuell sehr viele Beobachtungen haben (grosses N , vor allem in der Finanzmathematik), aber nur *eine Realisation*. Wir werden von jetzt an aber *unterstellen*, dass diese Zeitreihe eben eine *Realisation* eines stochastischen Prozesses X_t ist. Damit wenden wir uns also den erklärenden Modellen zu. In der Zeitreihenanalyse sind die ersten und zweiten Momente von stochastischen Prozessen (X_t) zentral: die Mittelwertfunktion, die Varianzfunktion und die Autokovarianzfunktion:

Definition 4.2 [Momente einer Zeitreihe] Die Mittelwertfunktion einer Zeitreihe X_t ist definiert als

$$\mu(t) := E[X_t].$$

Die Varianzfunktion einer Zeitreihe ist definiert als

$$\sigma^2(t) := V[X_t].$$

Die Autokovarianzfunktion einer Zeitreihe ist definiert als

$$\gamma(s, t) := E[(X_s - \mu(s))(X_t - \mu(t))].$$

Im Gegensatz zu Momenten von Zufallsgrössen muss man bei Zufallsprozessen auch die Autokovarianzfunktion angeben. Man beachte, dass $\gamma(t, t) = \sigma^2(t)$.

4.3.2 Stationäre Prozesse

Wir haben bereits in 4.2.4 eine intuitive Vorstellung von Stationarität bei Zeitreihen entwickelt. Es sollten a) der Mittelwert gleich bleiben, b) die Varianz gleich bleiben und c) keine periodischen Schwankungen vorkommen. Wir werden dies nun sauber definieren:

Definition 4.3 [Strenge Stationarität] Wir nennen einen stochastischen Prozess X_t streng stationär, wenn für jeden Lag τ und jedes $m \leq n$ gilt: die Vektoren

$$(X_m, X_{m+1}, \dots, X_n)$$

und

$$(X_{m+\tau}, X_{m+1+\tau}, \dots, X_{n+\tau})$$

haben die gleiche Verteilung.

Wenn nun die ersten und zweiten Momente existieren, so folgt aus der strengen Stationarität sofort, dass sowohl

a) der Mittelwert stationär ist: $\mu(t) = \mu$, (wähle $m = n = 1$ in Definition 4.3)

b) die Varianz stationär ist: $\sigma^2(t) = \sigma^2$, gleiches Argument wie a), und

c) die Autokovarianzfunktion stationär ist, das heisst nur vom Lag abhängt. Wir können dann schreiben: $\gamma(s, t) =: \gamma(t - s)$.

Da die Autokovarianzfunktion nicht skaleninvariant ist, betrachtet man meist die Autokorrelationsfunktion $\rho(\tau) := \gamma(\tau)/\gamma(0)$. Das empirische Pendant haben wir in 4.2 bereits kennengelernt.

Für Leser ohne profunde Kenntnisse der stochastischen Prozesse mag die Frage auftauchen, ob denn solche Prozesse überhaupt existieren. Triviale Beispiele sind i.i.d. Folgen von normalverteilten Zufallsgrössen. In der Praxis der Zeitreihenanalyse schwächt man die Forderung nach strenger Stationarität aber ab:

Definition 4.4 [Schwache Stationarität] Wir nennen einen stochastischen Prozess X_t schwach stationär (oder einfach stationär), wenn er sowohl mittelwert- wie auch autokorrelationsstationär (und damit auch varianzstationär) ist.

Bei Gauss'schen Prozessen reicht die Kenntnis der ersten und zweiten Momente bereits aus, um die Verteilung des gesamten Prozesses zu kennen. Viele Theoreme benötigen nur die ersten und zweiten Momente. Damit macht es also durchaus Sinn, die Forderung nach Stationarität derart abzuschwächen.

4.3.3 Die Autokorrelationsfunktion

Wir haben bereits in 4.2 bei der Beschreibung von Zeitreihen gesehen, dass die empirische Autokorrelationsfunktion wichtige Informationen über eine Zeitreihe liefern kann. Wir betrachten jetzt Modelle von Zeitreihen. Auch hier ist die Autokorrelationsfunktion zentral wichtig. Wir werden zuerst allgemeine Aussagen über die Autokorrelationsfunktion herleiten, bevor wir spezielle Modelle untersuchen.

Wir setzen voraus, dass ein stationärer stochastischer Prozess X_t Mittelwert μ , Varianz σ^2 , Autokovarianzfunktion $\gamma(\tau)$ und Autokorrelationsfunktion $\rho(\tau)$ hat. Es gilt damit

$$\rho(\tau) = \gamma(\tau)/\gamma(0) = \gamma(\tau)/\sigma^2.$$

Man beachte, dass immer $\rho(0) = 1$. Es folgen jetzt 3 zentrale Eigenschaften der Autokorrelationsfunktion.

Eigenschaft 1: $\rho(\tau) = \rho(-\tau)$ (die Autokorrelationsfunktion ist eine gerade Funktion bezüglich des Lags τ)

Diese Eigenschaft besagt einfach, dass die Korrelation zwischen X_t und $X_{t+\tau}$ gleich ist der Korrelation zwischen X_t und $X_{t-\tau}$. Da $\gamma(\tau) = \rho(\tau)\sigma^2$ beweist man dies am besten via

$$\gamma(\tau) = Cov[X_t, X_{t+\tau}] = Cov[X_{t-\tau}, X_t] = \gamma(-\tau).$$

Beim zweiten Gleichheitszeichen haben wir benutzt, dass der stochastische Prozess stationär ist.

Eigenschaft 2: $|\rho(\tau)| \leq 1$

Diese Eigenschaft gilt bereits bei "normalen" Korrelationen. Sie folgt sofort aus der Cauchy-Schwarz'schen Ungleichung.

Eigenschaft 3: Keine Eineindeutigkeit

Ein gegebener stochastischer Prozess hat selbstverständlich nur eine Autokorrelationsfunktion. Es ist aber leider im allgemeinen nicht so, dass es zu einer gegebenen Autokorrelationsfunktion nur einen stochastischen Prozess geben kann. Damit ist es also nicht klar, welches Modell man wählen soll, um eine gegebene Realisation einer Zeitreihe zu erklären.

4.3.4 Wichtige Modelle von Zeitreihen

4.3.4.1 Vollkommen zufälliger Prozess, "White Noise"

Ein vollkommen zufälliger Prozess (Z_t) ist eine Folge von i.i.d. Zufallsgrößen. Damit sind auch Mittelwerte, Varianzen und Kovarianzen gleich (falls sie existieren). Die Kovarianzen sind jeweils für Lags $\neq 0$ gleich 0 (wegen der Unabhängigkeit). Dieser Prozess ist stationär, sogar streng stationär. Diesen Prozess nennt man auch "White Noise". Es muss keine normalverteilte Folge sein!

4.3.4.2 Random Walk

Sei $(Z_i)_{1 \leq i \leq t}$ eine Folge von i.i.d. Zufallsgrößen (White Noise). Es gelte

$$E[Z_i] = \mu$$

und

$$V[Z_i] = \sigma_Z^2.$$

Dann definieren wir

$$S_t = S_{t-1} + Z_t, \tag{4.8}$$

wobei wir für $t = 0$ den Wert $S_0 = 0$ setzen. Eine analoge Definition dieses sogenannten Random Walks direkt über die Summe ist

$$S_t = \sum_{i=1}^t Z_i.$$

Wir erhalten $E[S_t] = t\mu$ und $V[S_t] = t\sigma_Z^2$. Da sowohl Varianz wie auch Mittelwert sich mit t ändern, ist S_t nicht stationär. Interessant ist jedoch hier (nicht ganz unerwartet), dass die ersten Differenzen eines Random Walks:

$$\Delta S_t := S_t - S_{t-1} = Z_t$$

einen White-Noise Prozess ergeben und damit insbesondere wieder stationär sind.

Eine bekannte, einfache Anwendung ist die Modellierung von Aktienkursen an aufeinanderfolgenden Tagen mittels eines Random Walks. Da ein Random Walk auch negative Werte annehmen kann, ist dieses Modell nur lokal zu gebrauchen. Wenn man aber die Logreturns anschaut und Z_i normalverteilt wählt (mit den richtigen Parametern), so hat man genau das Modell von Samuelson (siehe Kapitel 3).

In der Mathematik wird unter einem Random Walk meist ein Prozess verstanden, der nur "einen Schritt nach oben gehen kann oder einen nach unten" (diskreter Zustandsraum). Wir haben aber hier an Z_i nicht derart einschränkende Anforderungen formuliert.

4.3.4.3 Moving Average (MA)

Sei (Z_t) ein White Noise Prozess; der Mittelwert sei 0 und die Varianz σ_Z^2 . Wir nennen dann (X_t) einen Moving Average Prozess der Ordnung q (MA[q]), wenn er sich in der Form

$$X_t = \beta_0 Z_t + \beta_1 Z_{t-1} + \beta_2 Z_{t-2} + \dots + \beta_q Z_{t-q} \tag{4.9}$$

darstellen lässt. Dabei sind die β_i 's Konstanten; die Z 's werden meist derart normiert, dass $\beta_0 = 1$. Wir erhalten sofort, dass

$$E[X_t] = 0$$

und

$$V[X_t] = \sigma_Z^2 \sum_{i=0}^q \beta_i^2,$$

da die Z_t unabhängig sind. Schreiten wir zur Berechnung der Autokorrelationsfunktion: Es gilt für die Autokovarianzfunktion:

$$\gamma(k) = Cov(X_t, X_{t+k}) = Cov(\beta_0 Z_t + \beta_1 Z_{t-1} + \dots + \beta_q Z_{t-q}, \beta_0 Z_{t+k} + \beta_1 Z_{t+k-1} + \dots + \beta_q Z_{t+k-q}).$$

Es sind jetzt drei Fälle zu unterscheiden; man beachte, dass die (Z_t) i.i.d. sind:

$$= \begin{cases} 0 & \text{falls } k > q \\ \sigma_Z^2 \sum_{i=0}^{q-k} \beta_i \beta_{i+k} & \text{falls } 0 \leq k \leq q \\ \gamma(-k) & \text{falls } k < 0. \end{cases}$$

$\gamma(k)$ ist offenbar unabhängig von t . Zudem ist der Mittelwert konstant 0. Damit ist dieser Prozess stationär. Wenn die Z 's sogar normalverteilt sind, dann sind auch die X 's normalverteilt und wir haben sogar strenge Stationarität.

Die Autokorrelationsfunktion ist demnach

$$\rho(k) = \begin{cases} 0 & \text{falls } k > q \\ \sum_{i=0}^{q-k} \beta_i \beta_{i+k} / \sum_{i=0}^q \beta_i^2 & \text{falls } 0 \leq k \leq q \\ \rho(-k) & \text{falls } k < 0. \end{cases}$$

Für spätere Untersuchungen zentral ist, dass die Autokorrelationsfunktion eines $MA(q)$ -Prozesses bei Lag q *abbricht*. Betrachten wir die Autokorrelationsfunktion eines $MA(1)$ -Prozesses mit $\beta_0 = 1$. Wir erhalten

$$\rho(k) = \begin{cases} 1 & \text{falls } k = 0 \\ \beta_1 / (1 + \beta_1^2) & \text{falls } k = \pm 1 \\ 0 & \text{sonst.} \end{cases}$$

Wenn wir nur fordern, dass ein MA -Prozess endlicher Ordnung ($q < \infty$) stationär ist, so müssen wir keine einschränkenden Anforderungen an die β 's machen. Hingegen möchte man oft eine Eigenschaft namens *Invertierbarkeit* garantiert wissen. Dann müssen aber Einschränkungen an die β 's gemacht werden. Was ist unter Invertierbarkeit zu verstehen? Betrachten wir dazu die beiden $MA(1)$ -Prozesse

$$A : X_t = Z_t + \theta Z_{t-1}$$

und

$$B : X_t = Z_t + \frac{1}{\theta} Z_{t-1}.$$

Diese beiden (verschiedenen) Prozesse haben die gleiche Autokorrelationsfunktion (überprüfen)! Folglich können wir nicht mal innerhalb der Klasse der $MA(1)$ -Prozesse aufgrund der Autokorrelationsfunktion eindeutig auf den Prozess selber (die ihn definierende Gleichung) schließen. Wenn wir jetzt aber in A und B die Gleichungen dahingehend invertieren, dass wir nicht die X 's durch die Z 's ausdrücken, sondern umgekehrt die Z 's durch die X 's, so erhalten wir

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

und

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

Sei nun $|\theta| < 1$. Dann konvergiert die Serie in A , aber nicht diejenige in B (X ist stationär). Wenn man also die Residuen schätzen möchte (aus den X 's; man beobachtet ja nur die X 's), so ist dies nur mit einem Modell vom Typ A möglich. Wir werden in 4.4 tatsächlich die Residuen schätzen müssen. Wir nennen Modell A invertierbar; Modell B ist nicht invertierbar. Wenn wir Invertierbarkeit fordern, dann gibt es für eine gegebene Autokorrelationsfunktion eines MA -Prozesses nur einen MA -Prozess.

