

Univariate Box/Jenkins-Modelle in der Zeitreihenanalyse

Thome, Helmut

Veröffentlichungsversion / Published Version
Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:
GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:

Thome, H. (1994). Univariate Box/Jenkins-Modelle in der Zeitreihenanalyse. *Historical Social Research*, 19(3), 5-77.
<https://doi.org/10.12759/hsr.19.1994.3.5-77>

Nutzungsbedingungen:

Dieser Text wird unter einer CC BY Lizenz (Namensnennung) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier:
<https://creativecommons.org/licenses/by/4.0/deed.de>

Terms of use:

This document is made available under a CC BY Licence (Attribution). For more information see:
<https://creativecommons.org/licenses/by/4.0>

Univariate Box/Jenkins-Modelle in der Zeitreihenanalyse

Helmut Thome *

Abstract: This is the second in a series of articles which introduces basic concepts and models of time series analysis. Whereas the first paper (HSR 3/1992) presented elementary descriptive concepts and traditional techniques of decomposing a time series into trend, season, and irregular fluctuations, this second paper offers an introduction into the Box-Jenkins approach to modelling univariate processes. The basic concept underlying this methodology is the idea to treat observed time series data as (generally non-independent) realizations of a "stochastic process". This concept is discussed (after some introductory remarks) in section 2. In order to actually model a stochastic process, a number of restrictive assumptions need to be made regarding the "stationarity" of the process and the nature of the dependencies linking the time ordered realizations. The latter set of assumptions leads to three types of basic models which are outlined in subsequent sections: the *autoregressive (AR)*, the *moving-average (MA)*, and the *mixed (ARMA)* model. These models are constructed in a three step procedure: the *identification* (based on empirical autocorrelation and partial autocorrelation functions) of the model (section 7), the *estimation* of the model parameters (section 8), and the evaluation or *diagnosis* of the model (section 9). Since the assumption of stationarity is often unrealistic, Box and Jenkins have extended the repertoire of models in a way to include certain types of non-stationary processes, the so-called *integrated* processes, for which they invented their *ARIMA* models (section 10). In a further extension, *seasonal components* may be incorporated thereby creating *SARIMA* models (section 11). - The methods and models presented in this paper remain within the confines of univariate analysis. Strategies for modelling possible relationships between two or more series (dynamic regression, transfer-function models) will be outlined in one of the forthcoming issues of *HSR*.

* Address all communications to Helmut Thome, Institut für Soziologie, Universität Halle-Wittenberg, 06099 Halle (Saale), e-mail: thome@soziologie.uni-halle.d400.de, Fax: (0345) 3883130.

In Heft 3/1992 dieser Zeitschrift erschien ein Artikel zur Komponentenzerlegung (Thome 1992a), mit dem eine umfassende Einführung in die Zeitreihenanalyse begonnen werden sollte. Später als erwartet erscheint nun der zweite Artikel dieser Serie, der in die Modellierungsstrategien einführt, die als "Box/Jenkins-Methode" bekannt geworden sind. Es werden vor allem die Basis-Modelle (ARMA-Modelle) vorgestellt, die in den beiden letzten Abschnitten erweitert werden durch die Berücksichtigung bestimmter Formen der Nicht-Stationarität (ARIMA-Modelle) und durch den Einbau saisonaler Komponenten (SARIMA-Modelle).

Die Basis-Modelle, die hier vorgestellt werden, sind für praktische Anwendungen in der Sozialforschung in der Regel viel zu simpel. Man kann sich die komplexen, praxis-relevanten Modelle und Methoden jedoch nicht aneignen, bevor man die Grundkonzepte geklärt hat. Sie erscheinen dem Anfänger häufig als unnötig abstrakt ohne erkennbaren Bezug zu seiner substantiellen Arbeit. Das erschwert die Rezeption dieser Methode; der Leser braucht eine gehörige Portion Geduld und Ausdauer. Ansonsten werden lediglich Grundkenntnisse der Inferenzstatistik und der Regressionsanalyse vorausgesetzt (s. Thome 1990).

1. Einleitung

Wir beginnen unsere Betrachtung mit einem Blick auf zwei Zeitreihen. Abbildung 1 zeigt die Entwicklung des realen Bruttoinlandprodukts, Abbildung 2 den Verlauf der Kriminalitätsrate in Schweden.

Die beiden Reihen können mit den Mitteln der deskriptiven Statistik beschrieben und in einzelne Komponenten zerlegt werden (s. Thome 1992a). Der Historiker oder Soziologe stellt sich aber häufig auch die Frage, ob zwischen zwei Zeitreihen ein struktureller Zusammenhang besteht. Er möchte z. B. prüfen, ob die Entwicklung der Kriminalitätsrate etwas zu tun hat mit der Entwicklung des Wohlstandes oder anderer Variablen. Er mag daran denken, eine derartige Beziehung mit Hilfe eines Regressionsmodells zu untersuchen. Eine solche Untersuchung stieße jedoch auf besondere Probleme (und Möglichkeiten), die in der "Natur" der Zeitreihen liegen, also von Daten, die durch ihren Zeitindex (den Meßzeitpunkt) geordnet sind. Für solche Daten kann man z. B. nicht annehmen, daß die "Fehler" (die "Residuen") unkorreliert sind, wie wir das in einer "normalen" Regressionsanalyse mit Querschnittsdaten voraussetzen. Es wird also nötig sein, die Abhängig-

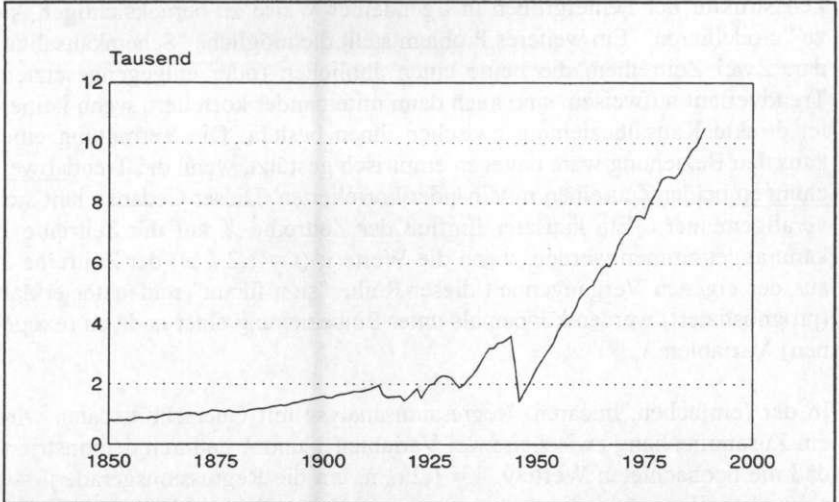


Abb. 1: Entwicklung des realen Bruttoinlandprodukts in Schweden 1850-1989

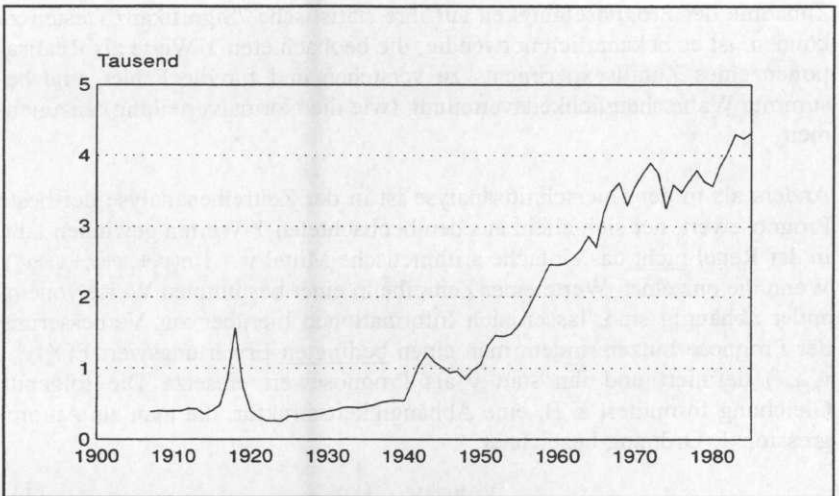


Abb. 2: Verlauf der Kriminalitätsrate in Schweden 1907-1985

keitsstruktur der Fehlergrößen in irgendeiner Weise zu berücksichtigen, sie zu "modellieren". Ein weiteres Problem stellt die mögliche "Scheinkausalität" dar. Zwei Zeitreihen, die beide einen ähnlichen (oder entgegengesetzten) Trendverlauf aufweisen, sind auch dann miteinander korreliert, wenn keinerlei direkte Kausalbeziehung zwischen ihnen besteht. Die Vermutung einer kausalen Beziehung wäre dagegen empirisch gestützt, wenn die Trendabweichungen beider Zeitreihen miteinander korrelierten. Dieser Gedanke läßt sich verallgemeinern: Ein kausaler Einfluß der Zeitreihe X_t auf die Zeitreihe Y_t kann angenommen werden, wenn die Werte y_t ($t = 1, 2, \dots, n$) der Zeitreihe Y_t aus der eigenen Vergangenheit dieser Reihe "signifikant" schlechter erklärt (prognostiziert) werden können als unter Einbeziehung einer anderen (exogenen) Variablen X_t .

In der (einfachen, linearen) Regressionsanalyse mit Querschnittsdaten wird ein Zusammenhang zwischen zwei Variablen Y und X dadurch demonstriert, daß die beobachteten Werte y_i , $i = 1, 2, \dots, n$, um die Regressionsgerade $\hat{y}_i = a + bx_i$ "signifikant" weniger streuen als um das arithmetische Mittel \bar{y} . Das heißt, die einzelnen y_i lassen sich mit Hilfe der entsprechenden Werte x_i besser prognostizieren als ohne deren Kenntnis. Man bezeichnet \hat{y}_i als den bedingten Erwartungswert der Zufallsvariablen Y : $\hat{y}_i = E(Y|x_i)$. Um die Zunahme der Prognosefähigkeit auf ihre statistische "Signifikanz" testen zu können, ist es bekanntlich notwendig, die beobachteten K -Werte als Realisationen eines Zufallsexperiments zu verstehen und für die Fehler eine bestimmte Wahrscheinlichkeitsverteilung (wie die Normalverteilung) anzunehmen.

Anders als in der Querschnittsanalyse ist in der Zeitreihenanalyse der beste Prognosewert, der sich allein aus den beobachteten Y -Werten gewinnen läßt, in der Regel nicht das einfache arithmetische Mittel $\bar{y} = 1/n(y_1 + \dots + y_t + \dots + y_n)$. Wenn die einzelnen Werte einer Zeitreihe in einer bestimmten Weise voneinander abhängig sind, lassen sich Informationen hierüber zur Verbesserung der Prognose nutzen, indem man einen bedingten Erwartungswert $E(Y_t | y_{t-1}, y_{t-2}, \dots)$ definiert und ihn statt \bar{y} als Prognosewert einsetzt. Die folgende Gleichung formuliert z. B. eine Abhängigkeitsstruktur, die man als Autoregression 1. Ordnung bezeichnet:

$$y_t = \phi y_{t-1} + a_t \quad (1)$$

Wir werden uns mit diesem Modell in späteren Abschnitten noch näher beschäftigen. An dieser Stelle genügen einige wenige Bemerkungen. Die Gleichung besagt, daß sich jeder Wert y_t aus zwei Komponenten zusammensetzt: dem unmittelbar vorangegangenen, mit einem bestimmten Gewicht ϕ versehenen Wert y_{t-1} und einer reinen Zufallskomponente a_t . Es wird also

kein "deterministischer" Zusammenhang zwischen aufeinanderfolgenden Beobachtungen angenommen, sondern ein zufallsgeprägter, ein "stochastischer" Zusammenhang. Wenn wir für die Zufallskomponenten einen Erwartungswert $E(a_t) = 0$ annehmen, ergibt sich in dem vorliegenden Beispiel für Y_t der Erwartungswert Φy_{t-1} - unter der Bedingung, daß y_{t-1} bereits vorliegt. Für den Zeitreihenanalytiker stellt sich also die Aufgabe, aus einer beobachteten Zeitreihe die interne Abhängigkeitsstruktur zu ermitteln, sie in einem Modell wie Gleichung (1) zu formalisieren und die entsprechenden Parameter zu schätzen. Er kann dieses Modell zur Prognose über den Beobachtungszeitraum hinaus benutzen oder als Basis, um den Einfluß exogener Variablen untersuchen zu können: Die Prognose müßte sich signifikant verbessern, wenn in den "Bedingungsset" für den Erwartungswert $E(Y_t)$ entsprechende Informationen über die Werte der exogenen Variablen aufgenommen würden. All dies setzt voraus, daß die beobachtete Reihe als Realisation eines stochastischen Prozesses betrachtet wird. Dieses Konzept ist nun etwas eingehender zu erörtern.

2. Zum Begriff stochastischer Prozesse

Wenn wir in der Regressionsanalyse mit Querschnittsdaten den strukturellen Zusammenhang zwischen zwei Variablen Y und X ermitteln, betrachten wir die abhängige Variable Y als eine Zufallsvariable, für die n (die Anzahl der Fälle) unabhängige Realisationen vorliegen. Auch wenn wir die interne Abhängigkeitsstruktur der einzelnen Zeitreihenwerte feststellen und sie als nicht-deterministisch ansehen wollen, müssen wir jeden einzelnen Beobachtungswert $y_t, t \in \{1, 2, \dots, n\}$ als Realisation einer Zufallsvariable Y_t auffassen, die grundsätzlich noch andere, nicht beobachtete Realisationsmöglichkeiten beinhaltet. Das ist in Gleichung (1) durch die stochastische Komponente a_t ausgedrückt. Zwei Fragen drängen sich an dieser Stelle auf: 1. Was kann man über die Wahrscheinlichkeitsverteilung einer Zufallsvariablen aussagen, für die zu einem bestimmten Zeitpunkt nur eine einzige Realisation vorliegt? 2. Wenn man eine Zeitreihe wie das Bruttosozialprodukt eines Landes über einen bestimmten Zeitraum beobachtet hat, warum soll man sich dann noch Gedanken darüber machen, daß jedes einzelne y_t auch anders hätte ausfallen können? Gibt es überhaupt eine "Grundgesamtheit", aus der eine "Stichprobe" gezogen worden ist oder eine andere sich hätte ziehen lassen? Mit der ersten Frage werden wir uns weiter unten relativ ausführlich beschäftigen. Die zweite Frage berührt Grundkonzepte der Statistik, die hier nur knapp angerissen werden können.