Wenn man das Konzept der Invertierbarkeit auf MA -Prozesse beliebiger Ordnung ausdehnen will, so bietet sich die Notation mit Hilfe des sogenannten Backshift Operators B an: er wird folgendermassen definiert:

$$B^j X_t := X_{t-j}$$

für alle j . In der Tat verschiebt (*shift*) er die Zeitreihe X um j Einheiten in die Vergangenheit (*back*). Mit Hilfe des Backshift-Operators können wir Gleichung (4.9) folgendermassen kompakt umschreiben:

$$X_t = (\beta_0 + \beta_1 B + \dots + \beta_q B^q) Z_t =: \theta(B) Z_t.$$

Dabei ist θ ein Polynom der Ordnung q in B . Ein MA[q]-Prozess ist invertierbar, wenn alle Wurzeln der Gleichung

$$\theta(B) = \beta_0 + \beta_1 B + \dots + \beta_q B^q = 0$$

ausserhalb des Einheitskreises liegen (siehe "Time Series forecasting and control" von G. Box und G. Jenkins, 1994). Dabei betrachtet man B als eine komplexe Variable. Beispielsweise hat man für $q = 1$ die Gleichung

$$1 + \theta B = 0.$$

Dies hat eine Lösung $B = -1/\theta$. Sobald $|\theta| < 1$, liegt diese Wurzel ausserhalb des Einheitskreises. Dies haben wir weiter oben schon gesehen.

Der Einsatz von MA-Prozessen in der Ökonomie ist vielfältig und kann theoretisch begründet werden. Entscheide der Regierungen oder Notenbanken, Änderungen von Rohstoffpreisen und so weiter haben nicht nur einen direkten Effekt auf wichtige Kennziffern (Bruttosozialprodukt), sondern wirken eventuell lange nach, eh sie vollständig ausklingen.

In Gleichung (4.9) kann man auch eine Konstante μ auf der rechten Seite hinzufügen, damit der Prozess Mittelwert μ hat. Die Autokovarianzfunktion wird nicht verändert.

4.3.4.3 Autoregressive Prozesse (AR)

Sei (Z_t) wieder ein White-Noise-Prozess. Der Mittelwert sei jeweils 0 und die Varianz $V(Z_t) = \sigma_Z^2$. Dann nennen wir einen Prozess (X_t) Autoregressiver Prozess der Ordnung p (AR[p]), wenn er eine Darstellung der Form

$$X_t = \alpha_1 X_{t-1} + \dots + \alpha_p X_{t-p} + Z_t. \quad (4.10)$$

hat. Dies ist in der Tat etwas wie eine "Regression". Nur sind die Regressoren alte Werte der Zeitreihe; X wird auf sich selbst regressiert, deshalb *Autoregressiver* Prozess. Wir werden immer p für die Ordnung der autoregressiven Prozesse benutzen und q für die MA-Prozesse; ebenso Parameter α für AR-Prozesse und β für MA-Prozesse; ebenso θ für das Polynom des Backshift-Operators bei einem MA-Prozess und ϕ bei einem AR-Prozess.

4.3.4.3.1 AR[1]

Der Einfachheit halber werden wir zuerst die AR[1]-Prozesse ausführlich untersuchen. Wir haben also

$$X_t = \alpha X_{t-1} + Z_t. \quad (4.11)$$

Wir werden jetzt mit Hilfe von (4.11) X_t sukzessive mit Hilfe von immer mehr Gliedern des zu Grunde liegenden White Noise Prozesses Z ausdrücken und erhalten:

$$X_t = \alpha(\alpha X_{t-2} + Z_{t-1}) + Z_t = \alpha^2(\alpha X_{t-3} + Z_{t-2}) + \alpha Z_{t-1} + Z_t.$$

Wenn wir so fortfahren, erhalten wir eine Darstellung von X_t als MA[∞]-Prozess. Wir fordern noch, dass $|\alpha| < 1$. Dann gilt:

$$X_t = Z_t + \alpha Z_{t-1} + \alpha^2 Z_{t-2} + \dots$$

Die Forderung nach $|\alpha| < 1$ ist offensichtlich notwendig, damit der Prozess (X_t) nicht explodiert. Diese Dualität zwischen MA und AR-Prozessen wird uns noch dienlich sein. Gleichung (4.11) kann auch unter Benutzung des Backshift-Operators B folgendermassen dargestellt werden:

$$(1 - \alpha B)X_t = Z_t.$$

Da Polynome in B die gleichen Eigenschaften wie gewöhnliche Polynome haben, können wir folgendermassen fortfahren:

$$X_t = Z_t / (1 - \alpha B) = (1 + \alpha B + \alpha^2 B^2 + \dots)Z_t = Z_t + \alpha Z_{t-1} + \alpha^2 Z_{t-2} + \dots$$

Jetzt ist klar, dass gelten muss:

$$E[X_t] = 0,$$

$$V[X_t] = \sigma_Z^2(1 + \alpha^2 + \alpha^4 + \dots).$$

Man sieht auch hier, dass die Varianz nur dann endlich ist, wenn $|\alpha| < 1$. Genauer gilt dann

$$V[X_t] := \sigma_X^2 = \sigma_Z^2/(1 - \alpha^2).$$

Wir berechnen noch die Autokovarianzfunktion.

$$\gamma(k) = E[X_t X_{t+k}] = E\left[\left(\sum_{i=0}^{\infty} \alpha^i Z_{t-i}\right)\left(\sum_{j=0}^{\infty} \alpha^j Z_{t+k-j}\right)\right] = \sigma_Z^2 \sum_{i=0}^{\infty} \alpha^i \alpha^{k+i}$$

für $k \geq 0$. Für $|\alpha| < 1$ konvergiert dieser Ausdruck gegen

$$\gamma(k) = \alpha^k \sigma_Z^2 / (1 - \alpha^2) = \alpha^k \sigma_X^2.$$

Für $k < 0$ gilt natürlich $\gamma(k) = \gamma(-k)$ wegen Eigenschaft 1 der theoretischen Autokorrelationsfunktion. Die Autokovarianz hängt nicht von t ab. Damit ist ein AR[1]-Prozess insbesondere stationär (für $|\alpha| < 1$). Die Autokorrelationsfunktion lässt sich kompakt folgendermassen hinschreiben:

$$\rho(k) = \alpha^{|k|}.$$

In Figur 4.12 sind die Autokorrelationsfunktionen von drei AR[1]-Prozessen mit unterschiedlichem α (0.8, 0.3, -0.8) aufgezeichnet. Bezeichnenderweise fällt die Autokorrelationsfunktion geometrisch ab; zudem alterniert sie bei negativem α .

4.3.4.3.2 AR[p]

Wir haben den AR[1]-Prozess als MA-Prozess darstellen können. Genau gleich kann man auch einen AR[p]-Prozess als MA-Prozess darstellen: Entweder man substituiert sukzessive eine Darstellung mit verschiedener Zeit in die andere, oder man stellt den AR-Prozess mit Hilfe des Backshift-Operators B dar. (4.10) wird dann zu

$$(1 - \alpha_1 B - \dots - \alpha_p B^p) X_t = Z_t$$

oder

$$X_t = Z_t / (1 - \alpha_1 B - \dots - \alpha_p B^p) = f(B) Z_t,$$

wobei

$$f(B) = (1 - \alpha_1 B - \dots - \alpha_p B^p)^{-1} =: (1 + \beta_1 B + \beta_2 B^2 + \dots).$$

Die β 's müssen dann mittels der α 's durch Koeffizientenvergleich ausgedrückt werden. Dies kann sehr kompliziert werden. Aus der MA-Darstellung folgt dann sofort, dass $E[X_t] = 0$. Die Varianz ist endlich, vorausgesetzt dass $\sum_{i=0}^{\infty} \beta_i^2 < \infty$ (diese Voraussetzung ist somit notwendig für Stationarität). Mit den exakt gleichen Schritten wie bei einem AR[1]-Prozess erhalten wir auch einen Ausdruck für die Autokovarianzfunktion eines AR[p]-Prozesses ($\beta_0 = 1$)

$$\gamma(k) = \sigma_Z^2 \sum_{i=0}^{\infty} \beta_i \beta_{i+k}.$$

Eine hinreichende Bedingung, damit diese Reihe konvergiert (und damit für Stationarität), ist $\sum_i |\beta_i| < \infty$.

Wir haben bereits darauf hingewiesen, dass die Berechnung der β 's schwierig sein kann. Damit ist auch die Berechnung der Autokorrelationsfunktion eventuell sehr umständlich. Ein alternativer Weg geht über die Yule-Walker-Gleichungen. Dabei *geht man davon aus*, dass der autoregressive Prozess stationär *ist*. Man multipliziert dann Gleichung (4.10) mit X_{t-k} , bildet den Erwartungswert und teilt noch durch σ_X^2 , wieder

annehmend dass die Varianz von X_t endlich ist und $E[X_t] = 0$. Wenn man dann noch berücksichtigt dass $\rho(k) = \rho(-k)$, so ergibt sich für alle $k > 0$:

$$\rho(k) = \alpha_1 \rho(k-1) + \dots + \alpha_p \rho(k-p). \quad (4.12)$$

Diese Gleichungen heissen Yule-Walker-Gleichungen. Es ist dies eine Menge von Differenzgleichungen. Die allgemeine Lösung lautet (von der allgemeinen Theorie über Differenzgleichungen)

$$\rho(k) = A_1 \pi_1^{|k|} + \dots + A_p \pi_p^{|k|}; \quad (4.13)$$

dabei sind die (π_i) die Wurzeln der Hilfsgleichung

$$y^p - \alpha_1 y^{p-1} - \dots - \alpha_p = 0. \quad (4.14)$$

Die Konstanten A_i werden derart gewählt, dass $\rho(0) = 1$. Damit muss gelten $\sum A_i = 1$. In (4.12) haben wir aber für $k \in \{1, 2, \dots, p-1\}$ weitere $p-1$ Gleichungen, welche erfüllt sein müssen. Dabei verwendet man jeweils, dass $\rho(0) = 1$ und $\rho(k) = \rho(-k)$. Ein Beispiel dazu folgt gleich nachfolgend. In (4.13) sieht man sofort, dass die Autokorrelationen geometrisch abklingen, wenn k grösser wird, vorausgesetzt dass die $|\pi_i| < 1$ für alle i . Dies ist denn auch eine sowohl hinreichende wie auch notwendige Bedingung für Stationarität.

Eine äquivalente Aussage zur Stationarität ist, dass die Wurzeln der Gleichung

$$\phi(B) = 1 - \alpha_1 B - \dots - \alpha_p B^p = 0 \quad (4.15)$$

alle *ausserhalb* des Einheitskreises liegen müssen. Dabei wird B wieder als komplexe Veränderliche aufgefasst. Dieses Resultat findet man mit Beweis wieder in "Time Series forecasting and control" von G. Box und G. Jenkins, 1994. Gleichungen (4.14) und (4.15) sind äquivalent: in (4.14) muss man einfach durch y^p teilen und den Kehrwert der Wurzeln suchen. Deshalb müssen die Wurzeln in (4.14) auch innerhalb des Einheitskreises liegen und in (4.15) ausserhalb.

Betrachten wir als Beispiel den AR[2]-Prozess

$$X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + Z_t.$$

Dann sind also die π_1 und π_2 die Wurzeln der Gleichung

$$y^2 - \alpha_1 y - \alpha_2 = 0.$$

Wir wollen einen stationären AR-Prozess. Dazu muss gelten, dass sowohl $|\pi_1| < 1$ wie auch $|\pi_2| < 1$. Also muss gelten, dass

$$\left| \frac{\alpha_1 \pm \sqrt{\alpha_1^2 + 4\alpha_2}}{2} \right| < 1.$$

Man kann zeigen (an der Tafel in der Vorlesung), dass die stationäre Region für die Wahl der Parameter eines AR[2]-Prozesses durch die drei folgenden Ungleichungen begrenzt wird:

$$\alpha_1 + \alpha_2 < 1,$$

$$\alpha_1 - \alpha_2 > -1,$$

$$\alpha_2 > -1.$$

Dazu gilt, dass wir reelle Lösungen haben, sobald $\alpha_1^2 + 4\alpha_2 > 0$. Wir haben dann einen exponentiellen Abfall der Autokorrelationen. Andernfalls sind die Wurzeln komplex; es kann gezeigt werden, dass man dann für die Autokorrelationsfunktion eine gedämpfte Kosinusschwingung erhält (aus trigonometrischen Gründen ist die Autokorrelationsfunktion natürlich doch wieder reellwertig!). In Figur 4.13 sind vier Beispiele, ein Beispiel

pro Region, aufgezeichnet. Seien nun die Wurzeln π_1 und π_2 reell. Wir wollen die Autokorrelationsfunktion vollständig berechnen; was uns noch fehlt sind die Anfangsbedingungen, das heisst die A_1 und A_2 . Da $\rho(0) = 1$ muss gelten

$$A_1 + A_2 = 1.$$

Dazu kommt noch von der ersten Yule-Walker-Gleichung, dass

$$\rho(1) = \alpha_1\rho(0) + \alpha_2\rho(-1) = \alpha_1 + \alpha_2\rho(1).$$

Damit erhalten wir

$$\rho(1) = \alpha_1/(1 - \alpha_2) = A_1\pi_1 + A_2\pi_2 = A_1\pi_1 + (1 - A_1)\pi_2.$$

Daraus ergibt sich

$$\begin{aligned} A_1 &= [\alpha_1/(1 - \alpha_2) - \pi_2]/(\pi_1 - \pi_2) \\ A_2 &= 1 - A_1. \end{aligned}$$

Im Fall wo $\pi_1 = \pi_2$, kann man jede Kombination von (A_1, A_2) nehmen, sodass $A_1 + A_2 = 1$. Alle führen ja zur selben Gleichung (4.13). Für die Autokorrelationsfunktion eines AR[2]-Prozesses erhält man also im Allgemeinen eine umständliche, komplizierte Formel.