In Einführungskursen werden die Begriffe des Zufallsexperiments und der Zufallsvariablen oft allzu konkretistisch allein auf den Vorgang der Stich-

probenziehung aus einer empirischen Population bezogen. Hier liegen für jede Stichprobenvariable mehrere Realisationen vor, auf deren Basis man Rückschlüsse auf die entsprechende Verteilung in der Grundgesamtheit ziehen kann. Will man etwas über bestimmte Kennwerte ("Parameter") der Populationsverteilung aussagen, z. B. Konfidenzintervalle für das arithmetische Mittel schätzen, ist man jedoch genötigt, auf theoretische Modelle zurückzugreifen: Funktionen, die Stichprobenkennwerten (von denen man nur eine Realisation beobachtet hat) Wahrscheinlichkeiten zuordnen. Diese theoretischen Modelle sind gedankliche Konstruktionen über Häufigkeitsverteilungen, die man unter bestimmten Voraussetzungen erhielte, wenn man "unendlich oft" eine Stichprobenziehung wiederholen würde. In diesem Sinne stellen sie hypothetische (konzeptuelle) Populationen dar, und "Stichproben" sind "Ziehungen" aus einer Verteilungsfunktion $f(X)$. Bei "Signifikanztests" (oder anderen Formen des Überprüfens von Hypothesen) testen wir theoretische Annahmen (z. B. über den Zusammenhang zweier Variablen), indem wir sie mit beobachteten Daten vergleichen und dabei gewisse "zufallsbedingte" Abweichungen in Rechnung stellen. Die zu berücksichtigenden Zufallseinflüsse resultieren aber nicht nur aus der Stichprobenziehung, sondern auch aus Meßfehlern und anderen unbekanntem Einflußgrößen. Unsere theoretischen Annahmen müssen wiederum nicht auf bestimmte Populationen begrenzt sein, sondern können gesetzesartige Aussagen darstellen, die zeitlich und räumlich nicht (eng) begrenzt sind. Zum Beispiel wäre die Hypothese "Menschen neigen um so eher zum Selbstmord, je stärker sie ungewollt sozial isoliert sind" nichts wert, sie wäre zumindest unvollständig, wenn sie nur für Deutschland (oder irgendein anderes einzelnes Land) gelten sollte. Mit anderen Worten, die Hypothese formuliert ein hypothetisches Universum, in dem die Selbstmordfälle in einer bestimmten Weise auf sozial unterschiedlich stark isolierte Menschen verteilt sind. Ein einfaches Beispiel für Modelle in Gestalt von Wahrscheinlichkeitsfunktionen, das in den Anfangslektionen zur Inferenzstatistik häufig auftaucht, ist das Modell des Münzwürfens: Die theoretische Wahrscheinlichkeit ein "Wappen" zu werfen ist mit $P(W) = 0.5$ genauso groß wie die theoretische Wahrscheinlichkeit $P(Z) = 0.5$ für das Werfen der "Zahl". Das ist eine theoretische Annahme a priori. Bei 20 Würfeln "erwarten" wir demgemäß, daß das Ereignis "Wappen" ebenso zehnmals eintritt wie das Ereignis "Zahl". Auch hier betrachten wir das wiederholte Zufallsexperiment als eine Stichprobe von $n = 20$ Versuchen, obwohl wir eine empirische "Population" nicht angeben können. Die theoretische Verteilung $P(W) = 0.5$, $P(Z) = 0.5$ konstituiert eine hypothetische Population, eine Art Erzeugungsmechanismus, der bestimmte Ereignisse mit unterschiedlicher Häufigkeit hervorbringt. In der Regressionsanalyse formulieren wir einen Erzeugungsmechanismus für die Werte einer abhängigen Variablen Y , wobei wir eine deterministische Beziehung zwischen den bedingten Mittelwerten von Y und den Regressor-

variablen X_1, X_2, \dots, X_n annehmen und die Einflußgewichte der einzelnen Regressoren unter bestimmten Verteilungsannahmen für die stochastische Fehlerkomponente schätzen.

Auch in der Zeitreihenanalyse versucht man, einen Mechanismus zu identifizieren, der die beobachteten Werte generiert haben könnte, oder eine Hypothese hierüber zu testen - wie z.B. das Modell eines autoregressiven Prozesses 1. Ordnung. Wie bereits erwähnt, kann man nicht davon ausgehen, daß die Realisation der Zufallsvariable Y_t unabhängig ist von der Realisation der anderen Zufallsvariablen Y_{t-k} ($k=1,2,\dots$). Das bedeutet, es sind die Parameter (Kennwerte) einer multivariaten Verteilungsfunktion zu schätzen. Darauf bezieht sich der Begriff des stochastischen Prozesses: "A model which describes the probability structure of a sequence of observations is called a stochastic process" (Box u. Jenkins: 1976: 21). Schlingen u. Streitberg (1989, S. 70) weisen daraufhin, daß ein stochastischer Prozeß auf zweierlei Weise interpretiert werden kann: (1) als "Ensemble" von Zeitreihen, von denen eine durch den Zufallsvorgang ausgewählt wird, (2) "als Folge von Zufallsvariablen, wobei jedem Zeitpunkt t eine Zufallsvariable zugeordnet ist". Die Autoren fahren fort: "In den meisten zeitreihenanalytischen Anwendungen wird es schwierig sein, den zugrundeliegenden Zufallsvorgang substantiell zu spezifizieren. Ein stochastischer Prozeß dient dann als **Modellvorstellung**, in die bestimmte theoretische Annahmen über den Mechanismus eingehen, der die beobachtete Zeitreihe produziert hat" (ebd.).

Die Kennwerte, die bestimmte Eigenschaften einer Verteilungsfunktion darstellen, werden bekanntlich als "Momente" bezeichnet. Es sollen hier nur die beiden ersten Momente interessieren, also das arithmetische Mittel und die Varianzen/Kovarianzen. Für eine Zeitreihe Z_t , die aus n zeitlich geordneten Werten besteht, müßten folglich n Mittelwerte und n^2 Varianzen bzw. Kovarianzen der Z_t und Z_{t+k} geschätzt werden, die in der folgenden Matrix zusammengestellt sind:

$$\begin{bmatrix} \sigma_1^2 & \gamma_{1,2} & \gamma_{1,3} & \gamma_{1,4} & \dots & \gamma_{1,n} \\ \gamma_{2,1} & \sigma_2^2 & \gamma_{2,3} & \gamma_{2,4} & \dots & \gamma_{2,n} \\ \gamma_{3,1} & \gamma_{3,2} & \sigma_3^2 & \gamma_{3,4} & \dots & \gamma_{3,n} \\ \vdots & & \ddots & & & \vdots \\ \gamma_{n,1} & \gamma_{n,2} & \gamma_{n,3} & \gamma_{n,4} & \dots & \sigma_n^2 \end{bmatrix} \quad (2)$$

Die Varianzen sind mit σ_t^2 ($t = 1,2,\dots,n$), die Autokovarianzen mit $\gamma_{t,k}$ ($t,k = 1,2,\dots,n; k \neq t$) bezeichnet. Beide Kennwerte sind als deskriptive Maßzahlen

in dem ersten Artikel dieser Serie (Thome 1992a) schon definiert und erläutert worden. Als Schätzgröße ist die Autokovarianz wie folgt definiert:

$$\hat{\gamma}_k = \frac{1}{n} \sum_{t=1}^{n-k} (Z_t - \bar{Z}) (Z_{t+k} - \bar{Z}) \quad (3)$$

Daß man die Summe der Kreuzprodukte durch n und nicht durch $n-k$ dividiert, hat schätztechnische Gründe, die hier nicht zu erläutern sind. Wird $k=0$ gesetzt, wird aus der Autokovarianz die Varianz $\gamma_0 = \sigma^2$:

$$\hat{\gamma}_0 = \frac{1}{n} \sum_{t=1}^n (Z_t - \bar{Z})^2 \quad (4)$$

Soviele Parameter können natürlich mit nur n Beobachtungen gar nicht geschätzt werden. Folglich muß eine Reihe restriktiver Annahmen eingeführt werden, die die Zahl der zu schätzenden Parameter reduziert. Die erste Festlegung dieser Art ist, daß die arithmetischen Mittel $E(Z_t) = \mu$ für alle $t = 1, 2, \dots, n$ gleich sind, formal:

$$E(Z_1) = E(Z_2) = \dots = E(Z_n) = \mu \quad (5)$$

In gleicher Weise sollen die Varianzen konstant sein:

$$E(Z_t - E(Z_t))^2 = \sigma^2, \quad t = 1, 2, \dots, n \quad (6)$$

Darüber hinaus soll die Größe der Kovarianzen nur von der absoluten Größe des Verschiebeparameters (*Lag*) $|k|$ in Gleichung (3) abhängen. Das bedeutet z. B., daß die Kovarianzen zwischen z_t und z_{t+k} genauso groß sind wie die zwischen z_{t+1} und z_{t+1+k} , oder zwischen z_{t-1} und z_{t-1+k} ; die Autokovarianzen sind also symmetrisch.

Aus der Matrix (2) wird somit die Matrix (7), in der wir das Symbol σ^2 durch γ_0 ersetzen, da die Varianz formal nichts anderes ist als die Autokovarianz mit *Lag* $k=0$.

$$\begin{bmatrix} \gamma_0 & \gamma_1 & \gamma_2 & \dots & \gamma_{n-1} \\ \gamma_{-1} & \gamma_0 & \gamma_1 & \dots & \gamma_{n-2} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \gamma_{-(n-1)} & \gamma_{-(n-2)} & \gamma_{-(n-3)} & \dots & \gamma_0 \end{bmatrix} \quad (7)$$

Die Indices geben jetzt nur noch den Verschiebeparameter k an, der positiv oder negativ sein kann. Wegen der schon festgestellten Symmetrie der Autokovarianzen sind die Einträge unterhalb der Diagonalen spiegelbildlich identisch mit denen oberhalb der Diagonalen. Die Zahl der zu schätzenden Parameter hat sich also auf $n + 1$ reduziert: Mittelwert, Varianz und $n - 1$ Kovarianzen. Das sind immer noch zu viele, so daß weitere Restriktionen einzuführen sind, die zweierlei beinhalten können:

(a) Autokovarianzen, die über einen bestimmten *Lag* hinausgehen, werden gleich Null gesetzt. Beispielsweise kann man annehmen, daß Zeitreihenwerte, die 6 oder mehr Meßzeitpunkte voneinander entfernt liegen, voneinander unabhängig sind.

(b) Autokovarianzen "höherer Ordnung", also mit größerem *Lag*, lassen sich analytisch aus Autokovarianzen niedrigerer Ordnung ableiten.

Die erste Gruppe von Restriktionen wird in den *moving average* (MA) Modellen, die zweite Gruppe in den autoregressiven (AR) Modellen formuliert, die im nächsten Abschnitt vorgestellt werden. Die genaue Form dieser Modelle wird nicht willkürlich oder allein auf Grund theoretischer Überlegungen gewählt, sondern anhand der vorliegenden Daten identifiziert.

Die MA- und AR-Modelle lassen sich nur dann adäquat identifizieren und schätzen, wenn die zuvor eingeführten Voraussetzungen konstanter Erwartungswerte und konstanter Varianzen/Kovarianzen erfüllt sind. Zeitreihen, die diese Voraussetzung erfüllen, bezeichnet man als "schwach stationär" oder "kovarianz-stationär". Zeitreihen, die nicht nur in den beiden ersten, sondern in allen Verteilungsmomenten konstant sind, bezeichnet man als (stark) stationär. Die statistische Theorie stochastischer Prozesse fußt nicht nur auf den Stationaritätsannahmen, sondern setzt auch darüber hinaus die sog. "Ergodizität" voraus. Verkürzt ausgedrückt bedeutet dies, daß sich die Stichprobenmomente beobachteter Zeitreihen (d. h. zeitlich begrenzter Realisationen eines stochastischen Prozesses) mit zunehmender Länge (mit $n \rightarrow \infty$) den entsprechenden Populationsmomenten annähern. Diese Voraussetzung ist mit einer endlichen Reihe nie zu prüfen, und soll uns deshalb

auch nicht weiter beschäftigen. Man betrachtet sie in der Forschungspraxis als erfüllt, wenn die beobachtete Zeitreihe stationär ist. Viele sozialwissenschaftlich relevante Zeitreihen sind weder mittelwert- noch kovarianzstationär. Mit diesem Problem werden wir uns erst in Abschnitt 10 beschäftigen, so daß wir uns nun den schon erwähnten autoregressiven Modellen (Prozessen) zuwenden können.

3. Autoregressive Prozesse: AR-Modelle

Wir beginnen mit dem Modell eines autoregressiven Prozesses 1. Ordnung, in dem lediglich unmittelbar benachbarte Werte direkt voneinander abhängen:

$$z_t = c + \phi z_{t-1} + a_t \quad (8)$$

Im Unterschied zu Gleichung (1) haben wir hier noch eine Konstante "c" eingeführt, die dem Ordinatenabschnitt in der Regressionsgleichung für Querschnittsdaten entspricht. Die stochastische Komponente a_t unterliegt Voraussetzungen, die ebenfalls denen entsprechen, die wir aus der statischen Regressionsanalyse kennen:

$$\begin{aligned} E(a_t) &= 0 \\ E(a_t^2) &= \sigma_a^2 \\ \text{Cov}(a_t, a_{t+k}) &= 0 \text{ für } k \neq 0 \end{aligned} \quad (9)$$

Die Residuen sollen sich im Mittel ausgleichen, eine konstante Varianz aufweisen und weder untereinander noch mit der Regressorvariablen Z_{t-1} korrelieren. Zeitreihen (ganz gleich, ob es sich dabei um Residuen oder originäre Beobachtungswerte handelt), die diese Voraussetzungen erfüllen, bezeichnet man als "weißes Rauschen", *white noise*.

Im Unterschied zur üblichen Regressionsanalyse haben wir es in der univariaten Zeitreihenanalyse aber nicht mit zwei substantiell unterschiedlichen Variablen zu tun. Dennoch läßt sich eine Regression durchführen, wenn wir die beobachtete Zeitreihe zweifach hinschreiben und dabei um ein Beobachtungsintervall verschieben (s. Abb. 3). Aus der Zeitreihe $\{z_1 = 3, z_2 = 5, z_3 = 4, z_4 = 6, z_5 = 3\}$ ergeben sich so die Datenpaare $P_1 = (5;3)$, $P_2 = (4;5)$, $P_3 = (6;4)$ und $P_4 = (3;6)$, wobei die ersten Zahlen Werte der Variablen Z_t und die zweiten Zahlen Werte der Variablen Z_{t-1} darstellen. Daraus läßt sich das folgende Streudiagramm konstruieren (s. Abb. 4), das eine negative Bezie-

hung und somit einen negativen Phi-Koeffizienten in Gleichung (8) nahelegt.

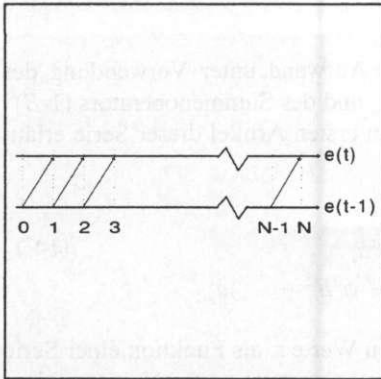


Abb. 3: Verschobene Zeitreihen

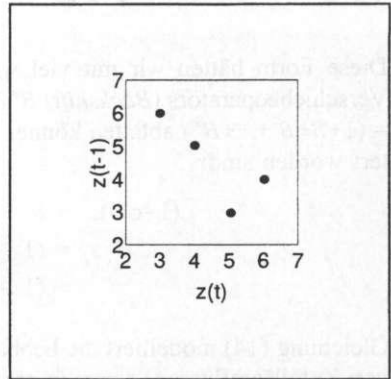


Abb. 4: Streudiagramm zwischen z_t und z_{t-1}

Wenn Gleichung (8) gilt, gelten auch die folgenden Gleichungen:

$$\begin{aligned} z_{t-1} &= c + \phi z_{t-2} + a_{t-1} \\ z_{t-1} &= c + \phi z_{t-2} + a_{t-1} \\ z_{t-2} &= c + \phi z_{t-3} + a_{t-2} \\ z_{t-3} &= c + \phi z_{t-4} + a_{t-3} \end{aligned} \tag{10}$$

$$\begin{aligned} z_t &= c + \phi(c + \phi z_{t-2} + a_{t-1}) + a_t \\ &= c + \phi c + \phi^2 z_{t-2} + \phi a_{t-1} + a_t \end{aligned} \tag{11}$$

Somit lautet die Gleichung (8) nun:

$$\begin{aligned} z_t &= c + \phi(c + \phi z_{t-2} + a_{t-1}) + a_t \\ &= c + \phi c + \phi^2 z_{t-2} + \phi a_{t-1} + a_t \end{aligned} \tag{11}$$

$$\begin{aligned} z_t &= c(1 + \phi + \phi^2 + \dots + \phi^{k-1}) + \phi^k z_{t-k} + \phi^{k-1} a_{t-(k-1)} + \dots \\ &\quad \dots + \phi^2 a_{t-2} + \phi a_{t-1} + a_t \end{aligned} \tag{12}$$

Unter der Voraussetzung $|\phi| < 1$ und $k \rightarrow \infty$ läßt sich der erste Teil der

$$c(1 + \phi + \phi^2 + \dots) = \frac{c}{1 - \phi} \tag{13}$$

Unter der gleichen Voraussetzung geht der Summand $\phi^k z_{t-k}$ gegen Null, so daß Gleichung (8) schließlich folgende Form annimmt:

$$z_t = \frac{c}{1 - \phi} + a_t + \phi a_{t-1} + \phi^2 a_{t-2} + \dots \quad (14)$$

Diese Form hätten wir mit viel weniger Aufwand unter Verwendung des Verschiebeoperators (*Backshift*) $B^m Z_t = Z_{t-m}$ und des Summenoperators $(1-B)^{-1} = (1+B+B^2+\dots+B^\infty)$ ableiten können, die im ersten Artikel dieser Serie erläutert worden sind:

$$\begin{aligned} (1-\phi B)z_t &= a_t \\ z_t &= (1 - \phi B)^{-1} a_t \\ &= (1 + \phi B + \phi^2 B^2 + \dots) a_t \end{aligned} \quad (14^*)$$

Gleichung (14) modelliert die beobachteten Werte z_t als Funktion einer Serie von Zufallseinflüssen, denen je nachdem, wie weit sie in die Vergangenheit zurückreichen, unterschiedliche Gewichte zugeordnet sind, die in diesem Falle ein geometrisch abfallendes Muster bilden. Alle nicht-deterministischen, schwach stationären stochastischen Prozesse, lassen sich als Linearkombination einer Folge nicht-korrelierter Zufallsvariablen im Sinne des *white noise* darstellen (Wold-Theorem). Sie unterscheiden sich nur in den Gewichten, die den verschiedenen a_{t-k} ($k=1,2,\dots,\infty$) zugeordnet sind. Gelegentlich interpretiert man das Muster dieser Gewichte als "Gedächtnisspur" des Systems. Im Falle stationärer autoregressiver Prozesse hat das System ein "unendlich langes" Gedächtnis, wobei die Einflußgewichte aber rasch gegen Null konvergieren, auch dann, wenn es sich um autoregressive Prozesse höherer Ordnung (s. unten) handelt. Da die univariaten Box/Jenkins-Modelle (von denen wir bisher in Gestalt des AR(1)-Modells nur eine spezifische Ausformung kennengelernt haben) Zeitreihenwerte stets als Funktion reiner Zufallseinflüsse darstellen, hat man sie polemisch auch "models of ignorance" genannt. Es sei jedoch schon an dieser Stelle darauf hingewiesen, daß diese Modelle um exogene Variablen, also substantiell interpretierbare Einflußgrößen erweitert werden können, was in diesem Artikel aber noch nicht erläutert wird.

Gleichung (14) erlaubt die rasche Ableitung des Erwartungswerts¹ des Prozesses, wenn man sich erinnert, daß laut Modellvoraussetzung $E(a_{t,k}) = 0$

1 Die Regeln des Rechnens mit Erwartungswerten sind knapp dargestellt in Thome (1990, S. 264 ff.)

für alle $k \geq 0$:

$$E(z_t) = E \left(\frac{c}{1 - \phi} + a_t + \phi a_{t-1} + \phi^2 a_{t-2} + \dots \right) \quad (15)$$

$$= \frac{c}{1 - \phi}$$

Der (nicht-bedingte) Erwartungswert ist also eine Konstante (vermöge der Restriktion $\phi < |1|$). Spätestens bei der Prognose (die ebenfalls erst in einem späteren Artikel behandelt wird) benötigen wir aber auch den bedingten Erwartungswert, den wir oben schon eingeführt hatten: $E(Z_t | z_{t-1}) = \phi z_{t-1}$.

Rechnerisch etwas komplizierter gestaltet sich die Ableitung der Varianz/-Kovarianz:

$$E(z_t - E(z_t))^2 \quad (16)$$

$$= E \left(\frac{c}{1 - \phi} + a_t + \phi a_{t-1} + \phi^2 a_{t-2} + \dots - \frac{c}{1 - \phi} \right)^2$$

$$= E \left[(a_t + \phi a_{t-1} + \phi^2 a_{t-2} + \dots) (a_t + \phi a_{t-1} + \phi^2 a_{t-2} + \dots) \right]$$

$$= E \left(\sum_{i=0}^{\infty} \phi^i a_{t-i} \right) + E \left(\sum_{i=0}^{\infty} \phi \phi^i a_{t-1-i} a_{t-i} \right) + E \left(\sum_{i=0}^{\infty} \phi^2 \phi^i a_{t-2-i} a_{t-i} \right)$$

Da laut Voraussetzung die $\{a_i\}$ voneinander unabhängig sind, gilt für alle Erwartungswerte $E(a_i a_j) = 0$, wenn $k \neq i$. Die meisten Terme der letzten Gleichungsseite können also gleich Null gesetzt werden und es bleibt

$$E(z_t - E(z_t))^2 = E(a_t^2) + \phi^2 E(a_{t-1}^2) + \phi^4 E(a_{t-2}^2) + \dots$$

$$= \sigma_a^2 (1 + \phi^2 + \phi^4 + \phi^6 + \dots) \quad (17)$$

$$\gamma_0 = \frac{\sigma_a^2}{1 - \phi^2}$$

Die bedingte Varianz ist mit

$$\begin{aligned}
 V(z_t | z_{t-1}) &= E[(z_t - E(z_t | z_{t-1}))^2] \\
 &= E[(z_t - \phi z_{t-1})^2] \\
 &= E(a_t^2) \\
 &= \sigma_a^2
 \end{aligned}
 \tag{18}$$

gegeben. Im Unterschied zum bedingten Mittelwert ist also auch die bedingte (ebenso wie die nicht bedingte) Varianz eines schwach stationären Prozesses eine Konstante.

Die Autokovarianzen lassen sich mathematisch in ähnlicher Weise ableiten wie die Varianzen, was schließlich zur Formel

$$\gamma_k = \frac{\phi^k \sigma_a^2}{1 - \phi^2} = \phi^k \gamma_0
 \tag{19}$$

führt. Wenn $k = 0$, ist diese Gleichung identisch mit der Varianzgleichung (17).

Damit lassen sich auch entsprechende Formeln für den Korrelationskoeffizienten unmittelbar angeben. Bekanntlich ist der Pearsonsche Korrelationskoeffizient als Quotient aus der Kovarianz und den beiden Standardabweichungen definiert, bezogen auf die Variablen Z_t und Z_{t-k} :

$$\rho_k = \frac{\text{Kov}(z_t, z_{t-k})}{\sigma_{z_t} \sigma_{z_{t-k}}}
 \tag{20}$$

Unter Berücksichtigung von Gleichung (19) ergibt sich somit

$$\rho_k = \frac{\phi^k \gamma_0}{\gamma_0} = \phi^k
 \tag{21}$$

Bei einem AR(1)-Prozeß ist das Muster der theoretischen Autokorrelationskoeffizienten identisch mit dem Muster der Gewichte für die Zufallseinflüsse in Gleichung (14). Bei einem negativen ϕ oszillieren die Korrelationskoeffizienten zwischen positiven und negativen Werten. Die einzelnen Autokorrelationskoeffizienten lassen sich als Funktion des Verschiebeparameters k

ausdrücken und dementsprechend in einem Diagramm darstellen (s. Abb. 5a und 5b)

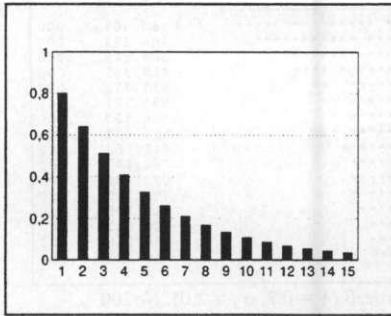


Abb. 5a: AKF des AR(1)-Prozesses: $\phi=0.8$

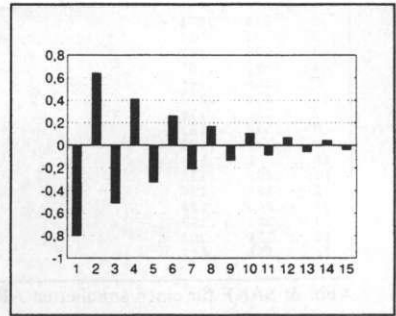


Abb. 5b: AKF des AR(1)-Prozesses: $\phi=-0.8$

Damit erhalten wir ein Instrument, mit dem wir prüfen können, ob eine beobachtete Zeitreihe die Realisation eines autoregressiven Prozesses 1. Ordnung ist bzw., ob das AR(1)-Modell als eine adäquate Darstellung des erzeugenden Prozesses gelten kann. Dazu müssen wir die empirischen Autokorrelationskoeffizienten gemäß

$$r_k = \frac{\frac{1}{n} \sum_{i=1}^{n-k} (z_i - \bar{z})(z_{i-k} - \bar{z})}{\frac{1}{n} \sum_{i=1}^n (z_i - \bar{z})^2} \quad (22)$$

ermitteln und die daraus resultierende Stichproben-Autokorrelationsfunktion (SAKF) mit der theoretischen Autokorrelationsfunktion (AKF) vergleichen. Die beiden Muster werden allerdings nie identisch sein, da die beobachtete Zeitreihe endlich ist, in der Regel nur eine geringe Zahl von Meßzeitpunkten umfaßt und die Modellvoraussetzungen in der Realität nie vollkommen erfüllt sind. Die folgende Abbildung 6 zeigt das Muster einer SAKF für einen simulierten AR(1)-Prozeß ($\phi = 0.9, \sigma_a = 2.0$) mit 200 Fällen. Da sich anhand der SAKF häufig nicht entscheiden läßt, welcher Prozeßtyp vorliegt, haben Box und Jenkins vorgeschlagen, die Partielle Autokorrelationsfunktion (PAKF) als weiteres Hilfsmittel heranzuziehen. Sie läßt sich aber besser mit Blick auf autoregressive Prozesse höherer Ordnung erläutern.

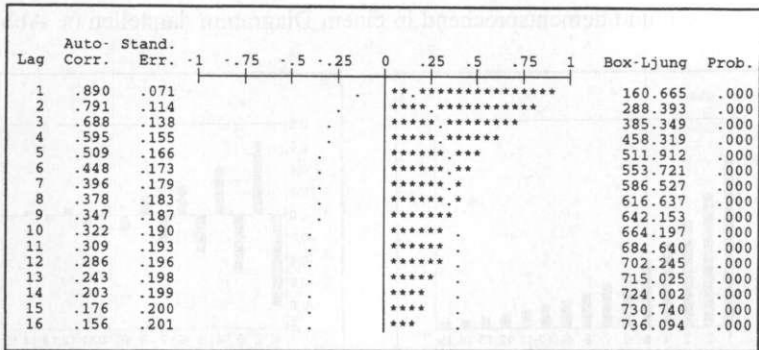


Abb. 6: SAKF für einen simulierten AR(1)-Prozeß ($\phi = 0.9, \sigma_a = 2.0$), $N=200$

Autoregressive Prozesse 2. Ordnung: AR(2)-Modelle

Der autoregressive Prozeß 2. Ordnung ist wie folgt definiert.

$$z_t = c + \phi_1 z_{t-1} + \phi_2 z_{t-2} + a_t \quad (23)$$

$$(1 - \phi_1 B - \phi_2 B^2)z_t = c + a_t$$

Wie sonst auch in einer Regressionsgleichung stellen die Gewichte ϕ_1 und ϕ_2 partielle Regressionskoeffizienten dar. In einem AR(2) Modell wird also angenommen, daß z_{t-2} einen direkten Beitrag zu dem zwei Intervalle später beobachtbaren Wert z_t leistet. Die Ableitung des Erwartungswertes sowie der Varianz und der Autokovarianzen verlangt jetzt einen etwas höheren Rechenaufwand und wird hier nicht dargestellt. Folgende Ergebnisse sind festzuhalten:

Der Erwartungswert ist

$$E(Z_t) = \frac{c}{1 - \phi_1 - \phi_2} \quad (24)$$

Die Varianz ist

$$\sigma_z^2 = \frac{\sigma_a^2}{1 - \rho_1 \phi_1 - \rho_2 \phi_2} \quad (25)$$

Zwischen den beiden ersten Autokorrelationskoeffizienten und den Phi-

Gewichten besteht folgende Beziehung:

$$\begin{aligned} \rho_1 &= \frac{\phi_1}{1 - \phi_2} \\ \rho_2 &= \phi_2 + \frac{\phi_1^2}{1 - \phi_2} \end{aligned} \quad (26)$$

Die Korrelationskoeffizienten für höhere *Lags* lassen sich rekursiv aus den beiden ersten Koeffizienten berechnen (s. Box u. Jenkins 1976, S. 59):

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} \quad ; \quad k > 0 \quad (27)$$

Die Stationaritätsbedingungen sind mit den folgenden Ungleichungen gegeben:

$$\begin{aligned} \phi_2 + \phi_1 &< 1 \\ \phi_2 - \phi_1 &< 1 \\ -1 &< \phi_2 < 1 \end{aligned}$$

Das bedeutet, daß $|\phi_1| < 2$ und $|\phi_2| < 1$. Die möglichen Kombinationen der beiden Koeffizienten, die die Stationaritätsbedingung erfüllen, sind in dem folgenden Diagramm enthalten (Abb. 7):

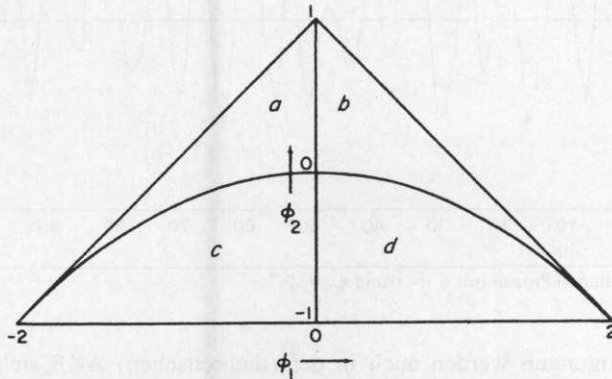


Abb. 7: Stationaritätsdreieck für AR(2)-Prozesse.