AR-Prozesse sind dort geeignet, wo die neuen Werte etwa den alten entsprechen plus ein Fehler. Man kann auch hier einen Mittelwert einbauen, welcher von 0 verschieden ist. Die Bestimmungsgleichung (4.10) wird dann zu

$$X_t - \mu = \alpha_1(X_{t-1} - \mu) + \dots + \alpha_p(X_{t-p} - \mu) + Z_t.$$

Die Autokorrelationsfunktion wird dadurch nicht tangiert.

4.3.4.4 (Gemischte) ARMA-Prozesse

Die MA- und AR-Prozesse kann man als *die* Grundbausteine der Zeitreihenanalyse bezeichnen. Wir haben dort auch je viele Freiheiten: die Ordnung kann frei gewählt werden; je grösser die Ordnung, desto mehr Parameter stehen uns zur Verfügung, um ein Modell an eine gegebene empirische Reihe anzupassen. Was ist jedoch von einem AR[20]- oder MA[63]-Prozess zu halten? Wenn viele Parameter zur Verfügung stehen, kann man fast jeden endlichen Datensatz an ein Modell anpassen. Man möchte aber ein Modell, bei dem möglichst wenig Parameter vorkommen, welche zudem auch eine reale Bedeutung (Interpretation) haben. Die zentrale Bedeutung von ARMA[p, q]-Prozessen liegt darin, dass wir zum Beispiel mit einem ARMA[1,1]-Prozess (mit 2 Parametern!) viel umfassender Eigenschaften von empirischen Zeitreihen einfangen können als mit den einfachen MA- oder AR-Prozessen. Ein ARMA[p, q]-Prozess besteht aus p AR-Termen und aus q MA-Termen. Die Bestimmungsgleichung lautet

$$X_t = \alpha_1 X_{t-1} + \dots + \alpha_p X_{t-p} + Z_t + \beta_1 Z_{t-1} + \dots + \beta_q Z_{t-q}. \quad (4.16)$$

Diese Gleichung stellt man normalerweise mit Hilfe des Backshift-Operators folgendermassen dar:

$$\phi(B)X_t = \theta(B)Z_t; \quad (4.17)$$

dabei seien

$$\phi(B) = 1 - \alpha_1 B - \dots - \alpha_p B^p$$

und

$$\theta(B) = 1 + \beta_1 B + \dots + \beta_q B^q.$$

Genau wie bei einem AR-Prozess gilt, dass ein ARMA-Prozess stationär ist, falls die α 's so gewählt sind, dass die Wurzeln von

$$\phi(B) = 0$$

ausserhalb des Einheitskreises liegen. Dann haben wir nämlich eine Darstellung als (unendlicher) MA-Prozess, welcher nicht explodiert (siehe Bemerkungen bei AR[p]-Prozessen). Gleichfalls gilt, dass ein ARMA-Prozess invertierbar ist (das heisst, die Residuen Z lassen sich mit Hilfe der gemessenen Werte X ausdrücken), wenn die Wurzeln der Gleichung

$$\theta(B) = 0$$

ausserhalb des Einheitskreises liegen.

Die Berechnung der Autokorrelationsfunktion eines allgemeinen ARMA[p, q]-Prozesses ist umständlich. Man findet sie in Box und Jenkins. Hingegen wollen wir den wichtigen Fall eines ARMA[1,1] doch angeben: Man erhält für die Autokovarianzfunktion die Werte

$$\begin{aligned}\gamma(0) &= \frac{1 + 2\alpha\beta + \beta^2}{1 - \alpha^2} \sigma_Z^2, \\ \gamma(1) &= \frac{(1 + \alpha\beta)(\alpha + \beta)}{1 - \alpha^2} \sigma_Z^2,\end{aligned}$$

und für $k \geq 2$ erhält man die Werte

$$\gamma(k) = \alpha\gamma(k-1)$$

und somit

$$\gamma(k) = \alpha^{k-1}\gamma(1). \quad (4.18)$$

Es ist bezeichnend, dass die Autokovarianzfunktion für Lags grösser als 1 geometrisch abfällt mit Parameter α , wie bei einem AR[1]-Prozess! Bei einem MA[1]-Prozess bricht die Autokovarianzfunktion ja auch ab; sie ist Null für Werte grösser als 1. Diese beiden Eigenschaften von MA- und AR-Prozessen schimmern also auch hier durch. Insbesondere ist dieses Resultat auch konsistent für die Fälle wo $\alpha = 0$ oder $\beta = 0$ (überprüfen!). In Figur 4.14 findet man die Autokorrelationsfunktion eines ARMA[1,1]-Prozesses für diverse Werte von α und β . Die Interpretation dieser Figur ist folgendermassen: Wenn man nur einen AR[1]-Prozess hat mit $\alpha = 0.8$, so würde die Autokorrelationsfunktion bekanntlich mit 0.8^k gegen 0 konvergieren (Fall 2 in Figur 4.14). Der MA-Teil dieses ARMA[1,1]-Prozesses hat aber bei $\beta = +0.5$ eine Verstärkung der Autokorrelation bei 1 zur Folge und für $\beta = -0.5$ eine Abschwächung (MA[1] mit $\beta = -0.5$ hat bei Lag 1 eine *negative* Autokorrelation!). Von Lag 2 weg gehen alle geometrisch gegen 0 (wegen (4.18)).

Bei Prognosemethoden kann es sehr nützlich sein, einen ARMA-Prozess als reinen MA-Prozess darzustellen. Dann wird Darstellung (4.17) zu

$$X_t = \psi(B)Z_t. \quad (4.19)$$

Dabei ist $\psi(B) = \sum \psi_i B^i$ der sogenannte MA-Operator, eventuell von unendlicher Ordnung. Wenn wir (4.19) mit (4.17) vergleichen, sieht man sofort, dass

$$\psi(B) = \theta(B)/\phi(B)$$

gelten muss. Wenn man einen ARMA-Prozess als reinen AR-Prozess darstellt, so wird (4.17) zu

$$\pi(B)X_t = Z_t, \quad (4.20)$$

wobei $\pi(B) = \phi(B)/\theta(B)$. Nach Konvention schreibt man $\pi(B) = 1 - \sum_{i \geq 1} \pi_i B^i$, weil man natürlicherweise einen AR-Prozess in der Form

$$X_t = \sum_{i=1}^{\infty} \pi_i X_{t-i} + Z_t$$

schreibt. Wenn man jetzt (4.19) und (4.20), oder direkt die Definitionen von ψ und π , vergleicht, so muss gelten:

$$\pi(B)\psi(B) = 1.$$

Die Gewichte ψ oder π erhält man direkt durch Koeffizientenvergleich in Gleichungen wie $\psi(B)\phi(B) = \theta(B)$.

Dazu lösen wir jetzt zum Abschluss ein Beispiel: Gesucht sind die Gewichte ψ und π (von oben) für den ARMA[1,1]-Prozess

$$X_t = 0.5X_{t-1} + Z_t - 0.3Z_{t-1}.$$

Damit gilt hier: $\phi(B) = (1 - 0.5B)$ und $\theta(B) = (1 - 0.3)B$. Der Prozess ist damit stationär und invertierbar (siehe AR-, resp. MA-Prozesse). Wir berechnen

$$\psi(B) = \theta(B)/\phi(B) = (1 - 0.3B)(1 - 0.5B)^{-1} = (1 - 0.3B)(1 + 0.5B + 0.5B^2 + \dots).$$

Ausmultipliziert ist dies

$$\psi(B) = 1 + 0.2B + 0.1B^2 + 0.05B^3 + \dots$$

Das Bildungsgesetz liefert uns allgemein

$$\psi_i = 0.2 \times 0.5^{i-1}$$

für $i \in \{1, 2, \dots\}$. Analog erhält man $\pi_i = 0.2 \times 0.3^{i-1}$ für $i \in \{1, 2, \dots\}$.

4.4 Statistische Analyse im Zeitbereich

Nachdem wir in 4.3 die wichtigsten Modelle der Zeitreihenanalyse vorgestellt haben, werden wir jetzt in 4.4 die Methoden beschreiben, mit denen man ein Modell an eine gegebene Zeitreihe anpasst. Wichtigstes Hilfsmittel wird die Autokorrelationsfunktion sein. Die Analyse mit Hilfe der Autokorrelationsfunktion nennt man auch Analyse im Zeitbereich (time domain). Der Gegensatz ist die Spektralanalyse mit Hilfe des Spektrums; sie wird Analyse im Frequenzbereich genannt und aus Zeitgründen in dieser Vorlesung *nicht* behandelt.

4.4.1 Schätzung der Autokovarianz- und Autokorrelationsfunktion

Wir haben bereits in 4.2 die empirische Autokorrelationsfunktion angegeben. Dann folgte in 4.3 die theoretische Autokorrelationsfunktion, welche wir in MA-, AR- und ARMA-Prozessen explizit angegeben haben. Wenn wir eine empirische Zeitreihe haben, so wollen wir auf Grund der empirischen Autokorrelationsfunktion (mit unseren Kenntnissen der theoretischen Autokorrelationsfunktionen) ein geeignetes theoretisches Modell auswählen. Dieses Verfahren ist deshalb erfolgversprechend, weil die empirischen Autokorrelationsfunktionen für $N \rightarrow \infty$ im Modell gegen die richtigen Autokorrelationsfunktionen konvergieren. Das dies keine Selbstverständlichkeit ist, wird klar, wenn man sich vor Augen führt, dass man ja nicht - wie in der gewöhnlichen Statistik - mit dem Stichprobenumfang gegen unendlich geht (mit unabhängigen Realisationen!), sondern mit einer Realisation von immer grösserem Umfang arbeitet. Dies gilt jedoch nicht ohne Einschränkungen. Insbesondere brauchen wir stationäre Zeitreihen. Im Folgenden setzen wir voraus, dass die Zeitreihe stationär ist.

Wir rufen wieder in Erinnerung, wie wir aus einer empirischen Zeitreihe die empirischen Autokovarianzen in (4.7) berechnet haben:

$$c_k = \frac{1}{N} \sum_{t=1}^{N-k} (x_t - \bar{x})(x_{t+k} - \bar{x}). \quad (4.21)$$

Dies ist der gewöhnliche Schätzer unserer theoretischen Grösse $\gamma(k)$. Wir werden ihn hier auch nicht durch verbesserte Schätzer ersetzen, sondern beschränken uns darauf, die Eigenschaften dieses Schätzers zusammenzutragen. Die Beweise der folgenden Aussagen finden sich zum Beispiel in Schlittgen/Streitberg. Als erstes muss festgehalten werden, dass dieser Schätzer einen Bias der Ordnung $1/N$ aufweist (auch ein Schätzer c_k , in dem man durch $(N - K)$ anstatt durch N teilt, hat einen Bias). Daraus folgt jedoch sofort, dass der Schätzer immerhin asymptotisch erwartungstreu ist:

$$\lim_{N \rightarrow \infty} E[c_k] = \gamma(k).$$

Weiters kann gezeigt werden, dass gilt

$$\text{Cov}(c_k, c_m) \simeq \sum_{r=-\infty}^{\infty} [\gamma(r)\gamma(r+m-k) + \gamma(r+m)\gamma(r-k)]/N. \quad (4.22)$$

Mit $m = k$ erhalten wir die Varianz von c_k ; weil der Bias fast Null ist (von Ordnung $1/N$), entspricht dies auch in etwa dem Mean-Square-Error. (4.22) hat eine wichtige Konsequenz: Schätzungen von aufeinanderfolgenden Lags sind unter Umständen hochgradigst korreliert. Das heisst, dass wenn zum Beispiel c_1 zu gross ist, so wird auch c_2 eher zu gross ausfallen. Dies zeigt, wie schwierig und gefährlich die Interpretation eines empirischen Korrelogramms sein kann.

Wenn man die Autokovarianzfunktion geschätzt hat, wird man mit

$$r_k := c_k/c_0 \quad (4.23)$$

die Autokorrelationsfunktion schätzen. Weil dieser Schätzer ein Quotient zweier Zufallsgrössen ist, sind die statistischen Eigenschaften von r_k viel schwieriger zu untersuchen. Wir haben auch hier einen Bias, welcher gegen Null geht. Wir werden in dieser Vorlesung nur den Fall benutzen, wo wir eine völlig zufällige Zeitreihe haben (also eine gewöhnliche Stichprobe); dann haben wir nur bei Lag 0 eine Korrelation ungleich 0. Damit werden wir dann entscheiden können, ob geschätzte Werte einer Autokorrelation signifikant von 0 verschieden sind. Seien also x_1, \dots, x_N Realisationen von i.i.d. Zufallsgrössen; wir haben also gar keine echte Zeitreihe mehr. Man kann zeigen (zum Beispiel in Schlittgen/Streitberg), dass für $k \geq 1$ gilt:

$$E[r_k] \simeq -1/N$$

und

$$V[r_k] \simeq 1/N.$$

Unter sehr schwachen Bedingungen (stationäre ARMA-Prozesse erfüllen diese Bedingungen) ist r_k asymptotisch normalverteilt. Damit können wir einen einfachen Test auf White-Noise machen: wenn die H_0 -Hypothese eine völlig zufällige Zeitreihe bedeutet, so müssen 95% der r_k 's innerhalb des Bandes $-1/N \pm 2/\sqrt{N}$ liegen. Dies wird meist zu $\pm 2/\sqrt{N}$ vereinfacht. Wenn Werte von r_k ausserhalb dieses Bereichs liegen, so haben wir Kandidaten dafür, dass dort, bei k , die theoretische Autokorrelation vielleicht doch nicht 0 ist. Aber jedes 20. Mal gibt es einen Fehlalarm. Dies ist jedoch nur ein approximativer Test. Wenn N klein ist, muss man aber offensichtlich mit voreiligen Schlüssen vorsichtig sein. Wenn ein signifikanter Wert bei r_1 oder (bei saisonalen Daten) bei r_{12} auftritt, so ist grosse Vorsicht geboten. Dort kann eher "was los sein" als bei r_{59} . Ein Beispiel eines Korrelogramms von 100 i.i.d.-Realisationen findet man in Figur 4.15. Dort sind sowohl r_{12} wie auch r_{17} signifikant von 0 verschieden (obschon wir unabhängige Realisationen haben).

Falls wir auf Grund des Korrelogramms einen reinen White Noise Prozess (vorläufig) ausschliessen, werden wir uns bei den MA-, AR- und ARMA-Prozessen umschaun. Dabei haben wir bereits gelernt, dass wir es höchstwahrscheinlich

- a) mit einem MA-Prozess zu tun haben, wenn das Korrelogramm plötzlich (bei q) abbricht
- b) mit einem AR-Prozess zu tun haben, wenn das Korrelogramm eine Mischung von gedämpft-exponentiell und gedämpft-sinusförmig ist
- c) mit einem ARMA-Prozess zu tun haben, wenn das Korrelogramm langsam ausklingt.

Am besten erkennt man MA-Prozesse, während die AR- und echten ARMA-Prozesse schwieriger zu identifizieren sind. Wir werden in 4.4.2 die partiellen Autokorrelationen kennenlernen, welche uns als Entscheidungshilfen zusätzlich zur Verfügung stehen. Dieser erste Schritt "welche Klasse von Prozessen?" ist eindeutig der schwierigste und bedarf der Erfahrung.

4.4.2 Anpassung eines AR-Prozesses

Falls wir uns in obigem Prozess, nach Betrachtung des Korrelogramms, für einen Versuch mit einem AR-Prozess entscheiden, so stellen sich 2 Fragen:

- a) Was ist die Ordnung des Prozesses (das p)?
- b) Wie können wir die Parameter des Prozesses schätzen?

Wir wenden uns zuerst der Frage b) zu und nehmen also an, dass wir das "richtige" p bereits kennen.

4.4.2.1 Schätzung der Parameter eines AR[p]-Prozesses

Wir haben also einen AR[p]-Prozess mit unbekanntem Mittelwert μ und unbekanntem Parametern $\alpha_1, \dots, \alpha_p$, wobei gelte:

$$X_t - \mu = \alpha_1(X_{t-1} - \mu) + \dots + \alpha_p(X_{t-p} - \mu) + Z_t \quad (4.24)$$

Wenn wir N Beobachtungen x_1, \dots, x_N haben, so kann man zum Beispiel die $\mu, \alpha_1, \dots, \alpha_p$ durch die Methode der kleinsten Quadrate schätzen, indem man die folgende Summe bezüglich der $\mu, \alpha_1, \dots, \alpha_p$ minimiert:

$$S := \sum_{t=p+1}^N [x_t - \mu - \alpha_1(x_{t-1} - \mu) - \dots - \alpha_p(x_{t-p} - \mu)]^2.$$

Zu Recht erinnert diese Methode an die Schätzung der Parameter in einem linearen Regressionsmodell. Formal führt man ja die genau gleichen Rechnungen aus. Wenn (Z_t) ein Normalprozess ist, so ist diese Methode gleich der Maximum-Likelihood-Schätzung (wie beim linearen Modell wenn die Varianzen gleich sind), wenn die ersten p Werte x_1, \dots, x_p fix gegeben sind.

Wenn man in einem AR[1]-Prozess diese Gleichungen löst, erhält man

$$\hat{\mu} = \frac{\bar{x}_{(2)} - \hat{\alpha}_1 \bar{x}_{(1)}}{1 - \hat{\alpha}_1} \quad (4.25)$$

und

$$\hat{\alpha}_1 = \frac{\sum_{t=1}^{N-1} (x_t - \hat{\mu})(x_{t+1} - \hat{\mu})}{\sum_{t=1}^{N-1} (x_t - \hat{\mu})^2}, \quad (4.26)$$

dabei sind $\bar{x}_{(1)}, \bar{x}_{(2)}$ die Mittelwerte der ersten bzw. letzten $(N-1)$ Beobachtungen. Bekanntlich haben wir bei stationären Zeitreihen

$$\bar{x}_{(1)} \doteq \bar{x}_{(2)} \doteq \bar{x},$$

womit man meist setzt:

$$\hat{\mu} = \bar{x}. \quad (4.27)$$

Da dieser Schätzer einleuchtet, wird er meist anstelle von (4.25) eingesetzt. Zusätzlich setzt man diesen Schätzer auch in (4.26) ein und erhält:

$$\hat{\alpha}_1 = \frac{\sum_{t=1}^{N-1} (x_t - \bar{x})(x_{t+1} - \bar{x})}{\sum_{t=1}^{N-1} (x_t - \bar{x})^2}. \quad (4.28)$$

Wenn wir die Schätzung für α_1 aus (4.28) mit der Definition der empirischen Autokorrelation r_1 von (4.5) vergleichen, so sehen wir, dass der einzige (kleine) Unterschied im Nenner ist. Damit ist also

$$\hat{\alpha}_1 \simeq c_1/c_0 = r_1.$$

Dies ist nicht allzu überraschend: Die Autokorrelation eines AR[1]-Prozesses bei Lag 1 ist ja gerade auch α_1 .

Mit ähnlichen asymptotischen Überlegungen erhält man in einem AR[2]-Prozess folgende Schätzungen:

$$\hat{\mu} = \bar{x}$$

und jeweils

$$\hat{\alpha}_1 = r_1(1 - r_2)/(1 - r_1^2) \quad (4.29)$$

und

$$\hat{\alpha}_2 = (r_2 - r_1^2)/(1 - r_1^2). \quad (4.30)$$

Die Schätzer in (4.29) und (4.30) haben einen gewaltigen Vorteil: Wenn man einen AR[2]-Prozess anpassen möchte, obschon die Daten aus einem AR[1]-Prozess stammen, so ist ja $\rho(2) = \rho(1)^2$, also (so die Hoffnung) auch $r_2 \doteq r_1^2$. Damit werden $\hat{\alpha}_1 \doteq r_1$ und $\hat{\alpha}_2 \doteq 0$, was erwünscht ist.

4.4.2.2 Schätzung der Ordnung eines AR-Prozesses

Die Bestimmung der Ordnung eines AR-Prozesses aus dem Korrelogramm ist ein schwieriges Unterfangen. Bei einem AR[1]-Prozess haben wir ein geometrisches Abklingen der theoretischen Autokorrelationen. Wir haben berechnete Hoffnung, dass auch die empirischen Autokorrelationen ein vergleichbares Muster aufweisen. Aber bereits bei einem AR[2]-Prozess können schwierigere Muster auftreten; bei AR[p] mit $p \geq 3$ sowieso.

Ein Vorgehen von Praktikern besteht dann darin, nacheinander AR-Prozesse von immer grösserer Ordnung an die Daten anzupassen und jeweils die Fehlerquadratsumme zu berechnen. Diese kann man dann gegen die Ordnung plotten. Wenn man auf einmal ein deutliches Abfallen der Fehlerquadratsumme beobachtet, kann man die Ordnung an der Stelle wählen.

Ein anderes, theoretischeres Vorgehen führt über die sogenannte *partielle Autokorrelationsfunktion*. Diese hat im Gegensatz zur gewöhnlichen Autokorrelationsfunktion die Eigenschaft, dass sie bei einem AR[p]-Prozess bei p abbricht; bei einem MA[q]-Prozess hingegen langsam ausklingt (und nicht bei q abbricht)! Also bietet sie sich geradezu an, um AR-Prozesse zu identifizieren. Was ist unter dieser partiellen Autokorrelationsfunktion zu verstehen? Wir müssen dazu ausholen und erst erklären, was man unter partieller Korrelation versteht. Wenn man 3 Zufallsgrössen X, Y, Z hat, so kann es sein, dass Z sowohl X wie auch Y weitgehend erklärt: sei Z zum Beispiel die Anzahl Stunden Mathematik, welche Schüler pro Woche besuchen. X sei dann bei einem Test die Punktzahl in Algebra, Y die Punktzahl in Geometrie. Es ist natürlich klar, dass wenn jemand viele Stunden Mathematik besucht (grosses Z), dann wird er/sie auch viele Punkte in X und Y ablegen; umgekehrt mit wenigen Wochenstunden. Wenn man jetzt nur noch X und Y mittels der Korrelation miteinander vergleicht, so wird man feststellen, dass SchülerInnen mit guten Algebranoten auch gleichzeitig (meist) gute Geometrienoten haben und umgekehrt; entsprechend mit schlechten Noten. Man könnte dann meinen, X und Y seien positiv korreliert. Das stimmt auch! Aber gefühlsmässig möchte man doch die Korrelation zwischen X und Y messen, nachdem man den Z -Effekt *irgendwie wegdividiert* hat. Mathematisch exakt ausformuliert haben wir folgendes Vorgehen (Beweise dazu in Schlittgen/Streitberg): Wir definieren \hat{X} und \hat{Y} als Schätzungen von X resp. Y mit Hilfe von Z_1, \dots, Z_k derart, dass \hat{X} die beste lineare Approximation von X durch Z_1, \dots, Z_k ist: $\hat{X} := a_1 Z_1 + \dots + a_k Z_k$, sodass

$$E[(X - a_1 Z_1 - \dots - a_k Z_k)^2] \quad (4.31)$$

bezüglich a minimiert wird. \hat{Y} definiert man analog. Dann ist die partielle Korrelation von X und Y unter Konstanthalten von Z_1, \dots, Z_k per Definitionem der Ausdruck

$$\text{Corr}[X - \hat{X}, Y - \hat{Y}].$$

In der Tat hat man damit die Z -Einflüsse so gut wie möglich (im Sinne von (4.31)) wegdividiert und berechnet jetzt nur noch eine *Residualkorrelation*. Die partielle *Autokorrelation* wird nun folgendermassen definiert:

Definition 4.5 [partielle Autokorrelation] Sei X_t ein stationärer Prozess. Die partielle Autokorrelation π_k , $k \geq 2$, ist die partielle Korrelation von X_t und X_{t-k} unter Konstanthalten (s.o.) der dazwischen liegenden Zufallsgrößen X_u mit $t - k < u < t$. Dazu setzt man noch $\pi_0 = 1$ und $\pi_1 = \rho(1)$; des weiteren setzt man für $k < 0$ einfach $\pi_{-k} = \pi_k$.

Betrachten wir dazu als Beispiel einen AR[1]-Prozess. Die schematischen Abhängigkeitsverhältnisse werden an der Tafel entwickelt und zeigen sofort, dass die partielle Autokorrelation von beispielsweise X_t und X_{t-2} unter Konstanthalten von X_{t-1} gleich 0 sein muss. Aller Einfluss von X_{t-2} auf X_t ist ja bei einem AR[1]-Prozess bereits von X_{t-1} mitgegeben worden. Die selbe Überlegung gilt erst recht für grössere Lags. Also haben wir bei einem AR[1]-Prozess die folgenden partiellen Autokorrelationen: $\pi_0 = 1, \pi_1 = \rho(1), \pi_k = 0$ für $k \geq 2$. Mit derselben Überlegung folgt, dass die theoretischen partiellen Autokorrelationen π_k bei einem AR[p]-Prozess ab $k > p$ gleich 0 sein müssen. Dies lässt sich auch nachrechnen.

Zentral ist das folgende Theorem:

Theorem 4.6 [über die Berechnung der partiellen Autokorrelationen] Die partielle Autokorrelation π_p ist der Koeffizient $\alpha_p(p)$ in der besten linearen Anpassung (im Sinne von (4.31)) von X_t mit Hilfe von X_{t-1}, \dots, X_{t-p} , wobei $\hat{X}_t = \alpha_1(p)X_{t-1} + \dots + \alpha_p(p)X_{t-p}$.

Beweis von Theorem 4.6 Schlittgen/Streitberg.

Theorem 4.6 ist verwirrend, in Worten: Wenn man die partielle Autokorrelation an der Stelle p (oder einfach π mit Index p) berechnen möchte, so nimmt man die beste lineare Approximation von X_t durch die letzten p Folgenglieder (deshalb das $\alpha(p)$). Dann gilt, dass der Koeffizient des letzten Folgengliedes X_{t-p} , also $\alpha_p(p)$ genau π_p entspricht. Wenn man π_{p-1} berechnen möchte, so muss man mit weniger Folgengliedern, genauer mit genau $(p - 1)$, die beste Approximation machen und wieder den Koeffizienten des letzten Folgengliedes nehmen. In der Praxis berechnet man die partielle Autokorrelationsfunktion mit Hilfe der Levinson-Durbin-Rekursion (siehe Schlittgen/Streitberg).

Man beachte, dass wir bei der Einführung der partiellen Autokorrelation bis jetzt allgemeine stationäre Prozesse behandelt haben. Erst jetzt wenden wir uns wieder den AR-Prozessen zu.