Der in das Dreieck eingezeichnete Halbkreis gibt Wertekombinationen an, die zu komplexen "Wurzeln" (Lösungen) der sog. charakteristischen Gleichung $(1 - \phi_1 B - \phi_2 B^2) = 0$ führen, wobei man den *backshift* "B" wie eine Variable "x" behandeln kann (quadratische Gleichung mit einer Unbekann-

ten). Die Lösungsformel, die man in jedem mathematischen Lehrbuch nachschlagen kann, lautet:

$$(1 - \phi_1 B - \phi_2 B^2) = 0$$

$$B_{1,2} = \frac{-\phi_1 \pm \sqrt{\phi_1^2 + 4\phi_2}}{2\phi_2} \quad (28)$$

Wenn die Wurzeln komplexe Zahlen darstellen (also Zahlen mit der imaginären Größe $i = \sqrt{-1}$), enthält der Prozeß eine stochastische Schwingung. Man bezeichnet sie auch als "pseudo-periodisch", weil die Periode (also die Länge der Schwingung) aufgrund von Zufallseinflüssen schwankt, nicht deterministisch festgelegt ist. Das gleiche gilt für die Amplitude. Abb. 8 zeigt einen mit $\phi_1 = 1$ und $\phi_2 = -0.75$ simulierten Prozeß dieser Art. Ein Beispiel mit realen Daten wird im nächsten Abschnitt präsentiert.

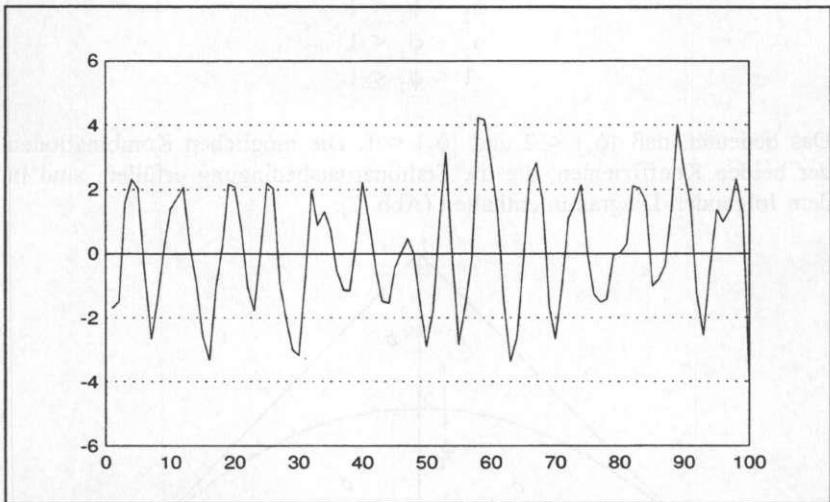


Abb. 8: Simulierter Prozeß mit $\phi_1 = 1$ und $\phi_2 = -0.75$

Die Schwingungen werden auch in der (theoretischen) AKF sichtbar, die ganz unterschiedliche Formen annehmen kann, je nachdem welche Gewichte ϕ_1 und ϕ_2 vorliegen. In Abbildung 9 ist die SAKF des simulierten AR(2)-Prozesses wiedergegeben. Die SP AKF erscheint weiter unten in Abb. 10.

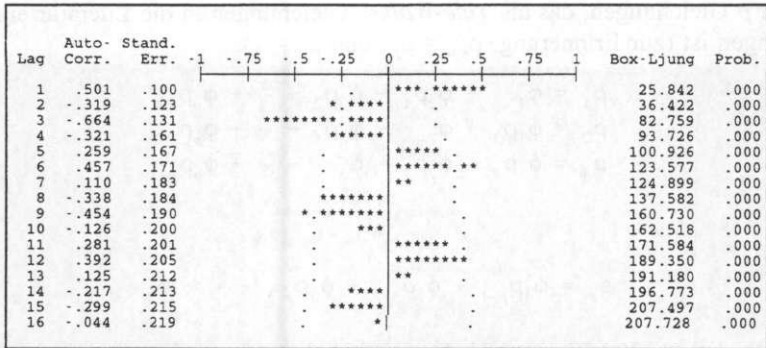


Abb. 9: SAKF des simulierten AR(2)-Prozesses

Autoregressive Prozesse höherer Ordnung: AR(p)-Modelle

Gelegentlich hat man es auch in den Sozialwissenschaften mit autoregressiven Prozessen höherer Ordnung zu tun, wobei wir von nun an den Ordnungsgrad allgemein mit dem Parameter "p" darstellen wollen:

$$z_t = c + \phi_1 z_{t-1} + \phi_2 z_{t-2} + \dots + \phi_p z_{t-p} + a_t$$

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) z_t = a_t \quad (29)$$

Erwartungswert und Varianz ergeben sich aus Verallgemeinerungen der entsprechenden Gleichungen für den AR(2)-Prozeß:

$$E(Z_t) = \frac{c}{1 - \phi_1 - \phi_2 - \dots - \phi_p} \quad (30)$$

$$\sigma_z^2 = \frac{\sigma_a^2}{1 - \rho_1 \phi_1 - \rho_2 \phi_2 - \dots - \rho_p \phi_p} \quad (31)$$

Die Autokorrelationskoeffizienten und die Phi-Gewichte sind durch die Gleichung

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \dots + \phi_p \rho_{k-p} ; k > 0 \quad (32)$$

miteinander verbunden. Bevor man sie rekursiv zur Berechnung der ρ_k , $k > p$ einsetzen kann, müssen die ersten p Autokorrelationskoeffizienten bekannt sein. Substituiert man in Gleichung (31) $k = 1, 2, \dots, p$, erhält man ein System

mit p Gleichungen, das als *Yule-Walker* Gleichungen in die Literatur eingegangen ist (zur Erinnerung: $\rho_{k-p} = \rho_{p-k}$ und $\rho_0 = 1$):

$$\begin{aligned}
 \rho_1 &= \phi_1 + \phi_2\rho_1 + \phi_3\rho_2 + \dots + \phi_p\rho_{p-1} \\
 \rho_2 &= \phi_1\rho_1 + \phi_2 + \phi_3\rho_1 + \dots + \phi_p\rho_{p-2} \\
 \rho_3 &= \phi_1\rho_2 + \phi_2\rho_1 + \phi_3 + \dots + \phi_p\rho_{p-3} \\
 &\vdots \\
 \rho_p &= \phi_1\rho_{p-1} + \phi_2\rho_{p-2} + \phi_3\rho_{p-3} + \dots + \phi_p
 \end{aligned}
 \tag{33}$$

Falls die Phi-Koeffizienten bekannt sind, kann man daraus die p ersten Korrelationskoeffizienten bestimmen. Andererseits: wenn man die theoretischen Autokorrelationskoeffizienten durch die empirischen Autokorrelationskoeffizienten ersetzt, lassen sich aus diesem Gleichungssystem nach den Regeln der Matrixrechnung vorläufige Schätzer für die Phi-Gewichte ableiten.

Die Stationaritätsbedingungen eines autoregressiven Prozesses lassen sich nun ebenfalls allgemein angeben: Ein AR(p)-Prozeß ist stationär, wenn die Wurzeln des Polynoms $(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)$ alle außerhalb des Einheitskreises liegen,² wenn sie in ihrem Absolutbetrag also größer 1 sind. Einige Programmpakete zur Zeitreihenanalyse bieten Routinen an, mit denen man die Wurzeln beliebiger Polynome errechnen kann. Das ist u. a. deshalb nützlich, weil die Schätzalgorithmen in der Regel instabil werden und verzerrte Ergebnisse liefern, wenn die Schätzer nahe an die Stationaritätsgrenzen heranreichen.

Auch die AR-Prozesse höherer Ordnung lassen sich umformen zu (s. Gleichung (14*))

$$\begin{aligned}
 (1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) z_t &= a_t \\
 z_t &= (1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)^{-1} a_t \\
 z_t &= (1 - \psi_1 B - \psi_2 B^2 - \dots - \psi_\infty B^\infty) a_t
 \end{aligned}
 \tag{34}$$

Diese sog. "Psi-Gewichte-Form" wird u. a. zur Ableitung von Prognoseglei-

2 Wie wir weiter oben (s. Gleichung (28)) sahen, können die Lösungen der charakteristischen Gleichungen konjugiert komplexe Zahlen ergeben, die sich bekanntlich nicht auf einem (eindimensionalen) Zahlenstrahl abtragen lassen, sondern einer zweidimensionalen Darstellung bedürfen. Daher rührt die Rede vom "Einheitskreis".

chungen und zur Modellierung von Ausreißereffekten benötigt. In unserem jetzigen Kontext ist sie nützlich, weil sich aus den Psi-Gewichten für beliebige stationäre Prozesse die Varianzen/Kovarianzen und damit auch die (theoretischen) Autokorrelationskoeffizienten ableiten lassen. Die Autokovarianz ist gegeben durch

$$\gamma_k = \sigma_a^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+k} \quad (35)$$

(s. Box u. Jenkins 1976, S. 48 f.). Daraus erhält man unmittelbar die Varianz

$$\gamma_0 = \sigma_a^2 \sum_{j=0}^{\infty} \psi_j^2 \quad (36)$$

Diese Gleichung verdeutlicht nochmals die Bedeutung der Stationaritätsvoraussetzung: Nur wenn die Psi-Gewichte rasch gegen Null konvergieren, ist die Varianz des stochastischen Prozesses endlich.

Der Quotient beider Ausdrücke liefert eine allgemeine Formel für den Autokorrelationskoeffizienten $\rho_k = \gamma_k/\gamma_0$, $k > 0$.

4. Die partielle Autokorrelationsfunktion

Das Konzept der partiellen Korrelation ist aus der Elementarstatistik bekannt. Die partielle Korrelation $r_{1,2}$ zwischen den Variablen X_1 und X_2 "unter Konstanthalten" von X_3 ist die einfache Korrelation der Residuen von X_1 und X_2 , die man erhält, wenn man diese beiden Variablen auf die Kontrollvariable X_3 regrediert. Inhaltlich kann dieses Konzept ohne weiteres auf die Autokorrelation von Zeitreihen übertragen werden. Nehmen wir als Beispiel einen autoregressiven Prozeß zweiter Ordnung:

$$z_t = \phi_1 z_{t-1} + \phi_2 z_{t-2} + a_t \quad (37)$$

Die Variablen X_1 und X_2 sind durch die Variablen Z_t und Z_{t-2} , die Variable X_3 durch die Variable Z_{t-1} ersetzt worden. Gesucht wird folglich ein Koeffizient, der den partiellen (direkten) Zusammenhang zwischen den Zeitreihenwerten angibt, die um ein Intervall voneinander getrennt sind. Gefragt wird: Wie stark ist der Einfluß von Z_{t-2} auf Z_t soweit er nicht durch Z_{t-1} vermittelt ist. In der Zeitreihenanalyse tritt nun die Besonderheit auf, daß die (partiellen) Regressionskoeffizienten, also die Phi-Gewichte, immer schon Korrelationskoeffizienten sind. Bekanntlich unterscheiden sich Pearsons Korrelationskoeffizient und der Regressionskoeffizient nur dadurch, daß im ersten Fall

die Kovarianz durch die Standardabweichungen der beiden Variablen, im zweiten Fall durch die Varianz der Regressorvariable dividiert wird. Die Standardabweichungen der Regressoren Z_i und Z_{i-2} (oder allgemein $Z_{i,k}$, $k < n$) sind aber identisch, so daß das Produkt der beiden Standardabweichungen gleich der Varianz jeder der beiden Variablen ist. Das bedeutet, daß in einem AR(2)-Prozeß der (partielle) Regressionskoeffizient ϕ_2 gleichzeitig der partielle Autokorrelationskoeffizient für den direkten Zusammenhang zwischen $Z_{i,2}$ und Z_i ist. Im Unterschied zur vertrauten Terminologie aus der Querschnittsanalyse bezeichnet man ihn als partiellen Regressionskoeffizienten 2. Ordnung (zum Lag $k = 2$), obwohl hier nur eine einzige Kontrollvariable, $Z_{i,1}$, vorliegt. In einem AR(3)-Prozeß

$$z_t = c + \phi_1 z_{t-1} + \phi_2 z_{t-2} + \phi_3 z_{t-3} + a_t \quad (38)$$

ist ϕ_3 der partielle Autokorrelationskoeffizient 3. Ordnung, denn er repräsentiert die direkte Beziehung zwischen Z_i und $Z_{i,3}$ unter Konstanthalten von $Z_{i,1}$ und $Z_{i,2}$. Dagegen ist ϕ_2 in der Gleichung (38) nicht der partielle Autokorrelationskoeffizient zweiter Ordnung, denn er repräsentiert ja die direkte Beziehung zwischen Z_i und $Z_{i,2}$ unter Konstanthalten von $Z_{i,1}$ und $Z_{i,3}$, ist also ebenfalls ein partieller Koeffizient 3. Ordnung. Den partiellen Auto-Korrelationskoeffizienten 2. Ordnung für einen AR(3)-Prozeß erhält man, indem man mit den Daten dieses Prozesses die Regressionskoeffizienten der Gleichung $z_t = c + \phi_1 z_{t-1} + \phi_2 z_{t-2} + a_t$ bestimmt. Das letzte Phi-Gewicht, ϕ_2 , ist der gesuchte partielle Autokorrelationskoeffizient 2. Ordnung.