Bemerkung 4.7: Es sei noch erwähnt, dass man in einem AR[p]-Prozess die beste lineare Approximation von X_t durch X_{t-1}, \dots, X_{t-p} mittels

$$\hat{X}_t = \alpha_1 X_{t-1} + \dots + \alpha_p X_{t-p}$$

erhält. Siehe dazu Schlittgen/Streitberg.

Für die AR[p]-Prozesse zentral ist nun

Theorem 4.8 [Partielle Autokorrelationsfunktion eines AR[p]-Prozesses] X_t ist genau dann ein AR[p]-Prozess, erfüllt also die Beziehung

$$X_t = \alpha_1 X_{t-1} + \dots + \alpha_p X_{t-p} + Z_t$$

mit $\alpha_p \neq 0$, wenn $\pi_p \neq 0$ und $\pi_k = 0$ für $k > p$.

Beweis von Theorem 4.8 Barndorff-Nielsen, O. and G. Schou (1973): On the parametrization of autoregressive models by partial autocorrelations; J. Multivariate Analysis, 3, 408-419, **199**.

Das für die weitere Arbeit zentrale Resultat ist also, dass die partiellen Autokorrelationen bei autoregressiven Prozessen der Ordnung p ab Lags grösser als p verschwinden. Es sei noch erwähnt, dass dies bei MA[q]-Prozessen nicht der Fall ist. In Aufgabe 28 auf Blatt 12 ist folgendes zu beweisen: die theoretische partielle Autokorrelation der Ordnung 2, π_2 , eines beliebigen, stationären Prozesses ist gleich

$$[\rho(2) - \rho(1)^2]/[1 - \rho(1)^2]. \quad (4.32)$$

In einem AR[1]-Prozess gilt $\rho(1)^2 = \rho(2)$; damit wird (4.32) null, was mit obiger Theorie konsistent ist. Wenn man (4.32) mit (4.30) vergleicht, so sieht man eine frappante Ähnlichkeit dieser beiden Ausdrücke. Dies bringt uns zum Punkt der Schätzung der partiellen Autokorrelationsfunktion. Aus Theorem 4.6 und Bemerkung 4.7 folgt, dass bei einem AR[p]-Prozess $\pi_p = \alpha_p$ gilt, falls der Prozess ein *echter* AR[p]-Prozess ist, das heisst $\alpha_p \neq 0$. Damit bietet sich für uns jetzt folgendes Vorgehen an: Wir wollen die π_k 's schätzen. Wir passen an die Daten zuerst einen AR[1]-Prozess mit den Methoden aus 4.4.2.1 an und setzen $\hat{\pi}_1 = \hat{\alpha}_1$. Dann passen wir an die Daten einen AR[2]-Prozess mit den gleichen Methoden an; wir setzen $\hat{\pi}_2 = \hat{\alpha}_2$. Dies können wir bis zu beliebiger Ordnung machen. Theoretisch (bis auf zufällige Fluktuationen) müssen jetzt bei einem AR[p]-Prozess wegen Theorems 4.8 von der Stelle $k = (p + 1)$ an alle $\hat{\pi}_k$ beinahe null sein. Den Entscheid, ob Fluktuationen für $k > p$ zufällig sind, wird wieder mit einem Band von $\pm 2/\sqrt{N}$ auf dem 5%-Niveau gemacht. Dieser letzte Punkt ist ein nichttriviales Resultat; es ist zwar identisch mit dem früheren Band aus 4.4.1, die Herleitung ist aber ungemein schwieriger und wird hier nicht ausgeführt.

4.4.3 Anpassung eines MA-Prozesses

Falls wir nach einer ersten Untersuchung eher für einen MA[q]-Prozess plädieren oder einen AR[p]-Prozess von "normaler" Ordnung einfach nicht anpassen konnten, werden wir es vielleicht mit einem MA[q]-Prozess versuchen. Auch hier müssen zwei Fragen beantwortet werden:

- a) Was ist die Ordnung des Prozesses (das q)?
- b) Wie können wir die Parameter des Prozesses schätzen?

Wieder wenden wir uns zuerst Frage b) zu, gehen also davon aus, dass wir das wahre q kennen.

4.4.3.1 Schätzung der Parameter eines MA[q]-Prozesses

Die Schätzung der Parameter eines MA[q]-Prozesses ist schwieriger als bei einem AR[p]-Prozess. Brauchbare geschlossene Formeln für die Parameterschätzung existieren nicht. Dies liegt unter anderem daran, dass man bei einem MA[q]-Prozess ja nur die X 's sieht, nicht aber die Z -Werte. Bei einem MA[q]-Prozess sind die Z -Werte in der Bestimmungsgleichung (4.9) aber zentral, während bei einem AR[p]-Prozess (siehe (4.10)) der einzige Z -Wert ein Störterm ist. Man muss sich bei MA[q]-Prozessen mit numerischen Iterationsverfahren begnügen.

Wir beginnen mit der Untersuchung eines MA[1]-Prozesses:

$$X_t = \mu + Z_t + \beta_1 Z_{t-1}. \quad (4.33)$$

Wir wollen kurz zwei Methoden antippen, welche sich aufdrängen aber leider nicht funktionieren:

1. Man drückt die Fehlerquadratsumme $\sum Z_t^2$ durch die beobachteten X 's und die Parameter μ und β_1 aus und minimiert diese Summe. Leider sind diese Rechnungen schwierig und führen nicht auf eine explizite, geschlossene Formel.
2. Wir wissen ja, dass bei einem MA[1]-Prozess die theoretische Autokorrelation gleich

$$\rho(1) = \beta_1 / (1 + \beta_1^2)$$

ist. Die empirische Autokorrelation r_1 können wir auch aus den Daten schätzen. Also könnten wir doch einfach

$$r_1 = \hat{\beta}_1 / (1 + \hat{\beta}_1^2) \quad (4.34)$$

setzen und nach $\hat{\beta}_1$ auflösen. Es gibt dann bekanntlich deren 2 Werte; man nimmt denjenigen mit $|\beta_1| < 1$. Leider ist dieser Schätzer sehr schlecht (ineffizient).

Box und Jenkins schlagen folgenden Weg vor: Man startet mit den sich trotzdem aufdrängenden Startwerten $\mu = \bar{x}$ und mit dem β_1 , welches man aus (4.34) schätzen könnte. Dazu setzt man noch $z_0 = 0$, wobei z_0 die

nicht direkt beobachtbare Realisation des White-Noise Prozesses Z an der Stelle 0 ist. Rekursiv berechnen wir jetzt die Residuen unter Verwendung von (4.33):

$$Z_t = X_t - \mu - \beta_1 Z_{t-1}. \quad (4.35)$$

Dies führt also zu $z_1 = x_1 - \mu$, da wir ja $z_0 = 0$ gesetzt haben. $z_2 = x_2 - \mu - \beta_1 z_1$, weiterfahrend erhalten wir am Schluss noch $z_N = x_N - \mu - \beta_1 z_{N-1}$. Hierin besteht die *Iteration!* Dann berechnet man $\sum_{t=1}^N z_t^2$. Wir können dann dieses Verfahren mehrmals anwenden und für verschiedene Wertepaare von (μ, β_1) die Restquadratsumme berechnen, indem man beispielsweise ein Gitter in der μ, β_1 -Ebene aufspannt. Dann kann man das Wertepaar auswählen, welches die Restquadratsumme minimiert. Es sei noch erwähnt, dass man der Kleinste-Quadrat-Schätzer auch der ML-Schätzer ist, gegeben $z_0 = 0$ und die Z_i 's seien normalverteilt.

Bei MA-Prozessen höherer Ordnung wendet man ein analoges Verfahren an wie oben: Beispielsweise kann man bei einem MA[2]-Prozess folgendermassen vorgehen: man startet mit (irgendwie) geschätzten Werten μ, β_1, β_2 und berechnet dann die Residuen wieder iterativ mit Hilfe der Gleichung

$$z_t = x_t - \mu - \beta_1 z_{t-1} - \beta_2 z_{t-2}.$$

Dann berechnet man auch hier die Fehlerquadratsumme. Dieses Verfahren wendet man auf ein (dreidimensionales) Gitter von Parametern μ, β_1, β_2 an. Dann sucht man den kleinsten Wert der Fehlerquadratsumme. Hier sind elektronische Rechner unerlässlich (dies gilt erst recht für noch grössere Ordnungen!). Des weiteren sollte man spezielle numerische Verfahren anwenden.

Wenn man auf einem Gitter solche Berechnungen vorgenommen hat, so ist es sinnvoll, die Fehlerquadratsumme graphisch darzustellen (geht bei MA[1]-Prozessen problemlos). Dabei sollte man auf zwei Punkte achten:

1. wie "flach" ist diese Oberfläche?
2. sind die Schätzungen für μ und β_1 stark korreliert?

4.4.3.2 Bestimmung der Ordnung eines MA[q]-Prozesses

Im Gegensatz zu den AR-Prozessen haben wir bereits mehrmals gesehen, dass wir wegen des Abbruchs bei Lag q für die MA-Prozesse ein hervorragendes Identifikationsmittel besitzen. Die partielle Autokorrelationsfunktion ist jedoch kaum eine weitere Hilfe bei MA[q]-Prozessen.

4.4.4 Anpassung eines ARMA-Prozesses

Wir haben in (4.18) gesehen, dass bei einem ARMA[1,1]-Prozess die Autokorrelationen ab Lag 1 geometrisch ausklingen. Falls wir beim empirischen Autokorrelogramm eines Datensatzes ein ähnliches Verhalten beobachten, können wir versuchen einen ARMA[1,1]-Prozess anzupassen.

Nach Besprechung der Probleme bei MA[q]-Prozessen ist es wohl nicht allzu überraschend, dass wir bei ARMA[p, q]-Prozessen auch iterative Verfahren anwenden müssen. Auch hier wird man auf einem Gitter die Restquadratsumme berechnen und diejenigen Parameter auswählen, welche den kleinsten Wert der Restquadratsumme hervorbringen. Da wir uns in dieser Vorlesung vor allem auf die ARMA[1,1]-Prozesse beschränken, werden wir dies nur für diesen Fall ausführen. Der allgemeine Fall wird jedoch nach den gleichen Prinzipien gelöst.

Die Bestimmungsgleichung für einen ARMA[1,1]-Prozess lautet

$$X_t - \mu = \alpha(X_{t-1} - \mu) + Z_t + \beta Z_{t-1}.$$

Gegeben seien also N Beobachtungen x_1, \dots, x_N ; wir suchen μ, α, β . Wir beginnen mit Startwerten, angenommen diese seien μ, α, β und setzen $z_0 = 0$ sowie $x_0 = \mu$. Dann berechnen sich die Residuen iterativ aus

$$z_t = x_t - \mu - \alpha(x_{t-1} - \mu) - \beta z_{t-1}$$

wie folgt: $z_1 = x_1 - \mu$, $z_2 = x_2 - \mu - \alpha(x_1 - \mu) - \beta z_1$ bis zur letzten Gleichung $z_N = x_N - \mu - \alpha(x_{N-1} - \mu) - \beta z_{N-1}$. Mit diesen Werten für die z 's berechnet man dann die Restquadratsumme. Danach nimmt man neue Werte für μ, α, β aus einem dreidimensionalen Gitter. Dies führt man dann in einem sinnvoll gewählten Bereich auf dem Gitter durch und sucht dann die Parameter, welche die Restquadratsumme minimieren. Mehr dazu findet man im Box und Jenkins.

4.4.5 Residuenanalyse

Nachdem man an die Daten ein Modell angepasst hat, will man untersuchen, ob dieses Modell die Daten genügend gut erklärt. Die statistische Methode dazu heisst *Residuenanalyse* und wird bei praktisch allen Modellierungen eingesetzt. Die Residuen sind definiert als

$$\text{Residuen} := \text{Daten} - \text{„gefittete“ Werte.}$$

Zum Beispiel gilt bei einem AR[1]-Prozess mit $\mu = 0$

$$\hat{z}_t = x_t - \hat{\alpha}x_{t-1}.$$

Bei einem "gut angepassten Modell" wird man erwarten, dass die so berechneten Residuen "zufälligen" Charakter haben und nahe bei Null liegen. Falls das Modell stimmt, hat man ja in Theorie einen White Noise Prozess mit unabhängigen Realisationen. Man wird die Residuen also in vielfältiger Weise plotten. Wir rufen jedoch in Erinnerung, dass wir es mit einer *Zeitreihe* zu tun haben und die Reihenfolge der Beobachtungen (und bei einer schlechten Modellwahl auch der Residuen) durchaus eine Rolle spielt. Damit macht es also Sinn, die Residuen ihrerseits vorläufig wieder als Zeitreihe aufzufassen.

Als erste Schritte wird man also wieder einen Plot der Daten - jetzt der Residuen - vornehmen und in einem zweiten Schritt sollte das empirische Korrelogramm berechnet werden. Der Plot sollte uns schon "Ausreisser" und offensichtliche Autokorrelation oder zyklische Effekte offenbaren. Benennen wir die Autokorrelationen der Residuen mit q_k für Lag k . Falls wir weiter oben das richtige Modell angepasst haben (weil wir die richtigen Werte von irgendwoher gekannt haben), so werden die Residuen eine Realisation eines White Noise Prozesses sein und das Korrelogramm, die q_k 's, sollten bei 5% Fehler erster Art in einem Band von $\pm 2/\sqrt{N}$ liegen. Aber wir haben ja aus den Daten zuerst die Parameter schätzen müssen. Dort kommt mehr Unsicherheit in's Spiel. Man kann zeigen, dass bei einem AR[1]-Prozess mit $\alpha = 0.7$ die Schranken für q_1 die Werte $\pm 1.3/\sqrt{N}$ haben, bei q_2 die Werte $\pm 1.7/\sqrt{N}$ und von $k \geq 3$ haben wir für q_k mit $\pm 2/\sqrt{N}$ wieder die alten Werte.

Bei einem ARMA-Prozess haben unter anderem Box und Jenkins bewiesen, dass $1/\sqrt{N}$ eine *obere Schranke* der Standardabweichung der q_k 's ist. Damit kann man folgern, dass Werte von q_k , welche ausserhalb des Bandes $\pm 2/\sqrt{N}$ liegen, sicher auf dem 5%-Niveau signifikant sind.

Viele weitere Tests sind entwickelt worden und auch in diversen Statistik-Paketen implementiert. Falls wir eine Modellwahl auf Grund der Residuenanalyse ablehnen müssen, wird man ein modifiziertes Modell - oder ein ganz neues Modell - ausprobieren müssen. Dieser Prozess kann sehr lange dauern und findet eventuell keinen befriedigenden Abschluss.

4.4.6 Abschliessende Bemerkungen zur Modellierung

Wir haben im Verlauf von 4.4 gesehen, dass das Modellieren ein schwieriger Prozess ist: Auf der einen Seite haben wir die Daten, auf der anderen Seite je nach Erfahrung und Vorwissen diverse Modelle, bei uns MA[q], AR[p] und ARMA[p, q]. Es geht allgemein darum, ein Modell für die Daten zu finden. Als seriöser Wissenschaftler oder Anwender sollte man nicht "die Daten foltern bis sie gestehen". Vielmehr kann man obigen Vorgang folgendermassen zusammenfassen:

Man untersucht die Art, wie die Daten gewonnen wurden (Zuverlässigkeit, Fehlerquellen, etc.)

Wieviele Daten habe ich zur Verfügung? (es sollten mehr als 50 sein)

Wozu betreiben wir eine Datenanalyse, Modellierung? Wozu wird das Modell gebraucht?

Plot der Daten, Korrelogramm

Modellvorschlag (als Resultat sagen wir dann zum Beispiel: MA[3], AR[1], ARMA[1,1])

Modellschätzung (die p , q , α_1, α_2)

Modellüberprüfung (Residuenanalyse)

Bücher über Zeitreihenanalyse für MathematikerInnen konzentrieren sich häufig auf die *Modellschätzung*. Da aber heute sehr gute Statistikpakete mit vollständig ausprogrammierten Schätzprogrammen existieren, kann man dies in einer anwendungsorientierten Einführung knapp bemessen. Der schwierigste Teil besteht sicher im Modellvorschlag. Dieser bedarf der Erfahrung mit realen Datensätzen. Der weiteren mündet obiges Prozedere oft in eine längere Schlaufe, wenn man bei der Modellüberprüfung einerseits feststellen muss, dass das Modell so nicht passt und andererseits (und hoffentlich) von der Residuenanalyse her auch gleich sich zu neuen Modellen inspirieren lässt.

4.5 Prognosemethoden

4.5.1 Einführende Bemerkungen

Im jetzt folgenden Teil werden wir uns mit der Vorhersage von zukünftigen Werten aus vergangenen Werten auseinandersetzen. In diesem Bereich gibt es naturgemäss viel Scharlatanerie mit fließenden Übergängen zu hochintelligenten Analysen, wobei wir uns jetzt nicht festlegen wollen, was wir als "hochintelligent" bezeichnen wollen. Davon zu unterscheiden ist das Denken in verschiedenen Szenarien, denen man nicht unbedingt eine Eintretens-Wahrscheinlichkeit zuordnet. Das Denken in verschiedenen Szenarien ist sicher sinnvoll, um die Tauglichkeit von Strategien/Sicherheitskonzepten zu untersuchen. Im folgenden Teil werden wir uns auf *mathematische* Prognose-Methoden konzentrieren, deren Tauglichkeit vor allem innerhalb von Modellen überprüft wird. Dies berechtigt zu methodischer Kritik!

Die Vorhersage von zukünftigen Werten aus vergangenen Werten einer Zeitreihe ist in den Wirtschaftswissenschaften, der Produktionsplanung, Finanzmathematik, Absatzplanung und vielen anderen Gebieten eine zentral wichtige Methode. Es geht also darum, aus Werten x_1, \dots, x_N einer Zeitreihe einen Wert x_{N+k} zu schätzen, wobei k der Prognosehorizont ist. Wir bezeichnen eine solche Prognose zum Zeitpunkt N mit Horizont k mit $\hat{x}(N, k)$.

Es gibt zwar sehr viele Prognosemethoden, aber leider muss gleich zu Beginn festgehalten werden, dass es keine universell anwendbare Prognosemethode gibt. In jedem Einzelfall muss neu entschieden werden, welche Methode wohl am meisten Erfolg verspricht. Vorhersagen basieren zudem immer auf Annahmen, welche sauber und transparent formuliert werden müssen. Diese Annahmen können sich als falsch erweisen. Ohne falschen Stolz muss man dann auch alle Prognosen wieder mit neuen Annahmen rechnen. Die alten Prognosen können sich also insbesondere als falsch erweisen. Die Endabnehmer (Volk, FinanzmarktteilnehmerInnen, etc.) sollten deshalb über die Unsicherheit bei Prognosen informiert werden; wenn man dies versäumt, verliert man schnell an Glaubwürdigkeit. Auch kann es sinnvoll sein, mit diversen Annahmen viele Szenarien durchzurechnen, was zusätzlich die Unsicherheit der MathematikerInnen/StatistikerInnen/OekonometrikerInnen unterstreicht und zu einer entkrampfteren Diskussion beitragen kann. Als Beispiel seien AHV-Prognosen zu nennen, bei denen man diverse Szenarien bzgl. BSP-Wachstum (-1%, -0.5%, 0%, 0.5%, 1%, 1.5%, 2%), Entwicklung der Lebenserwartung (+1 Jahr pro 10 Jahre), Zuwanderung (und was für Zuwanderung (Alter, Ausbildung)), Fruchtbarkeit und Finanzierungsmodus berechnet und als Diskussionsbasis den interessierten Kreisen (in der Schweiz dem Volk) vorlegt.

Vorhersagemethoden können grob in drei Kategorien unterteilt werden:

- a) Subjektiv Wir haben uns weiter oben bereits über Scharlatanerie kurz unterhalten. Es gibt einen fließenden Übergang vom Kaffeesatz hin zu subjektiven Vorhersagen basierend auf erfahrenem Urteilen, Intuition und Wirtschaftsnachrichten. Dieser leicht abschätzbare Unterton in obigen Formulierungen könnte darüber hinwegtäuschen, dass auch bei sogenannten "objektiven" mathematisch/statistischen Methoden bei der Modellwahl auch starke subjektive Momente einfließen, welche die Prognose stark beeinflussen können.
- b) Univariate Dabei basieren die Vorhersagen nur auf bisherigen Werten der gleichen Zeitreihe. Als Beispiel dienen Vorhersagen von Verkaufszahlen eines Produkts auf Grund der bisherigen Verkaufszahlen. Diese Methoden werden manchmal als naive Methoden oder Projektionsmethoden bezeichnet.
- c) Multivariate Dabei beruhen die Vorhersagen einer Zeitreihe mindestens zum Teil auch auf Werten von mindestens einer anderen Zeitreihe. Diese anderen Zeitreihen nennt man dann Prädiktoren oder erklärende Variablen. Beispielsweise kann man obige Verkaufsprognosen auch vom Wirtschafts- oder Lohnwachstum abhängig machen. Solche Methoden bezeichnet man auch als kausale Methoden.

In der Praxis werden oft, mit sehr gutem Erfolg, Kombinationen obiger Verfahren angewandt: einerseits wird man statistische Prognoseverfahren anwenden und Intuition und Kenntnisse von erfahrenen Anwendern des jeweiligen Gebiets einfließen lassen. Andererseits wird man eventuell ein gewichtetes Mittel der Prognosen von diversen Methoden als Prognose abgeben. Der Nachteil der letzten Methode ist, dass wir dann kein "sauberes", zu Grunde liegendes Modell mehr haben.

Eine alternative Art, die Prognosemethoden zu klassifizieren besteht darin, zu unterscheiden, ob die Prognoseverfahren automatisch ablaufen und keine menschliche Intervention nötig ist, oder ob subjektiver, menschlicher Input notwendig ist. Neben den subjektiven Methoden gehören insbesondere die multivariaten Methoden zur letzten Kategorie. Die meisten univariaten Methoden kann man automatisch ablaufen lassen; sie lassen aber auch mehr oder weniger viel menschliche Intervention zu.

Es gibt viele Punkte, welche man bei der Wahl der Methode berücksichtigen sollte; unter anderem:

- a) Wozu will man eine Vorhersage, wozu wird sie nachher gebraucht?
- b) Was für eine Zeitreihe haben wir: hat es einen Trend oder eine Saisonkomponente (in der ursprünglichen Zeitreihe)?
- c) Wie viele beobachtete, vergangene Werte stehen uns zur Verfügung?
- d) Wie weit ist der Prognosehorizont?
- e) Die Anzahl Zeitreihen, bei denen man Prognosen machen will und die Kosten von solchen Prognoseverfahren pro Reihe
- f) Die Fähigkeiten und die Erfahrung des Analysten sowie die zur Verfügung stehenden Statistikpakete und die dort vorprogrammierten Prozeduren. Der/die Analyst/in sollte sich mit der Methode wohl fühlen; zudem sollten mehrere Methoden ausprobiert werden.

Punktschätzungen sind zwar insofern von Vorteil, als dass sie nur eine Zahl als Ergebnis liefern. Wenn man aber zusätzlich zur Punktschätzung eine Intervallschätzung abgibt, so hat man auch eine Information über die mit der Prognose *modellinhärent* vorhandene Unsicherheit. Da man auch eine Modellwahlunsicherheit hat, liefert diese Intervallschätzung meist zu enge Intervalle.

4.5.2 Univariate Verfahren

4.5.2.1 Extrapolation von Trendkurven

Wenn man eine Prognose über einen sehr grossen Zeithorizont machen muss, ist es kaum sinnvoll, ein kompliziertes Modell aus der Zeitreihenanalyse anzupassen. Hingegen kann es sinnvoll sein, zumindest den

Trend zu prognostizieren. Dazu eignen sich Kurven, welche wir zum Teil in 4.2.7 kennengelernt haben (lineare, polynomiale, exponentielle, logistisch, Gompertz-Kurven, Splines und vieles mehr). Man passt dann diese Kurven an die bisherigen Daten nach einem Kriterium an (zum Beispiel "Kleinste Quadrate") und extrapoliert dann mit dieser Kurve die zukünftigen Werte. Die Daumenregeln dazu sind, dass man mindestens 7 historische Werte braucht und der Prognosehorizont nicht grösser als halb so gross wie der historische Zeithorizont (das N) sein darf. Der Nachteil dieser Methode ist der, dass es keine logische Begründung für die angepasste Kurve gibt, abgesehen vom auch willkürlichen "Kleinste-Quadrat-Kriterium" (KQK), mit dem man die Kurve auswählt. Leider kann man oft viele Trendmodelle an die Daten anpassen, welche fast gleich gute Anpassung an die historischen Daten sind (bzgl. KQK), aber völlig verschiedene Prognosen liefern.

4.5.2.2 Exponentielles Glätten

Das folgende Verfahren wurde 1958 von Holt vorgeschlagen und taugt in dieser einfachen Form nur für Zeitreihen ohne Trend und ohne Saisonkomponente. Sollte eine Zeitreihe zum Beispiel einen Trend aufweisen, wird man diesen zuerst entfernen und danach die folgende Prognosemethode anwenden. Danach muss man den Trend wieder hinzufügen.

Sei also x_1, \dots, x_N eine Zeitreihe ohne Trend und ohne Saisonkomponente. Wir werden jetzt den Wert von x_{N+1} mit Hilfe einer gewichteten Summe der bisherigen Werte prognostizieren:

$$\hat{x}(N, 1) = c_0 x_N + c_1 x_{N-1} + c_2 x_{N-2} + \dots \quad (4.36)$$

Die c_i sind dabei die Gewichte. Es erscheint gefühlsmässig harmonisch, nicht allzuweit zurückliegenden Werten mehr Gewicht zu geben als Werten, welche weit in der Vergangenheit liegen. Dazu eignen sich zum Beispiel geometrisch abklingende Gewichte, welche auf 1 summieren. In der Tat sind dies genau die Gewichte, welche man beim exponentiellen Glätten benutzt (der Name ist also eigentlich falsch; es müsste heissen *geometrisches* Glätten):

$$c_i = \alpha(1 - \alpha)^i$$

für $i = 0, 1, \dots$ und $0 < \alpha < 1$. Damit wird (4.36) zu

$$\hat{x}(N, 1) = \alpha x_N + \alpha(1 - \alpha)x_{N-1} + \alpha(1 - \alpha)^2 x_{N-2} + \dots \quad (4.37)$$

Nach dem bisher gesagten, vor allem damit die Gewichte auf 1 summieren, ist eigentlich eine unendlich lange Reihe von Werten aus der Vergangenheit von Nöten. Diese stehen natürlich in der Praxis nicht zur Verfügung. Man schreibt deshalb Gleichung (4.37) meist in rekursiver Form, welche auch andere Vorteile hat:

$$\hat{x}(N, 1) = \alpha x_N + (1 - \alpha)[\alpha x_{N-1} + \alpha(1 - \alpha)x_{N-2} + \dots] = \alpha x_N + (1 - \alpha)\hat{x}(N - 1, 1). \quad (4.38)$$

Wenn man jetzt noch $\hat{x}(1, 1) = x_1$ setzt, so kann man mit Hilfe von (4.38) fortlaufend Prognosen machen. (4.38) ist auch deshalb sehr praktisch, weil wir damit die neuen Prognosen machen können und dabei nur den zuletzt geschätzten Wert und den aktuellen benutzen und abspeichern müssen.