Allgemein ermittelt man die Autokorrelationskoeffizienten erster bis p-ter Ordnung für einen AR(p)-Prozeß, indem man die *Yule-Walker* Gleichungen schrittweise für $k = 1, 2, \dots, p$ nach den Phi-Koeffizienten löst. Auf diese Weise erhält man eine Serie von Regressionskoeffizienten $\hat{\phi}_{kj}$, wobei $k = 1, 2, \dots, p$ und $j = 1, 2, \dots$. Für einen AR(3)-Prozeß sind also folgende Gleichungssysteme zu lösen:

$$k = 1: \quad r_1 = \hat{\phi}_{11} \quad (39a)$$

$$k = 2: \quad \begin{bmatrix} r_1 \\ r_2 \end{bmatrix} = \begin{bmatrix} 1 & r_1 \\ r_1 & 1 \end{bmatrix} \times \begin{bmatrix} \hat{\phi}_{2,1} \\ \hat{\phi}_{2,2} \end{bmatrix} \quad (39b)$$

$$k = 3: \begin{bmatrix} r_1 \\ r_2 \\ r_3 \end{bmatrix} = \begin{bmatrix} 1 & r_1 & r_2 \\ r_1 & 1 & r_1 \\ r_2 & r_1 & 1 \end{bmatrix} \times \begin{bmatrix} \hat{\phi}_{3,1} \\ \hat{\phi}_{3,2} \\ \hat{\phi}_{3,3} \end{bmatrix} \quad (39c)$$

Die jeweils letzten Gewichte des Vektors $(\hat{\phi}_{k1}, \hat{\phi}_{k2}, \dots, \hat{\phi}_{kk})$ also die p -mal ermittelten $\hat{\phi}_{kk}$, sind die partiellen Auto-Korrelationskoeffizienten erster bis p -ter Ordnung. Für die Zwecke der Partiiellen Autokorrelationsfunktion PAKF bzw. SP AKF definiert man also den Koeffizienten $\hat{\phi}_{11}$ als einen partiellen Korrelationskoeffizienten 1. Ordnung (zum Lag 1), obwohl er die einfache Korrelation zwischen Z_t und Z_{t-1} ohne Konstanthalten irgendwelcher anderer Faktoren angibt. Die theoretische PAKF eines AR(p)-Prozesses hat Werte ungleich Null für $k = 1, 2, \dots, p$ und danach (für höhere Lags $k > p$) nur noch Werte gleich Null (s. Abb. 10).

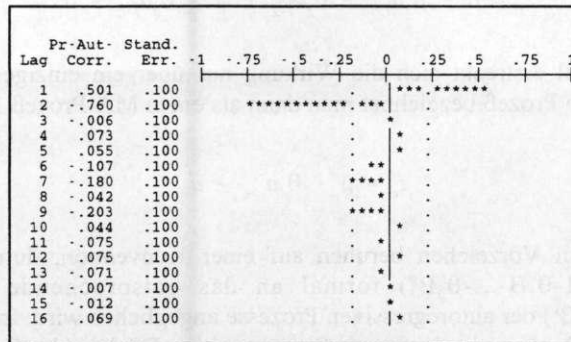


Abb. 10: SPAKF des simulierten AR(2)-Prozesses

Die Funktion hat somit theoretisch einen klaren Abbruchpunkt; das macht sie als Instrument der empirischen Identifikation eines AR(p)-Prozesses besonders geeignet, wenn auch die empirische SPAKF einer endlichen Reihe $\{Z_t\}$ Koeffizienten $\hat{\phi}_k \neq 0$ bei $k > p$ aufweist. In Abschnitt 7 unten wird dies weiter veranschaulicht.

5. Moving Average Prozesse: MA-Modelle

Autoregressive Modelle schreiben gegenwärtige Werte z_t als Funktion vergangener Werte z_{t-k} , $k = 1, 2, \dots, q$. Die treibende Kraft des Systems sind jedoch die Zufallsereignisse a_k , $k = 0, 1, 2, \dots, \infty$, ohne die sich die Werte der Zeitreihe nicht verändern würden und deren Einflußdauer in dieser Modellklasse (theoretisch) als "unendlich lang" formalisiert wird (s. Gleichungen (8) bis (14)). Indem man auf der rechten Gleichungsseite nicht unendlich viele "Input"-Terme $a_{t-1}, \dots, a_{t-\infty}$ einsetzt, sondern sie durch einen oder mehrere verzögerte Terme der "endogenen" Variable (z_{t-1} , z_{t-2} usw.) substituiert, erreicht man lediglich eine sparsamere Parametrisierung des Modells, ohne seinen analytischen Gehalt, wie er in der Psi-Gewichte-Form sichtbar wird, zu verändern. Nun kann man sich aber auch Systeme vorstellen, in denen ein Zufallsereignis, das zu einem beliebigen Zeitpunkt auftritt, nicht in alle Zukunft hinein wirksam bleibt, auch nicht mit ständig abnehmendem Gewicht. Solche Prozesse werden in den sog. *moving average* Modellen³ dargestellt:

$$z_t = \mu - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q} + a_t \quad (40)$$

Im Extremfall erstreckt sich die Wirkung nur über ein einziges Intervall; einen solchen Prozeß bezeichnet man dann als einen MA-Prozeß 1. Ordnung - MA(1):

$$z_t = \mu - \theta_1 a_{t-1} + a_t \quad (41)$$

Die negativen Vorzeichen beruhen auf einer Konvention, durch die das Polynom $(1 - \theta_1 B - \dots - \theta_q B^q)$ formal an das entsprechende Polynom $(1 - \phi_1 B - \dots - \phi_p B^p)$ der autoregressiven Prozesse angeglichen wird. In der Praxis muß man sich aber erst daran gewöhnen, positive Effekte (des Zufallsereignisses) mit Hilfe negativer θ 's darstellen zu sollen ("minus" mal "minus" gibt "plus").

Der substantielle Unterschied zwischen einem MA(q) und einem AR(p) Modell mag durchaus so gering sein, daß er praktisch nicht ins Gewicht fällt: Die Psi-Gewichte - und damit die Autokorrelationskoeffizienten - eines AR-Prozesses können sehr rasch gegen Null gehen, während andererseits das "endliche Gedächtnis" des MA-Prozesses eine große Zeitspanne umfassen

3 Der Ausdruck "moving average" (Gleitender Mittelwert) ist insofern irreführend, als sich die Gewichte nicht zum Betrag "1" summieren, insoweit also überhaupt keine Durchschnittsbildung vorliegt.

kann. Zwischen AR-Modellen und MA-Modellen entscheiden also kaum inhaltliche, sondern eher formal-statistische Gründe: Man wählt dasjenige Modell, das mit einer minimalen Anzahl von Parametern einen beobachteten Prozeß adäquat repräsentiert. Ob es überhaupt inhaltlich relevante Unterschiede zwischen verschiedenen Modellen gibt, hängt nicht nur von dem Modelltyp, sondern auch von der Größe und der Zahl der Parameter ab. Sichtbar wird ein etwaiger Unterschied, wenn das AR-Modell in seine Psi-Gewichte-Form übersetzt worden ist; die Psi-Koeffizienten des AR-Prozesses können dann mit den Theta-Koeffizienten des MA-Prozesses verglichen werden.

Unabhängig von ihrem Ordnungsgrad und der Größe ihrer Parameter sind MA-Prozesse mittelwertstationär. Dies wird formal durch die Bildung des Erwartungswertes bestätigt:

$$\begin{aligned}
 E(z_t) &= E(\mu - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q} + a_t) \\
 &= \mu - \theta_1 E(a_{t-1}) - \theta_2 E(a_{t-2}) - \dots - \theta_q E(a_{t-q}) + E(a_t) \quad (42) \\
 &= \mu, \text{ da } E(a_{t-k}) = 0 \text{ für alle } k = 0, 1, \dots, q
 \end{aligned}$$

Es ist also nicht nötig, die Größe der Theta-Koeffizienten zu begrenzen, um die Stationarität sicherzustellen (eine solche Restriktion wird aber, wie wir gleich sehen werden, aus anderen Gründen nötig). Die Konstante μ eines MA-Prozesses ist (im Unterschied zum AR-Prozeß) identisch mit dem arithmetischen Mittel.

Bei der folgenden Ableitung der Varianz gehen wir aus rechentechnischen Gründen von mittelwertbereinigten Daten aus (ohne Änderung der Notation), so daß die Varianz als $E(z_t^2)$ definiert werden kann:

$$\begin{aligned}
 E(z_t^2) &= E(a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q})^2 \\
 &= E(a_t^2) + \theta_1^2 E(a_{t-1}^2) + \theta_2^2 E(a_{t-2}^2) + \dots + \theta_q^2 E(a_{t-q}^2) \quad (43) \\
 &= \sigma_a^2 (1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2) = \gamma_0
 \end{aligned}$$

Die Varianz ist also ebenfalls endlich und konstant. Die Autokovarianz ist durch folgende Gleichung gegeben:

$$\begin{aligned}
 E(z_t z_{t-k}) &= E((a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q}) \cdot \\
 &\quad (a_{t-k} - \theta_1 a_{t-k-1} - \theta_2 a_{t-k-2} - \dots - \theta_q a_{t-k-q})) \\
 &= -\theta_k E(a_{t-k}^2) + \theta_1 \theta_{k+1} E(a_{t-k-1}^2) + \theta_2 \theta_{k+2} E(a_{t-k-2}^2) + \dots \quad (44) \\
 &\quad \dots + \theta_{q-k} \theta_q E(a_{t-k-q}^2) \\
 &= \sigma_a^2 (-\theta_k + \theta_1 \theta_{k+1} + \theta_2 \theta_{k+2} + \dots + \theta_{q-k} \theta_q) = \gamma_k
 \end{aligned}$$

Der Autokorrelationskoeffizient $\rho_k = \gamma_k / \gamma_0$ ist somit

$$\rho_k = \frac{-\theta_k + \theta_1\theta_{k+1} + \theta_2\theta_{k+2} + \dots + \theta_{q-k}\theta_q}{1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2} \quad (45)$$

Definitionsgemäß sind alle θ_k mit $k > q$ gleich Null. Daraus folgt, daß auch alle Autokorrelationskoeffizienten ρ_k bei $k > q$ gleich Null sind. Das heißt: die Autokorrelationsfunktion eines MA($\hat{\quad}$)-Prozesses hat einen scharfen Abbruch an der Lag-Stelle $k=q$. Die empirische Autokorrelationsfunktion (SAKF) sollte also ein geeignetes Instrument sein, die Ordnung eines MA(q)-Prozesses aufzudecken. Die folgende Abbildung 11 stellt einen mit $\theta_1 = 1,5$ und $\theta_2 = -0,8$ simulierten MA(2) Prozeß und dessen empirische Autokorrelationsfunktion (SAKF) dar.

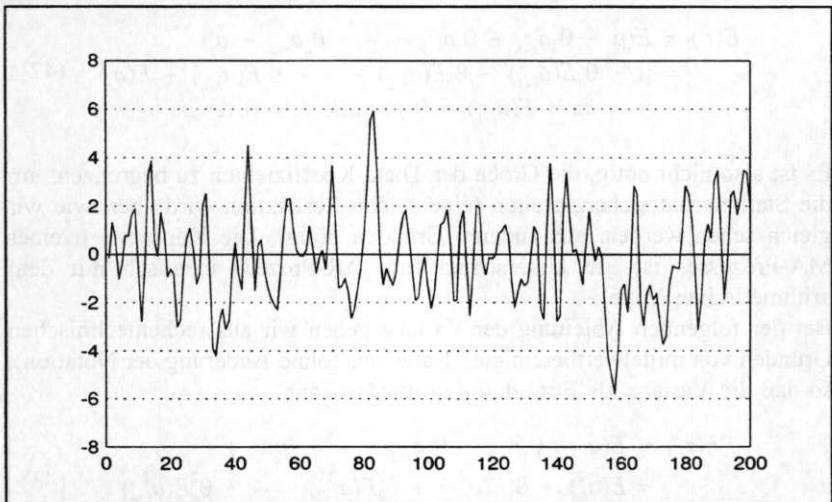


Abb. 11 a: Simulierter MA(2)-Prozeß mit $\theta_1 = 1,5$ und $\theta_2 = -0,8$

Lag	Auto-Corr.	Stand. Err.	-1	-.75	-.5	-.25	0	.25	.5	.75	1	Box-Ljung	Prob.
1	-.761	.071	*****				*					117.587	.000
2	-.354	.104					***	***				143.225	.000
3	-.108	.110					**	.	.			145.635	.000
4	.011	.110					.	.	.			145.661	.000
5	.020	.110					.	.	.			145.740	.000
6	-.029	.110					.	.	.			145.916	.000
7	.067	.110					*	.	.			146.851	.000
8	-.077	.111					**	.	.			148.095	.000
9	-.042	.111					.	*	.			148.475	.000
10	-.029	.111					.	*	.			148.654	.000
11	.060	.111					.	*	.			149.422	.000
12	-.108	.111					**	.	.			151.938	.000
13	.131	.112					.	***	.			155.622	.000
14	-.061	.112					.	*	.			156.425	.000
15	-.093	.113					.	**	.			158.307	.000
16	.195	.113					.	****	.			166.659	.000

Abb. 11 b: SAKF des simulierten Prozesses

Die Autokorrelationskoeffizienten r_i einer endlichen empirischen Reihe sind bei $k > q$ zwar nicht exakt Null (was sie bei einer endlichen, empirischen Zeitreihe nie sind), schwanken um diesen Betrag aber innerhalb gewisser Grenzen, die als Vertrauensintervall spezifizierbar sind (zur Berechnung der Standardfehler siehe Abschnitt 7).

Um die Gestalt der partiellen Autokorrelationsfunktion eines MA-Prozesses zu klären, muß zuvor das Konzept der "Invertibilität" eines MA-Prozesses eingeführt werden. Dazu betrachten wir zunächst die Autokovarianz eines MA(1)-Prozesses ($q=1$) zum Lag $k=1$, die sich unmittelbar aus Gleichung (44) ergibt:

$$\gamma_1 = -\theta\sigma_a^2 \quad (46)$$

da $\theta_0 = 1$ (definitionsgemäß) und $\theta_k = 0$ für $k > q=1$. Ebenso läßt sich der entsprechende Autokorrelationskoeffizient aus Gleichung (45) ablesen:

$$\rho_1 = \frac{-\theta}{1 + \theta^2} \quad (47)$$

Der Autokorrelationskoeffizient steht mit dem Theta-Koeffizienten in einer nicht-linearen Beziehung. Gleichung (47) weist aber noch eine weitere Besonderheit auf; sie ist auch für einen Koeffizienten $\theta^* = 1/\theta$ erfüllt, wie sich leicht zeigen läßt:

$$\frac{-\left(\frac{1}{\theta}\right)}{1 + \left(\frac{1}{\theta}\right)^2} = \frac{-\theta}{1 + \theta^2} \quad (48)$$

Das heißt, die Autokorrelationsfunktion ist durch die Thetas nicht eindeutig festgelegt; sie ist für θ und für $1/\theta$ gleich. Box und Jenkins (1976) haben vorgeschlagen, nur diejenigen Thetas als Modellparameter zu akzeptieren, die die sog. Invertibilitätsbedingung erfüllen. Was ist damit gemeint? Wenn ein MA(1)-Prozeß

$$a_t = z_t - \mu + \theta a_{t-1} \quad (49)$$

$$a_t = z_t - \mu + \theta a_{t-1} \quad (49)$$

vorliegt, gelten auch die folgenden Gleichungen

$$\begin{aligned}
 a_{t-1} &= z_{t-1} - \mu + \theta a_{t-2} \\
 a_{t-2} &= z_{t-2} - \mu + \theta a_{t-3} \\
 a_{t-3} &= z_{t-3} - \mu + \theta a_{t-4} \\
 &\vdots \quad \quad \quad \vdots \quad \quad \quad \vdots \quad \quad \quad \vdots
 \end{aligned}
 \tag{50}$$

Die erste dieser Gleichungen in (49) eingesetzt führt zu

$$\begin{aligned}
 a_t &= z_t - \mu + \theta(z_{t-1} - \mu + \theta a_{t-2}) \\
 &= z_t - \mu + \theta z_{t-1} - \theta\mu + \theta^2 a_{t-2}
 \end{aligned}
 \tag{51}$$

Der Term a_{t-2} läßt sich durch die zweite Gleichung in (50) ersetzen usw. Durch fortgesetzte Substitution (analog zu den Gleichungen (10) bis (14)) erhalten wir:

$$a_t = z_t + \theta z_{t-1} + \theta^2 z_{t-2} + \dots + \theta^{k-1} z_{t-k+1} + \theta^k a_{t-k} - \mu(1 + \theta + \theta^2 + \dots + \theta^{k-1})
 \tag{52}$$

Unter der Invertibilitätsbedingung $\theta < |1|$ wird daraus bei $k \rightarrow \infty$ ein infiniter autoregressiver Prozeß

$$\begin{aligned}
 a_t &= z_t + \theta z_{t-1} + \theta^2 z_{t-2} + \dots - \frac{\mu}{1 - \theta} \\
 z_t &= \frac{\mu}{1 - \theta} - \theta z_{t-1} - \theta^2 z_{t-2} - \dots + a_t
 \end{aligned}
 \tag{53}$$

Das heißt, ein MA(1)-Modell und das (entsprechende) AR(∞)-Modell sind - unter der genannten Voraussetzung - zwei unterschiedliche Darstellungsweisen desselben Prozesses. Das gilt auch für die MA(\wedge)-Prozesse höherer Ordnung. (Wir ziehen natürlich das sparsamere Modell, das mit der (weitaus) geringeren Zahl von Parametern, vor.) Analog zur Gleichung (34) lassen sich alle invertiblen MA(q)-Prozesse in die sog. "Pi-Gewichte-Form" transformieren:

$$MA(q) \rightarrow AR(\infty):$$

$$\begin{aligned}
 z_t &= (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) a_t \\
 (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q)^{-1} z_t &= a_t \\
 (1 + \pi_1 B + \pi_2 B^2 + \dots + \pi_\infty B^\infty) z_t &= a_t
 \end{aligned}
 \tag{54}$$

Im Falle eines MA-Prozesses 1. Ordnung erhalten wir

$$\begin{aligned}
 z_t &= (1 - \theta B)a_t \\
 (1 - \theta B)^{-1}z_t &= a_t \\
 (1 + \theta B + \theta^2 B^2 + \dots)z_t &= a_t \\
 (1 + \pi_1 B + \pi_2 B^2 + \dots)z_t &= a_t
 \end{aligned}
 \tag{55}$$

Zwischen den π -Gewichten und den θ -Gewichten besteht also die einfache Beziehung $\pi_j = -\theta^j$. Falls entgegen der Invertibilitätsvoraussetzung $|\theta| \geq 1$ wäre, bedeutete dies, daß gegenwärtige z-Werte von vergangenen z-Werten um so stärker abhingen, je weiter sie in die Vergangenheit zurückreichten - eine ziemlich unplausible Vorstellung, die den Sinn der Invertibilitätsvoraussetzung unterstreicht.

Die Pi-Gewichte-Form des MA(q)-Prozesses hat die gleiche Struktur wie die Psi-Gewichte-Form eines AR($p=q$)-Prozesses. Während die Psi-Gewichte-Form eines stationären stochastischen Prozesses die Gestalt der Autokorrelationsfunktion prägt, beeinflusst die Pi-Gewichte-Form die Gestalt der Partiellen Autokorrelationsfunktion. Daraus folgt, daß die Partielle Autokorrelationsfunktion eines MA(q)-Prozesses keinen Abbruch hat, sondern nach dem Lag $k-q$ gegen Null abfällt. Die AKF eines AR(p)-Prozesses hat also die gleiche Musterung wie die PAKF des MA($q=p$)-Prozesses. Abbildung 12 zeigt die SPAKF des simulierten MA(2) Prozesses aus Abbildung 11.

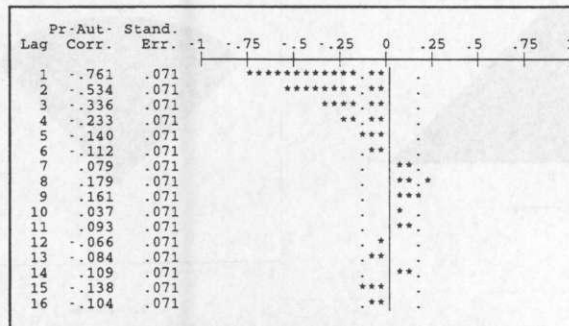


Abb. 12: SPAKF des simulierten MA(2)-Prozesses

Umgekehrt entspricht, wie wir schon sahen, das Muster der AKF eines MA(q)-Prozesses dem Muster der PAKF eines AR($p=q$)-Prozesses. Diese Beziehungen bezeichnet man auch als "Dualität" von AR- und MA-Prozessen (O.D. Anderson 1976, S. 42).

Eine interessante Facette sei noch nebenbei erwähnt. Betrachten wir noch einmal Gleichung (47), so erkennen wir, daß $\theta = 1$ ein $\rho = 0,5$ impliziert. Der maximale Autokorrelationskoeffizient eines invertiblen MA(1)-Prozesses liegt also unterhalb des Wertes von $\rho = 0,5$.

Wie wir sahen, stellt die Invertibilitätsbedingung des MA-Prozesses das mathematische Analogon zur Stationaritätsbedingung des AR-Prozesses dar. Das bedeutet, die Invertibilität eines MA-Prozesses ist gegeben, wenn alle Wurzeln des Theta-Polynoms, d.h alle Lösungen der Gleichung

$$(1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) = 0 \quad (56)$$

außerhalb des Einheitskreises liegen. Für einen MA(2)-Prozeß läßt sich diese Bedingung in die folgenden Ungleichungen übersetzen:

$$\begin{aligned} \theta_2 + \theta_1 &< 1 \\ \theta_2 - \theta_1 &< 1 \\ -1 &< \theta_2 < 1 \end{aligned} \quad (57)$$

Abbildung 13 stellt die "erlaubten" Wertebereiche für die Theta-Gewichte und die Autokorrelationskoeffizienten eines invertiblen MA(2)-Prozesses dar.

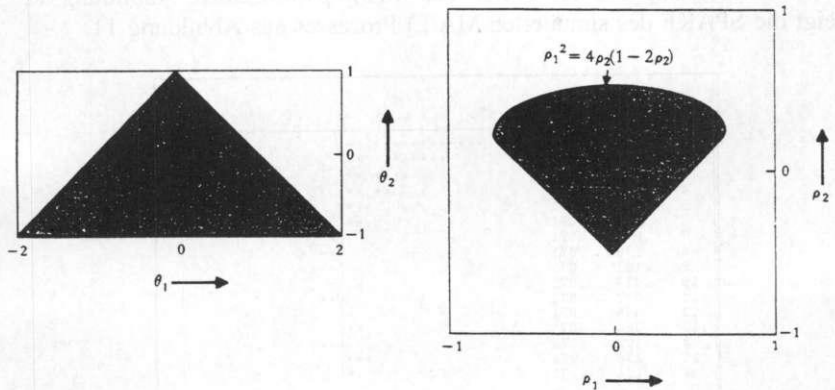


Abb. 13: Bereich erlaubter Parameter eines invertierbaren MA(2)-Prozesses.

MA-Prozesse höherer Ordnung ($q > 2$) werden in den Sozialwissenschaften selten identifiziert.

6. Mischprozesse: ARMA-Modelle

Die Zeitreihen, die dem Historiker oder Sozialwissenschaftler zur Verfügung stehen, sind in der Regel recht kurz, umfassen häufig weniger als 100 Meßzeitpunkte. Bei so wenigen Fällen wird die Schätzung der Modellparameter relativ unsicher; die Standardfehler werden mit jedem zusätzlichen Parameter, der zu schätzen ist, deutlich größer. Deshalb ist es wichtig, Modelle mit einer möglichst geringen Zahl von Parametern zu spezifizieren (Prinzip der *parsimony*). Bei einigen Prozessen kann die Zahl der Parameter weiter verringert werden, wenn man nicht "reine" AR- oder MA-Modelle, sondern sog. Mischmodelle verwendet, die sowohl eine (endliche) AR- als auch eine (endliche) MA-Komponente (man spricht auch von "Faktoren") enthalten. Man bezeichnet sie als ARMA(p,q)-Modelle. Die reinen AR- und MA-Modelle werden gelegentlich auch als ARMA($p,0$)- und ARMA($0,q$)-Modelle dargestellt. Sie sind wie folgt definiert:

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) z_t = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) a_t \quad (58)$$

Soll der ARMA-Prozeß stationär und invertierbar sein, müssen die Nullstellen der folgenden Gleichungen

$$\begin{aligned} \phi_p(B) &= (1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) = 0 \\ \theta_q(B) &= (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) = 0 \end{aligned} \quad (59)$$

alle außerhalb des Einheitskreises liegen, d. h. die AR-Komponente muß stationär und die MA-Komponente muß invertierbar sein. Unter diesen Voraussetzungen ist ein ARMA(p,q)-Modell sowohl in einen unendlichen AR-Prozeß als auch in einen unendlichen MA-Prozeß überführbar:

ARMA(p,q) \rightarrow AR(∞)

$$\begin{aligned} (1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) z_t &= (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) a_t \\ z_t &= \frac{(1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) a_t}{(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)} \\ &= \frac{\theta_q(B)}{\phi_p(B)} a_t = \psi(B) a_t \end{aligned} \quad (60)$$

ARMA(p,q) → MA(∞)

$$\frac{(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)}{(1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q)} z_t = a_t, \quad (61)$$

$$\frac{\phi_p(B)}{\theta_q(B)} z_t = \pi(B) z_t = a_t,$$

Vorauszusetzen ist ferner, daß die AR- und MA-Polynome keine gemeinsamen (gleichen) Wurzeln aufweisen. Um dieses Problem zu erläutern, betrachten wir einen ARMA(1,1)-Prozeß

$$(1 - \phi B) z_t = (1 - \theta B) a_t, \quad (62)$$

$$z_t = \frac{(1 - \theta B)}{(1 - \phi B)} a_t,$$

Falls $\phi = \theta$, ist der Quotient $(1-\theta B)/(1-\phi B) = 1$ mit der Folge, daß $z_t = a_t$. Das heißt, es läge in Wirklichkeit kein ARMA(1,1), sondern ein *white noise* Prozeß vor. Bei der Schätzung der Prozeßparameter mit empirischen Daten werden wir zwar kaum identische ϕ - und θ -Schätzer erhalten; möglicherweise liegen sie aber so nahe zusammen, daß eine (Beinahe-)Parameter-Redundanz gegeben ist, die bei kurzen Zeitreihen zu sehr instabilen Schätzergebnissen führen kann. Erläuterungen hierzu finden sich in Pankratz (1983, S. 203 ff.). Man sollte also in einem solchen Fall auf jeden Fall prüfen, ob ein Modell mit einer geringeren Zahl von Parametern zu akzeptablen Ergebnissen führt. (Zu verschiedenen Kriterien der Modellgüte siehe Abschnitt 9). Eine eventuelle Parameterredundanz ist nicht immer so leicht erkennbar wie im Falle des scheinbaren ARMA(1,1)-Modells. Box u. Jenkins (1976, S. 248 ff.) geben folgendes Beispiel eines ARMA(2,1) Modells:

$$(1 - 1.3B + 0.4B^2) z_t = (1 - 0.5B) a_t, \quad (63)$$

Die Parameterredundanz wird hier erst sichtbar, wenn das Polynom 2. Grades faktorisiert wird:

$$(1 - 0.5B)(1 - 0.8B) z_t = (1 - 0.5B) a_t, \quad (64)$$

Das heißt, es handelt sich in Wirklichkeit, um einen AR(1)-Prozeß. Wandelt man das Modell (63) leicht ab zu

$$(1 - 0.4B)(1 - 0.8B) z_t = (1 - 0.5B) a_t, \quad (65)$$

so heben sich die beiden Faktoren zwar nicht exakt auf; sie fallen aber kaum

ins Gewicht. Das wird deutlich, wenn man das ARMA(2,1)-Modell (63) in die Pi-Gewichte-Form überführt:

$$(1 - 0.700B - 0.030B^2 - 0.015B^3 - 0.008B^4 - \dots)z_t = a_t \quad (66)$$

Die Pi-Gewichte π_j , $j > 1$, liegen nahe bei Null, so daß Modell (63) durch das AR(1)-Modell

$$(1 - 0.7B)z_t = a_t \quad (67)$$

ersetzt werden kann, was zur Erzielung stabiler Schätzergebnisse (geringer Standardfehler) sinnvoller wäre. Schlittgen u. Streitberg (1989, S. 199 f.) empfehlen, als Teil der Modelldiagnose (s. unten Abschnitt 9) stets die Nullstellen der AR- und MA-Polynome zu berechnen (was aber nicht alle Standard-Programme zur Zeitreihenanalyse leisten). Liegen Nullstellen zu nahe beieinander oder sind sie sogar identisch, enthält das Modell zuviele Parameter (d. h., es ist unteridentifiziert). Es liegt dann eine ähnliche Situation vor wie im Falle der Multikollinearität im Regressionsmodell.

Mischprozesse sind relativ schwer identifizierbar, da, wie wir sahen, weder die AKF noch die PAKF an einer bestimmten Stelle abbrechen, sondern beide nur allmählich (mehr oder weniger rasch) gegen Null gehen. Das Muster der Autokorrelationskoeffizienten ρ_k ist ab Lag $k \geq q+1$ von der AR-Komponente und bei den Lags $0 < k \leq q$ von der MA-Komponente bestimmt.