Die Wahl von α hängt von den Eigenschaften der unterliegenden Zeitreihe ab. Aus (4.37) ist sofort ersichtlich, dass Werte von α nahe bei 1 vor allem den letzten Wert starkt gewichten und die weiter zurückliegenden kaum berücksichtigt. Anders bei Werten von α nahe bei null, wo weit zurück liegende Werte die Schätzung immer noch stark beeinflussen. Häufig werden Werte aus dem Intervall $\alpha \in [0.1, 0.3]$ benutzt. Man kann auch versuchen, ein "gutes" α aus den bisherigen Daten zu schätzen: dazu sei ein Kandidat α gegeben. Wir setzen dann $\hat{x}(1, 1) = x_1$ und $e_2 = x_2 - \hat{x}(1, 1)$. e_i ist also der Schätzfehler. Man fährt fort mit

$$\hat{x}(i, 1) = \alpha x_i + (1 - \alpha)\hat{x}(i - 1, 1) = \alpha[x_i - \hat{x}(i - 1, 1)] + \hat{x}(i - 1, 1) = \alpha e_i + \hat{x}(i - 1, 1),$$

wobei allgemein $e_i = x_i - \hat{x}(i-1, 1)$ gilt. Damit erhalten wir zum Beispiel $\hat{x}(2, 1) = \alpha e_2 + \hat{x}(1, 1)$. So fortfahrend erhält man alle Schätzfehler und kann diese dann quadriert aufsummieren als Fehlerquadratsumme: $\sum_{i=2}^N e_i^2$. Dasjenige α , welches die kleinste Restquadratsumme ergibt, sollte man dann wählen. Wenn man dies graphisch darstellt, sieht man schnell, dass die Oberfläche sehr flach, die Wahl des α 's also nicht so kritisch ist. In Figur 4.16 sind 4 verschiedene Situationen mit α 's von jeweils 0.2 und 0.8 eingetragen. Das sogenannte Holt-Winters-Verfahren berücksichtigt direkt Trend- und Saisonkomponenten (mehr dazu in Chatfield).

4.5.2.3 Box-Jenkins-Verfahren

Wir werden hier nur eine kurze Einführung zum Box-Jenkins-Verfahren geben. Des weiteren ist zu erwähnen, dass in praktisch allen Lehrbüchern das Box-Jenkins-Verfahren für sogenannte SARIMA-Modelle eingeführt wird. Was ist ein SARIMA-Modell? In SARIMA kommt versteckt (und nicht ganz zufällig) ARMA vor. ARMA-Prozesse sind eine Kombination von AR und MA-Prozesse und beinhalten als Spezialfälle diese zentral wichtigen Bausteine der Zeitreihenanalyse. Als wir die theoretischen Modelle MA, AR und ARMA besprochen haben, setzten wir voraus, dass die Zeitreihe bereits stationär ist. In der Realität muss man aber häufig zuerst einen Trend entfernen. Dies kann mit einem Differenzenfilter geschehen (vgl. 4.2.7.3):

$$y_t = x_{t+1} - x_t =: \Delta x_{t+1}.$$

Eventuell muss man mehrmals die Differenzen bilden (vgl. Figur 4.8). Wenn man von einem Prozess X d mal die Differenzen nehmen muss und danach einen ARMA[p, q]-Prozess Y hat, so sagt man X sei ein ARIMA[p, d, q]-Prozess. Es gilt dann:

$$Y_t = \Delta^d X_{t+1} = (1 - B)^d X_{t+1}.$$

Der Name "I=Integriert" kommt daher, dass man die Y aufsummieren muss, um wieder X zu erhalten. Man kann einen Random Walk auch als ARIMA[0,1,0]-Prozess auffassen. Was ist nun ein SARIMA-Prozess? Hier kommt (wieder nicht ganz zufällig) ARIMA vor. S steht dabei für Saisonal und bedeutet genauer multiplikativer, saisonaler Einfluss. Wir werden nicht weiter auf diese Modelle eingehen und verweisen dabei auf Chatfield. Wichtig ist einfach zu wissen, dass es Erweiterungen sind, welche ARIMA-, und damit auch ARMA-Prozesse als Spezialfälle beinhalten.

Die wichtigsten Schritte, um ein Modell für das Box-Jenkins-Verfahren aufzustellen, sind:

- a) Modellvorschlag Man untersucht die Daten und wählt unter den ARIMA-Prozessen den Prozess aus, welchem man die besten Chancen gibt, die Daten gut zu erklären.
- b) Schätzung Schätzen der Parameter im oben gewählten Modell wie in 4.4 beschrieben.
- c) Residuenanalyse Überprüfen des Modells.
- d) Goto a) Falls notwendig, muss man andere ARIMA-Modelle ausprobieren, bis man einen guten Fit erreicht.

Diese Schritte sind weiter oben in diesem Skript bereits aufgelistet worden. Das Verdienst und das zentrale am Box-Jenkins-Verfahren ist, dass allen vier Schritten gleiches Gewicht gegeben wird. Entsprechend viel Raum nehmen dann diese Schritte in Lehrbüchern auch ein. Des weiteren haben Box und Jenkins in diesem Verfahren auch die Differenzbildung zentral einfließen lassen und damit die ARMA Prozesse zu den ARIMA-Prozessen (sogar SARIMA-Prozessen) erweitert. Deshalb nennt man ARIMA-Modelle manchmal auch Box-Jenkins-Modelle.

Der erste Schritt im Box-Jenkins-Verfahren ist das Differenzieren der Daten. Dies wird so oft wiederholt, bis man glaubt, eine stationäre Zeitreihe zu besitzen. Dazu analysiert man die jeweiligen Korrelogramme der differenzierten Reihen, bis das Korrelogramm rasch abklingt (vgl. Figur 4.11 und Aufgabe 25). Wir gehen davon aus, dass die Zeitreihe keine saisonalen Effekte beinhaltet.

Wir nennen die neue Zeitreihe (nach Differenzenbildung) w_t , wo $t \in \{1, \dots, N - d\}$, wenn wir d mal die Differenzen gebildet haben. Jetzt kann man mit den Methoden von 4.4 ein Modell an w_t anpassen. Danach wird man anhand der Residuen überprüfen, ob es noch weitere Struktur in den Daten gibt, oder ob die Residuen als Realisationen von zufälligen Prozessen aufgefasst werden können.

Wenn wir glauben, eine vernünftige Modellierung gefunden zu haben, wenden wir uns der Prognose von zukünftigen Daten zu (wir bezeichnen die Daten wegen der besseren Vergleichbarkeit mit früheren Teilen dieses Kapitels wieder mit x_1, \dots, x_N und das Modell mit X_t). Die Vorhersagen berücksichtigen die bisherigen Daten und die bisherigen Prognosefehler bei 1-Schritt-Prognosen. Man kann zeigen, dass die k -Schritt Vorhersage, welche den erwarteten, quadratischen Fehler minimiert, gerade der bedingte Erwartungswert von X_{N+k} gegeben alle Information zur Zeit N ist:

$$\hat{x}(N, k) = E[X_{N+k} | X_N, X_{N-1}, \dots]. \quad (4.39)$$

Es werden jetzt 3 Verfahren vorgestellt, mit denen dieser bedingte Erwartungswert in der Praxis berechnet werden kann. Dabei wird die Relation

$$E[Z_{N+k} | X_N, X_{N-1}, \dots] = 0 \quad (4.40)$$

einfließen.

4.5.2.3.1 Vorhersage mit Hilfe der Bestimmungsgleichung

Dabei geht man davon aus, dass die Bestimmungsgleichung (4.16) exakt bekannt ist. Man wird dann

1. zukünftige Werte von Z mit 0 ersetzen,
2. zukünftige Werte von X mit deren bedingtem Erwartungswert ersetzen,
3. bekannte Werte von X und Z mit den bekannten Werte ersetzen; siehe jedoch weiter unten (Z ist unbeobachtet!).

Betrachten wir dazu die Bestimmungsgleichung (4.16) eines ARMA-Prozesses

$$X_t = \alpha_1 X_{t-1} + \dots + \alpha_p X_{t-p} + Z_t + \beta_1 Z_{t-1} + \dots + \beta_q Z_{t-q}.$$

Dann werden wir mit Daten bis zum Zeitpunkt N beispielweise eine 1-Schritt-Schätzung in der folgenden Form machen:

$$\hat{x}(N, 1) = \alpha_1 x_N + \dots + \alpha_p x_{N-p+1} + \beta_1 z_N + \dots + \beta_q z_{N-q+1}.$$

4.5.2.3.2 Vorhersage mit Hilfe der Gewichte ψ

Bei der Einführung der ARMA-Prozesse in 4.3.4.4 haben wir in Gleichung (4.17) die Bestimmungsgleichung (4.16) von ARMA-Prozessen kompakt in der Form

$$\phi(B)X_t = \theta(B)Z_t$$

zusammengefasst. Wenn wir nun voraussetzen, dass die Stationaritäts- und Invertierbarkeitsbedingungen aus 4.3.4.4 erfüllt sind, so können wir mit $\psi(B) := \theta(B)/\phi(B)$ folgende MA-Darstellung wählen:

$$X_t = \psi(B)Z_t. \quad (4.41)$$

Damit können wir nicht nur Vorhersagen berechnen, sondern auch die Varianz des Vorhersagefehlers berechnen. (4.41) heisst ausgeschrieben unter anderem

$$X_{N+k} = Z_{N+k} + \psi_1 Z_{N+k-1} + \dots$$

Damit schätzen wir

$$\hat{x}(N, k) = \sum_{j=0}^{\infty} \psi_{k+j} z_{N-j},$$

wir setzen also zukünftige Werte von Z gleich 0. Damit ist der k -Schritt-Vorhersagefehler gleich

$$Z_{N+k} + \psi_1 Z_{N+k-1} + \dots + \psi_{k-1} Z_{N+1}$$

und die Varianz des k -Schritt-Vorhersagefehlers ist gleich

$$(1 + \psi_1^2 + \dots + \psi_{k-1}^2) \sigma_z^2.$$

4.5.2.3.3 Vorhersage mit Hilfe der Gewichte π

Mit den Vorbereitungen von 4.5.2.3.2 können wir anstelle einer MA-Darstellung auch eine AR-Darstellung wählen und erhalten mit $\pi(B) = \phi(B)/\theta(B)$

$$\pi(B)X_t = Z_t.$$

Es gilt damit

$$X_{N+k} = \pi_1 X_{N+k-1} + \dots + \pi_k X_N + \dots + Z_{N+k}.$$

Also wird man die Vorhersage mit

$$\hat{x}(N, k) = \pi_1 \hat{x}(N, k-1) + \pi_2 \hat{x}(N, k-2) + \dots + \pi_k X_N + \pi_{k+1} X_{N-1} + \dots$$

machen. Diese Vorhersage wird rekursiv aufgebaut, da die dazwischenliegenden Werte ja auch fehlen.

In der Praxis sind die Modelle jedoch nicht bekannt: wir müssen die Parameter, die ψ 's, π 's und auch die Z 's schätzen. Letztere wird man durch 1-Schritt-Vorhersagefehler schätzen. Allgemein werden die Varianzen der Vorhersagefehler davon nicht zu stark betroffen.

4.5.3 Weitere Methoden

Im Bereich der univariaten Methoden gibt es noch die

- * schrittweise Autoregression (Spezialfall des Box-Jenkins-Verfahrens, automatisch)
- * Verallgemeinertes exponentielles Glätten
- * Bayesische Vorhersage und Strukturierte Modellierung; beide Verfahren benutzen die sogenannte Kalman-Filter-Methode um die Parameter aufzudatieren

Wir besprechen dies nicht in dieser Vorlesung. Des weiteren gibt es noch die multivariaten Verfahren (multiple Regression, ökonometrische Modelle). Beides ist kurz in Chatfield beschrieben.

Nach dieser kurzen Einführung in Prognoseverfahren, sollte eines klargestellt werden. Wir haben uns ja als Optimalitätskriterium die Minimierung des mittleren quadratischen Fehlers gegeben. In der Praxis entspricht dies der Minimierung der Restquadratsumme. Worin besteht dann der Unterschied diverser Prognose-Methoden? Dazu zitieren wir Priestley: *"There is no such thing as a forecasting "method"; there is no such thing as an ARMA (or Box-Jenkins) forecasting "method". There is something called a "least-squares"-forecasting method, and this, in fact, provides the basis for virtually all theoretical studies. What the authors refer to as the "Box-Jenkins Method" is simply a recursive algorithm for calculating least-squares forecast, and, if the model for the series was known precisely, Box-Jenkin's "method" would lead to exactly the same answer as the Wiener "method", the Kolmogorov "method" or even the Kalman filter "method". All these*

"methods" are simply different computational procedures for calculating the same quantity, namely, the least-squares forecast of a future value from linear combinations of the past data."