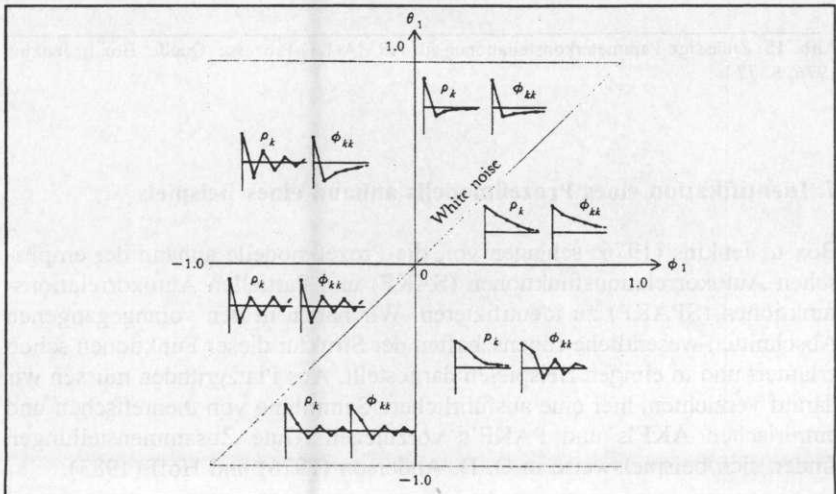


Abb. 14: Muster der AKF und PAKF verschiedener ARMA(1,1)-Prozesse. Quelle: Box u. Jenkins 1976, S. 78

Dies folgt aus dem oben erörterten Tatbestand, daß die AKF eines $MA(q)$ -Prozesses endlich ist, d. h. beim *lag* $k=q$ abbricht, während sie für einen AR-Prozeß unendlich ist. Allerdings ist der Übergang bei $k=q$ von einem unregelmäßigen MA-Muster zu einem exponentiell abfallenden AR-Muster in einer empirischen AKF kaum jemals mit Sicherheit erkennbar. Wir reproduzieren hier ein Schaubild aus Box u. Jenkins (1976, S. 78), das für unterschiedliche Kombinationen von Phi- und Theta-Gewichten eines ARMA(1,1)-Prozesses typische Muster der entsprechen AKF's und PAKF's aufzeigt (s. Abb. 14). Nützlich ist auch ein Schaubild (Abb. 15) aus gleicher Quelle (ebd., S. 77), das die zulässigen Parameterkonstellationen verdeutlicht.

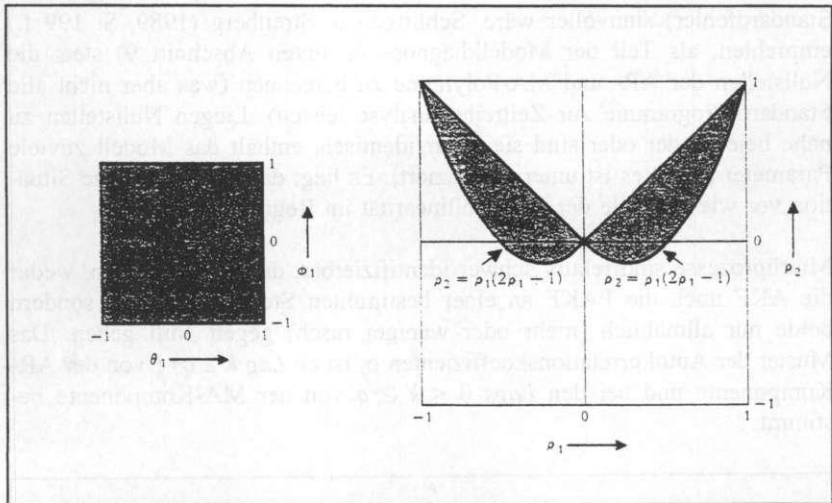


Abb. 15: Zulässige Parameterkonstellationen für ARMA(1,1)-Prozesse. Quelle: Box u. Jenkins 1976, S. 77

7. Identifikation eines Prozeßmodells anhand eines Beispiels

Box u. Jenkins (1976) schlagen vor, die Prozeßmodelle anhand der empirischen Autokorrelationsfunktionen (SAKF) und Partiellen Autokorrelationsfunktionen (SPAKF) zu identifizieren. Wir haben in den vorangegangenen Abschnitten wesentliche Eigenschaften der Struktur dieser Funktionen schon erläutert und in einigen Beispielen dargestellt. Aus Platzgründen müssen wir darauf verzichten, hier eine ausführlichere Sammlung von theoretischen und empirischen AKF's und PAKF's vorzulegen. Gute Zusammenstellungen finden sich beispielsweise in O. D. Anderson (1976) und Hoff (1983).

Bei der Interpretation der Autokorrelationskoeffizienten und der partiellen Autokorrelationskoeffizienten ist es nützlich, ihre Konfidenzintervalle zu kennen. Sie beruhen auf der Schätzung der entsprechenden Standardfehler. Unter der Voraussetzung, daß der Prozeß multivariat normalverteilt ist ("Gauß-Prozeß") und eine "große" Zahl von Meßzeitpunkten vorliegt, sind auch die empirischen Autokorrelationskoeffizienten r_i (als Schätzer der theoretischen Autokorrelationskoeffizienten ρ_k) normalverteilt um den Mittelwert ρ_k mit folgender Varianz (Bartlett-Formel):

$$Var(r_k) = \frac{1}{n} \sum_{i=-\infty}^{\infty} (\rho_i^2 + \rho_{i+k}\rho_{i-k} - 4\rho_k\rho_i\rho_{i-k} + 2\rho_k^2\rho_i^2) \quad (68)$$

Die Computerprogramme stützen sich bei der Berechnung der Varianz - und somit auch der Standardfehler $\sigma(r_k)$ - in der Regel auf eine Näherungsformel, die ebenfalls von Bartlett entwickelt wurde. Wir geben sie hier in einer Form wieder, in der die theoretischen Größen durch ihre Stichprobenschätzer ersetzt sind:

$$s(r_k) = \sqrt{\frac{1}{n}(1 + 2r_1^2 + \dots + 2r_m^2)} \quad , \quad k > m \quad (69)$$

Sie bezieht sich auf die Nullhypothese, daß $\rho_k = 0$ für $k > m$. Man weiß natürlich nicht im voraus, ab welchem Lag die wahren Autokorrelationskoeffizienten ρ_k gleich Null sind.⁴ Deshalb berechnet man die Standardfehler sukzessive für $m = 0, 1, 2, 3, \dots$. Das heißt die Varianzen für r_1, r_2, \dots, r_m werden nacheinander mit $n^{-1}, n^{-1}(1+2r_1^2), n^{-1}(1+2r_1^2+\dots+2r_{k-1}^2)$ geschätzt (s. Mills 1990 S.66). Die Schätzung der entsprechenden Konfidenzintervalle ist bei AR-Prozessen irrelevant; sie kann aber bei der Bestimmung der Ordnung eines MA(q)-Prozesses hilfreich sein. Die Konfidenzintervalle werden mit zunehmendem Lag breiter, eine Entwicklung, die in den Computerplots nicht immer sichtbar wird.⁵ Außerdem ist zu beachten, daß Formel (69) nur für große Lags eine gute Annäherung an den wahren Werte liefert (man spricht von der *large lag variance*). Dies ist vor allem beim Test auf *white noise* zu bedenken. Beim Test der Hypothese, daß $\rho_k = 0$ für alle $k = 1, 2, \dots$, reduziert sich Gleichung (69) zu

4 Box u. Jenkins (1976, S. 35) interpretieren den Lag $k = m$ als jenen Punkt, "beyond which the theoretical autocorrelation function may be deemed to have 'died out'".

5 SPSS/PC+ TREND^S berechnet die Standardfehler nach Formel (69) nur, wenn man die Voreinstellung durch den Befehl TSET ACFSE = MA ändert.

$$s(r_k) = \frac{1}{\sqrt{n}} \quad (70)$$

(Wei 1990, S. 21). Gleichung (70) liefert auch den Standardfehler für den partiellen Autokorrelationskoeffizienten,

$$s(\hat{\phi}_{k,k}) \approx \frac{1}{\sqrt{n}}, \quad k > p \quad (71)$$

unter der Hypothese, daß ein autoregressiver Prozeß *p-ter* Ordnung vorliegt (s. Box u. Jenkins 1976, S. 178).

Der Vorgang einer Modellidentifikation soll nun anhand eines konkreten Beispiels demonstriert werden.⁶ Abb. 16 zeigt die trendbereinigte Zeitreihe der Realinvestitionen im Vereinigten Königreich in Mrd £ von 1830 - 1979.⁷

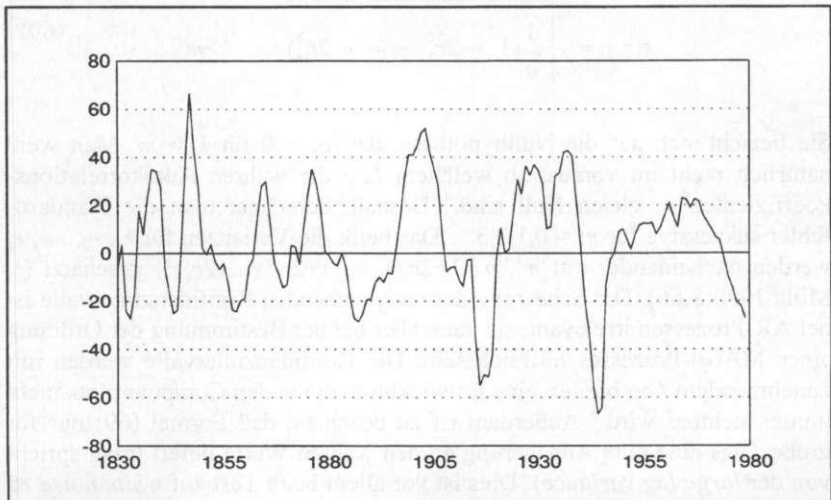


Abb. 16: Trendbereinigte Realinvestitionen im Vereinigten Königri. in Mrd £ von 1830 - 1979

- 6 Das Beispiel ist **einem** Aufsatz von Metz u. Stier (1992) entnommen. Ich danke Rainer Metz, der mir diese Daten zur Verfügung gestellt hat.
- 7 Das Verfahren der **Trendbereinigung** mittels **linearer Filter** ist in Metz u. Stier (1992) beschrieben. Es wurden alle Schwingungskomponenten mit einer Periodenlänge von über 80 Jahren aus den Originaldaten eliminiert. Zum Zwecke der Varianzstabilisierung wurden die **prozentierten Abweichungen vom Trend** als Basis für die Modellidentifikation verwendet.

Sie scheint eine zyklische Komponente zu enthalten. Dies wird durch die empirische Autokorrelationsfunktion bestätigt, wie die folgende Abb. 17a zeigt. In ihr sind die 95-Prozent-Konfidenzintervalle eingezeichnet. Die Autokorrelationskoeffizienten entwickeln sich in Form einer gedämpften, sinusähnlichen Schwingung. Zeitreihen-Werte, die etwa 26 bis 29 Lags (Jahre) auseinanderliegen, korrelieren relativ stark positiv miteinander; Werte, die etwa 14, 15 Jahre auseinanderliegen, korrelieren relativ stark negativ miteinander. Die partielle Autokorrelationsfunktion (SPAKF) ist in Abb. 17 b wiedergegeben. Sie zeigt signifikante partielle Autokorrelationskoeffizienten für die beiden ersten Lags und fällt danach abrupt ab, hat bei größeren Lags nur noch einzelne Koeffizienten, die größer sind als $2\sqrt{n}$.

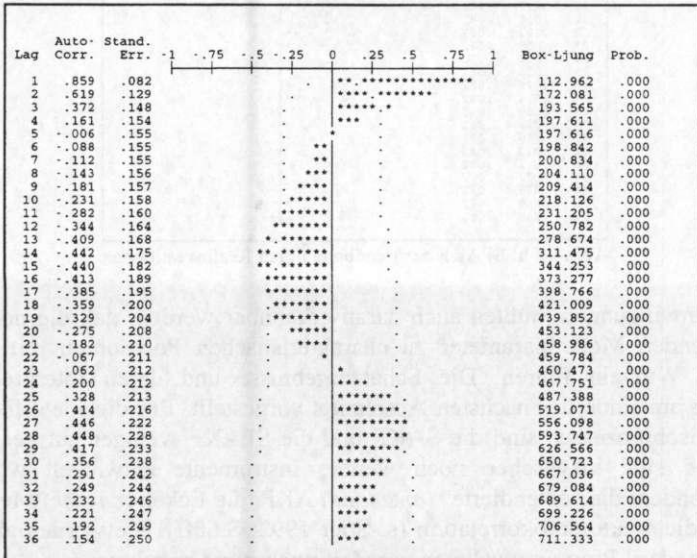


Abb. 17 a: SAKF der trendbereinigten Realinvestitionen

Die Muster von SAKF und SPAKF sprechen also für einen autoregressiven Prozeß zweiter Ordnung mit pseudo-periodischen Schwingungen, die sich über eine Basis-Periode von 26 bis 30 Jahren erstrecken.⁸

8 Zu beachten ist, daß in der Praxis (bei kurzen Zeitreihen), die SPAKF auch häufig dann keinen "signifikanten" Koeffizienten bei $k=2$ ausweist, wenn tatsächlich eine stochastische Schwingung vorliegt (s. Box u. Jenkins 1976, S. 66). Andererseits kann die SAKF ein schwingungsähnliches Muster annehmen, wenn der Prozeß keine relevante zyklische Komponente enthält. Für die Identifikation von Schwingungen stehen noch andere, sog. frequenzanalytische Verfahren zur Verfügung, die hier nicht behandelt werden können.

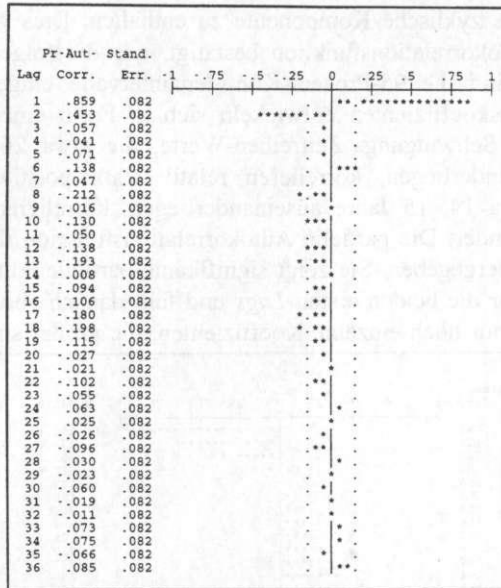


Abb. 17 b: SPAKF der trendbereinigten Realinvestitionen

Die Schwingungen müßten auch daran erkennbar werden, daß die noch zu schätzenden Modellparameter zu charakteristischen Polynomen mit komplexen Wurzeln führen. Die Schätzergebnisse und deren Interpretation werden am Ende des nächsten Abschnitts vorgestellt. Für die Identifikation von Mischprozessen sind die SAKF und die SPAKF weniger gut geeignet. Deshalb sind inzwischen noch weitere Instrumente entwickelt worden, insbesondere die "extended" (*extended*) AKF, die Ecken (*corner*) Methode sowie die Vektorautokorrelation (s. Stier 1992, S.68ff). Sie werden aber in den Standard-Programmpaketen zur Zeit noch nicht angeboten.