4.6 ARCH und GARCH

4.6.1 Einführung

Mit diesem letzten Teil der Zeitreihenanalyse schaffen wir eine Verbindung von Kapitel 3 und 4. Es lohnt sich, kurz die für die weitere Theorie wichtigsten Teile (am Schluss von Kapitel 3) nochmals durchzulesen. Repetieren wir stichwortartig diese Punkte beim Einsatz der Formel von Black-Scholes in der Praxis:

Im Modell von Samuelson - in diesem Modell gilt die Formel von Black-Scholes - genügt der Aktienkurs folgender sDGL:

$$dS_t = S_t(\mu dt + \sigma dX_t). \quad (4.42)$$

Dabei sind μ und σ zwei Konstanten und X_t die BB. Wir haben in Theorem 3.32 gesehen, dass wir zu dieser sDGL eine eindeutige Lösung der Form

$$S_t = S_0 \exp((\mu - \sigma^2/2)t + \sigma X_t) \quad (4.43)$$

haben. S_t hat eine Lognormalverteilung; das heisst $\log(S_t)$ hat eine Normalverteilung (genauer eine $\mathcal{N}(\log(S_0) + (\mu - \sigma^2/2)t, \sigma^2 t)$ -Verteilung. Für eine europäische Call-Option mit Strike K , Restlaufzeit θ , momentanem Preis des Basiswerts von x , risikolosem Zinssatz von r und Volatilität σ definieren wir

$$d_1 := \frac{\log(x/K) + (r + \sigma^2/2)\theta}{\sigma\sqrt{\theta}}$$

und $d_2 := d_1 - \sigma\sqrt{\theta}$.

Die Formel von Black-Scholes

$$C_t = xN(d_1) - Ke^{-r\theta}N(d_2),$$

wobei $N(d) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^d e^{-u^2/2} du$.

Man sieht sofort, dass σ der einzige unbekannt Parameter ist. Man kann σ aus historischen Daten schätzen oder die Formel von Black-Scholes bezüglich σ invertieren und erhält damit die implizite Volatilität. Aus Formel (4.43) können wir zudem schliessen, dass die sogenannten "Logreturns" folgende Verteilung haben müssten:

$$\log(S_t/S_{t-1}) \sim \mathcal{N}(\mu - \sigma^2/2, \sigma^2).$$

Wir werden jetzt eine ganze Folge von solchen Logreturns betrachten: $Y_i := \log(S_i/S_{i-1})$, $i \in \{1, \dots, T\}$, wobei T ohne Einschränkung eine ganze, grössere Zahl sei. Man kann dann untersuchen, ob diese Logreturns $(Y_i)_{1 \leq i \leq T}$ tatsächlich i.i.d. sind und zwar aus einer Normalverteilung. Man ist sich einig, dass dies nicht so ist. In der Tat hat man folgende Eigenschaften (siehe auch Figur 4.17):

1. **Heavy tails:** Es hat viel zu starke Ausschläge nach oben und nach unten (Kurtosis). Die Langschwanzigkeit nimmt sogar zu, wenn man die Zeitintervalle verkürzt. Zudem hat man auch eine asymmetrische Verteilung (**Schiefe (skewed)**). Sei $(x_i)_{1 \leq i \leq N}$ eine Stichprobe aus einer beliebigen Verteilung. Man definiert dann als empirisches Mass der Schiefe den folgenden Ausdruck:

$$T^S := \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^3 / \left(\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2 \right)^{3/2},$$

und für die empirische Kurtosis ist es

$$T^K := \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^4 / \left(\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2 \right)^2.$$

Für eine Stichprobe aus einer Normalverteilung müssten für T^S Werte um Null resultieren; für $T^K - 3$ ebenfalls. Damit ist insbesondere gesagt, dass Langschwänzigkeit von Verteilungen und Schiefe immer in Bezug auf die Normalverteilung angegeben werden. Die Formel von Black-Scholes ist relativ robust bezüglich dieser falschen Annahmen.

2. **Komplizierte Abhängigkeitsstrukturen:** Auf Grund der Unabhängigkeit der Folgenglieder würde man auch Unkorreliertheit dieser Folgenglieder erwarten. Dies trifft in der Tat zu. Hingegen hat man Abhängigkeitsstrukturen in den Logreturns derart, dass die Folge der $(|Y_i|)$ und der (Y_i^2) nicht unkorreliert sind. Bei i.i.d. müsste dies gelten. Man beobachtet eigentliche Volatility-Clusters, also eine Anhäufung hoher Volatilität in den Logreturns.

Die Varianz σ ist also nicht nur nicht bekannt, sondern auch alles andere als konstant. Leider ist die Formel von Black-Scholes nicht robust in Bezug auf diese Unsicherheit, das heisst, dass eine nichtadäquate Schätzung von σ einen ziemlich falschen Optionspreis liefert. Man versucht dann einerseits bei den Prozessen, welche das Modell von Samuelson ausmachen ((4.42) und die Gleichung für das Bankkonto), zeitabhängige Parameter μ, r, σ einzufügen und sogar auch zufällige Entwicklung dieser Prozesse zuzulassen. Andererseits will man auch die Logreturns gut modellieren können. Dazu eignen sich die ARCH und GARCH-Prozesse.

4.6.2 ARCH

Die folgenden Ausführungen basieren auf einem Text von Neil Shephard (1996): "Statistical aspects of ARCH and stochastic volatility", erschienen bei Chapman and Hall, Buchtitel: "The Time Series Models in Econometrics, Finance and other fields". Herausgegeben von Cox, Hinkley und Barndorff-Nielsen.

Zur Motivation denke man immer an die Logreturns einer Aktie: $Y_t = \log(S_t/S_{t-1})$. Wir haben also Daten y_1, \dots, y_N zur Verfügung. Wir suchen einen Prozess Y_t , welcher

1. Langschwänzig ist,
2. Autokorrelationen von Lag ungleich Null sollen verschwinden,
3. In den Y_t^2 sollten Autokorrelationen vorkommen, und zwar derart, dass sich Volatilitätsclusters bilden können. Y_t^2 ist ja ein Mass für die absolute Grösse der Ausschläge. Wenn wir bei den kleinen Lags grosse, positive Autokorrelationen bei den Y_t^2 erreichen können, so werden grosse Ausschläge angehäuft vorkommen (Volatilitätsclusters).

ARCH steht für "AutoRegressive Conditional Heteroscedasticity". Der einfachste ARCH-Prozess ist der ARCH[1], wobei dann gilt:

$$Y_t = Z_t \sigma_t, \tag{4.44}$$

wobei (Z_t) eine Folge von unabhängigen, standardnormalverteilten Zufallsgrössen sei. σ_t sollte die Volatilität des unterliegenden Basiswerts sein. Wir fordern:

$$\sigma_t^2 = \alpha_0 + \alpha_1 Y_{t-1}^2. \tag{4.45}$$

Da die Varianz immer positiv sein muss, werden wir fordern müssen, dass $\alpha_1, \alpha_2 \geq 0$. Gegeben alle bisherigen Werte y_1, \dots, y_{N-1} , so hat $Y_N | y_1, \dots, y_{N-1}$ eine $\mathcal{N}(0, \sigma_t^2)$ -Verteilung. Wir wollen jetzt die Stationaritätsanforderungen sowie die Autokorrelationen von Y_t und Y_t^2 untersuchen. Aus (4.44) und (4.45) erhält man

$$Y_t^2 = \sigma_t^2 + (Y_t^2 - \sigma_t^2) = \alpha_0 + \alpha_1 Y_{t-1}^2 + v_t, \tag{4.46}$$

dabei ist $v_t := \sigma_t^2 (Z_t^2 - 1)$. Falls $\alpha_1 \in (0, 1)$, so gilt $E[Y_t^2] = \alpha_0 / (1 - \alpha_1)$. Weiter kann man zeigen, dass

$$E[Y_t^4] / (E[Y_t^2])^2 = 3(1 - \alpha_1^2) / (1 - 3\alpha_1^2)$$

sobald $3\alpha_1^2 < 1$. Dieser Ausdruck ist immer grösser als 3; bei einer Normalverteilung haben wir genau 3. Damit haben wir also Langschwanzigkeit der Verteilung von Y_t . Unter der Bedingung $3\alpha_1^2 < 1$ ist Y_t^2 kovarianzstationär (ohne Beweis), die Autokorrelationsfunktion von Y_t^2 ist gegeben durch (Beweis in der Vorlesung an der Tafel)

$$\rho(\tau) = \alpha_1^{|\tau|}.$$

Damit ergeben sich Volatilitätsclusters. Falls $\alpha_1 < 1$, so folgt Y_t einem White Noise Prozess (Beweis in der Vorlesung an der Tafel) und Y_t^2 ist ein (verallgemeinerter, $\alpha_0!$) autoregressiver Prozess. Hier bezeichnen wir einen Prozess bereits dann als White Noise, wenn die Folgenglieder nur unkorreliert sind (Unabhängigkeit nicht nötig). Die Namensgebung ARCH wird in der Vorlesung an der Tafel begründet.

Im Text von Shephard findet man auch **Schätzmethoden** für α_0, α_1 . Da wir ja mit (4.46) einen AR-Prozess haben, könnten im Prinzip die dort entwickelten Methoden (kleinste Fehlerquadratsumme) eingesetzt werden. Leider sind diese Methoden ineffizient. In der Tat arbeitet man mit Maximum-Likelihood-Methoden. Sie sind in Shephard beschrieben.

Weiters findet man auch **Tests** für $\alpha_1 = 0$ (keine Volatilitätsclusters) in Shephard.

Zentral wichtig sind die ARCH (und GARCH) Modelle zur Vorhersage von Volatilitäten. Damit will man aber eben nicht etwa Y vorhersagen sondern Y^2 , welches ein Mass für die Volatilität ist. Wegen (4.39) muss der beste Schätzer (bzgl. MSE) der folgende bedingte Erwartungswert sein (Berechnung wie $E[Y_t^2]$ eine Seite vorher):

$$E[Y_{T+t}^2 | \Omega_T] = \alpha_0(1 + \alpha_1 + \alpha_1^2 + \dots + \alpha_1^{t-1}) + \alpha_1^t Y_T^2.$$

Dabei bezeichne Ω_T die Information bis zur Zeit T und t sei der Prognosehorizont.

4.6.3 GARCH

Es gibt viele Verallgemeinerungen von ARCH. Die bekannteste und in der Finanzwelt wichtigste ist der GARCH-Prozess, insbesondere der GARCH[1,1]-Prozess. Das **G** kommt von Generalized. AR mit einem MA Term wird ja zu ARMA. Genauso erhält man aus ARCH mit einem "MA"-Teil den GARCH. Es überrascht deshalb auch nicht besonders, dass während ein ARCH im Quadrat ein AR ist, ein GARCH im Quadrat ein ARMA ist. Genauer: (4.44) bleibt gleich, aber an Stelle von (4.45) tritt nun bei einem GARCH[1,1]

$$\sigma_t^2 = \alpha_0 + \alpha_1 Y_{t-1}^2 + \beta_1 \sigma_{t-1}^2. \quad (4.47)$$

Anstelle von (4.46) erhalten wir mit den gleichen Umformungen

$$Y_t^2 = \alpha_0 + \alpha_1 Y_{t-1}^2 + \beta_1 \sigma_{t-1}^2 + v_t = \alpha_0 + (\alpha_1 + \beta_1) Y_{t-1}^2 + v_t - \beta_1 v_{t-1}. \quad (4.48)$$

Damit haben wir also (bis auf die Konstante) einen ARMA-Prozess in den Quadraten. Die Autokorrelationen von Y_t^2 sind denn auch diejenigen des ARMA's. Damit ist auch gleich gesagt, dass man mit GARCH[1,1]-Modellen ein sehr grosses Spektrum von Verhalten modellieren kann. In GARCH-Modellen sind die Y selber White Noise (nur unkorreliert, nicht unabhängig!), die Y^2 erlauben Volatilitätsclusters. Damit sind die GARCH-Prozesse also sehr leistungsfähig. Dies erklärt auch ihren Erfolg in der Praxis.

Weitere Modelle sind IGARCH (integrated), Log GARCH, Exponential GARCH und weitere. Sie sind ausführlicher in Shephard beschrieben.

4.7 Weiterführende Fragen

Wir haben den ganz grossen Block der Analyse im Frequenzbereich, multivariate Zeitreihen und die ausführliche Behandlung der nichtlinearen Modelle (ARCH und GARCH) aus Zeitgründen ausgelassen. Mehr dazu findet man zum Beispiel im Hamilton.

4.8 Zusammenfassung von Kapitel 4

Die Zeitreihenanalyse ist ein immer wichtigeres Gebiet der Finanzmathematik. Ein eigenes Kapitel dazu war damit gerechtfertigt. Dieses Kapitel folgte einem klassischen Aufbau der Zeitreihenanalyse im Zeitbereich: Beschreibung von Zeitreihen, Vorstellen der wichtigsten Modelle: MA, AR, ARMA, später auch ARIMA, ARCH und GARCH. Mit den theoretischen Kenntnissen von Modellen wandten wir uns der Modellierung zu; wo Daten und Modelle sich gegenüberstehen. Ein wichtiges Anwendungsgebiet der Zeitreihenanalyse sind die Prognoseverfahren. Mit dem letzten Teil über ARCH und GARCH ist der Zusammenhang zwischen Kapitel 3 (Black-Scholes in der Praxis) und der Zeitreihenanalyse aufgezeigt worden.