8. Hinweise zu den Schätzverfahren und Fortsetzung des Beispiels

Bei der Zeitreihenanalyse sind einerseits die Momentfunktionen (Erwartungswerte, Varianzen, Auto-Kovarianzen, Auto-Korrelationen) und andererseits die Prozeßparameter (Phi- und Theta-Gewichte) zu schätzen. Mit der Schätzung der Momentfunktionen wollen wir uns in diesem Abschnitt nicht weiter beschäftigen. Sie wirft keine nennenswerten praktischen Probleme auf, wenn auch ihre theoretische Begründung (auf der Basis der eingangs erwähnten Ergodizitätsannahme) angesichts der fehlenden Unabhängigkeit der einzelnen

Realisationen der Zufallsvariable recht kompliziert ist.⁹

Für die Prozeßparameter lassen sich sog. *Momentenschätzer* angeben, da, wie wir (beispielsweise bei den Yule-Walker-Gleichungen) sahen, unter den Stationaritäts- bzw. Invertibilitätsvoraussetzungen eindeutige Beziehungen zwischen den theoretischen Momenten und den Parametern bestehen und die theoretischen durch die entsprechenden empirischen Momente ersetzt werden können. Für MA- und ARMA-Modelle ist diese Schätzmethode jedoch nicht empfehlenswert; sie ist zu kompliziert und führt zu erheblichen numerischen Problemen: "(Die Momentenschätzer) sind empfindlich gegenüber Rundungsfehlern und sollten nicht verwendet werden bei Prozessen, die nahe an der Grenze zur Nicht-Stationarität bzw. Nicht-Invertierbarkeit liegen, d. h. deren Polynomwurzeln dicht am Einheitskreis liegen" (Stier 1992, S. 57). Eine bessere Alternative stellen die *Maximum-Likelihood-Verfahren* dar (ML-Schätzung). Die "exakte" ML-Schätzung ist jedoch außerordentlich kompliziert und rechenaufwendig (s. Stier 1992, S. 57 ff.; Schlittgen u. Streitberg 1989, S. 181 ff.) und wird nicht von allen Programmpaketen zur Zeitreihenanalyse angeboten. Statt dessen wird in der Praxis häufig eine nicht-lineare Kleinst-Quadrate-Schätzung (KQS) vorgenommen, die die ML-Schätzung um so besser approximiert, je länger die Zeitreihe ist.¹⁰

Aus der gewöhnlichen Regressionsrechnung kennen wir das Verfahren, die Regressionskoeffizienten über die Minimierung der Fehlerquadratsumme zu schätzen. Bei linearen Modellen ("linear in den Parametern") ist das auf rein analytischem Wege mit Hilfe der Differentialrechnung möglich, die zu den sog. "Normalgleichungen" führt. Bei der Schätzung der Parameter stochastischer Prozesse läßt sich das Prinzip der Minimierung der Fehlerquadratsumme ($\sum_{t=1,n} a_t^2 = \min.$) beibehalten; es läßt sich aber nicht mehr in jedem Falle rein analytisch realisieren. Die Schätzgleichungen sind im allgemeinen nicht linear; zu ihrer Lösung bedarf es numerischer Verfahren, mit denen die Parameter "iterativ", Schritt für Schritt, bestimmt werden. Wir wollen die verschiedenen Varianten dieses Verfahrens (häufig verwendet man den sog.

-
- 9 Einschränkung ist allerdings darauf hinzuweisen, daß die **Schätzung** des Erwartungswertes $E(Z_t) = \mu$ durch das arithmetische Mittel der beobachteten Werte nicht die beste Schätzung ist, wenn Autokorrelation vorliegt. In diesem Falle ist es besser, μ als Modellparameter zu behandeln, der simultan mit den anderen Modellparametern nach dem *maximum likelihood* Prinzip zu schätzen ist (Schlittgen u. Streitberg 1989, S. 185).
 - 10 Die **gewöhnliche Kleinst-Quadrate-Schätzung** kommt nur bei reinen AR-Prozessen, nicht aber bei MA- oder Mischprozessen in Frage (s. Wei 1990, S. 147), wie man z.B. an Gleichung (47) unmittelbar sieht.

"Marquardt"-Algorithmus) nicht näher erläutern", sondern nur einige Punkte erörtern, bei denen der Anwender Entscheidungen treffen muß.

Verschiedene Computer-Programmsysteme erlauben eine Wahl zwischen der Methode der "Unconditional Least Squares" (ULS) und der "Conditional Least Squares" (CLS). Diese Alternative betrifft die Art und Weise, wie die Anfangswerte z_t , ($t=1,2,\dots,m$, $m < n$) einer Zeitreihe Z_t , ($t=1,2,\dots,n$) in den Schätzvorgang eingeführt werden. Das Problem läßt sich am Beispiel eines AR(1)-Prozesses verdeutlichen (wobei wir wiederum von mittelwertbereinigten Daten ausgehen):

$$z_t = \phi_1 z_{t-1} + a_t \quad (72)$$

Nehmen wir an, ein erster Schätzer $\hat{\phi}_1$ sei gefunden worden und die Summe der Fehlerquadrate sei zu ermitteln. Die geschätzten Fehler \hat{a}_t erhält man durch

$$\hat{a}_t = z_t - \hat{\phi}_1 z_{t-1} \quad (73)$$

Für das erste Residuum zum Zeitpunkt $t=1$, $\hat{a}_1 = z_1 - \hat{\phi}_1 z_0$, ist aber der Wert z_0 nicht bekannt (er wurde nicht beobachtet). Erst für \hat{a}_2 (bei Prozessen höher Ordnung erst für \hat{a}_3 , \hat{a}_4 usw.) sind die zur Fehlerschätzung benötigten Z -Werte bekannt:

$$\hat{a}_2 = z_2 - \hat{\phi}_1 z_1 \quad (74)$$

Ein ähnliches Problem entsteht bei MA- und bei Mischprozessen. Betrachten wir noch den Fall eines (mittelwertbereinigten) MA(1)-Prozesses:

$$\begin{aligned} z_t &= a_t - \theta a_{t-1} \\ \hat{a}_t &= z_t + \hat{\theta} a_{t-1} \\ \hat{a}_1 &= z_1 + \hat{\theta} a_0 \end{aligned} \quad (75)$$

Hier ist a_0 nicht bekannt. Generell fehlen in unseren Beobachtungen die zur Fehlerschätzung eigentlich benötigten Größen $z_0, z_{-1}, \dots, z_{-n}$, sowie a_0, a_{-1}, \dots ,

11 Eine recht anschauliche Darstellung des Verfahrens geben Pindyck u. Rubinfeld (1981, S. 539 ff.) oder Pankratz (1983, S. 192 ff., 209 ff.). Im Prinzip werden nach einem bestimmten Suchprinzip verschiedene Kombinationen von Parameterwerten "ausprobiert", bis man diejenige Kombination gefunden hat, die die minimale Fehlerquadratsumme liefert.

a_{t-g+1} (Pindyck u. Rubinfeld 1981, S. 540).¹² Eine Möglichkeit, dieses Problem zu lösen, besteht darin, für die fehlenden Größen deren nicht-bedingte Erwartungswerte einzusetzen: $E(a_t) = E(a_{t-m}) = 0$, $m=1,2,\dots$ und $E(z_t) = 0$ bei mittelwertbereinigten Daten.¹³ Die weiteren Werte lassen sich dann rekursiv mit Hilfe der entsprechenden Modellgleichungen ermitteln. Dieses Verfahren läuft im allgemeinen unter der Bezeichnung "Conditional Least Squares" (CLS) Methode. Diese Bezeichnung mag zunächst mißverständlich klingen, da die unbeobachteten Vergangenheitswerte durch ihre **nicht**-bedingten, von $\{z_t, t=1,2,\dots,n\}$ unabhängigen Erwartungswerte ersetzt werden. Dadurch wird aber die Kleinstquadratfunktion eine "bedingte" Funktion, nämlich abhängig von nicht beobachteten Werten, die unabhängig von den beobachteten Werten substituiert werden.

Der alternative Verfahrensvorschlag knüpft an den Tatbestand an, daß bei stationären Prozessen zwischen den beobachteten Werten $\{z_1, z_2, \dots, z_n\}$ und den zukünftigen Werten $\{z_{n+m+1}; m=0,1,2,\dots\}$ die gleiche Wahrscheinlichkeitsbeziehung besteht wie zwischen der Menge der beobachteten Werte und der Menge der unbeobachteten Werte der Vergangenheit $\{z_{-m}; m=0,1,2,\dots\}$ (s. Box u. Jenkins 1976, S. 199 f.). Folglich kann man auf der Basis der beobachteten Zeitreihenwerte und vorläufiger Schätzgrößen für die Modellparameter (zur Wahl der Anfangsschätzer siehe unten) die fehlenden Werte z_0, z_{-1}, \dots bzw. a_0, a_{-1}, \dots substituieren. Es werden so viele Werte z_0, z_{-1}, \dots "rückwärts prognostiziert" bis sie sich Null nähern. (Bei stationären Prozessen müssen sich die Voraussagen dem Erwartungswert der Reihe nähern). In der Box/Jenkins-Literatur wird dieser Vorgang als "backforecasting" oder "backcasting" bezeichnet. Die Fehlerquadratsumme wird auf diese Weise zu einer "Unconditional Least Squares Function" (ULS). Mit Hilfe der rückwärts prognostizierten Werte lassen sich die Modellparameter erneut schätzen. Mit ihnen kann man einen neuen Satz von "backward forecasts" errechnen usw. bis ein Konvergenzkriterium erfüllt ist (s. Pindyck u. Rubinfeld 1981, S. 553). Bei langen Reihen, die keine Saisonkomponente enthalten, ist der Unterschied zwischen ULS und CLS-Schätzern relativ gering; andernfalls ist das ULS-Verfahren dem CLS-Verfahren im allgemeinen vorzuziehen (s. Stier 1992, S. 61).¹⁴

-
- 12 Ein stationärer ARMA(p,q)-Prozeß kann in guter Näherung als ein MA(r)-Prozeß ($r>q$) dargestellt werden (s. Sehlingen u. Streitberg 1984, S. 193 f.)
 - 13 Bei nicht mittelwertbereinigten Zeitreihen wird das aus den beobachteten Werten errechnete arithmetische Mittel als (geschätzter) Erwartungswert eingesetzt oder man schätzt μ simultan mit den anderen Modellparametern (Box u. Jenkins 1976, S. 210 f.).
 - 14 Wieso Mills (1990, S. 119) zu der Aussage kommt, ULS scheine (im Gegensatz zu CLS) von keinem Autor empfohlen zu werden, ist mir unerfindlich. Immerhin favorisieren Box u. Jenkins (1976, S. 211 ff.) die ULS- gegenüber der CLS-Methode.

Einige Computer-Programmsysteme (wie z. B. SCA, siehe Liu u. Hudack 1992) bieten exakte Maximum-Likelihood-Schätzer an. Nach Schlittgen u. Streitberg (1989, S. 193) sollten sie in folgenden Situationen verwandt werden a) bei kurzen Reihen, b) bei Modellen hoher Ordnung, c) bei saisonalen Modellen, d) bei Modellen, "die am Rand des Stationaritäts- oder Invertierbarkeitsbereichs liegen".¹⁵

Iterative Schätzverfahren benötigen Vorgaben, Startwerte für die Schätzer, mit denen eine erste Fehlerquadratsumme ausgerechnet werden kann. Bei AR(1)- oder MA(1)-Prozessen erhält man diese Startwerte, wie gezeigt, unmittelbar aus der empirischen Autokorrelationsfunktion. Für Prozesse 2. Ordnung bzw. ARMA(1,1)-Modelle liefern Box u. Jenkins (1976, S. 518 - 520) Graphiken, aus denen sich Startwerte bei gegebenen Autokorrelationskoeffizienten ablesen lassen. In der Praxis verläßt man sich meist auf die Voreinstellungen des benutzten Programm-Systems und vertraut auf die Güte des dort implementierten Schätzalgorithmus. Es empfiehlt sich, Schätzungen mit verschiedenen Startwerten vorzunehmen; sie sollten bei konstant gehaltenem Konvergenzkriterium zu den gleichen Endschätzern führen. Einige Programmsysteme bieten eine GRID-Option an, mit deren Hilfe man überprüfen kann, ob die Endschätzer (innerhalb der näheren Umgebung) auch tatsächlich die niedrigste Fehlerquadratsumme geliefert haben.

Die Algorithmen erlauben auch eine approximative Schätzung der Standardfehler, so daß *t*-Werte berechnet und Signifikanztests durchgeführt werden können (Pankratz 1983, S.202).

Die Parameter des im vorigen Abschnitt identifizierten AR(2)-Modells wurden sowohl mit einem exakten ML-Verfahren als auch mit dem CLS-Verfahren geschätzt. Tabelle 1 bringt die Ergebnisse (Standardfehler in Klammern).

Tab. 1:
Parameterschätzungen des AR(2)-Modells

	ML	CLS
$\hat{\phi}_1$	1.280 (0.072)	1.283 (0.073)
$\hat{\phi}_2$	-.474 (0.072)	-.475 (0.073)

15 Es gilt ganz allgemein, daß "bei 'fast instationären' Prozessen häufig ein radikal verändertes Verhalten der Statistiken zutage tritt" (Schlittgen u. Streitberg 1989, S. 185).

In diesem Fall sind die ML-Schätzer und CLS-Schätzer praktisch identisch.¹⁶

Im nächsten Abschnitt werden verschiedene Kriterien vorgestellt, mit denen man die Güte dieses spezifischen AR(2)-Modells evaluieren kann. Auch ohne formale Berechnung können wir Abb. 7 entnehmen, daß diese Parameterschätzer zu Komplexwurzeln führen, daß der Prozeß also pseudo-periodische Schwingungen aufweisen muß.

Box u. Jenkins (1976, S. 58ff.) erläutern in ihrem Standardwerk wie die Länge der Basis-Periode aus den Wurzeln der charakteristischen Gleichung bzw. direkt aus den Phi-Gewichten berechnet werden kann. Dabei ist allerdings zu beachten, daß die geschätzten Parameter nicht identisch sind mit den wahren Prozeßparametern. Dies kann u. U. zu erheblichen Verzerrungen in der Bestimmung der Periodenlänge P führen. Dies gilt auch für unser Beispiel. Der entsprechende Rechengang läßt sich in unterschiedlichen "Formeln" zusammenfassen, z. B. (s. Box u. Jenkins 1976, S. 60, 63)

$$P = \frac{2\pi}{\cos^{-1}\left(\frac{|\phi_1|}{2\sqrt{-\phi_2}}\right)} = \frac{6,28}{\cos^{-1}\left(\frac{1,28}{2\sqrt{+0,474}}\right)} \quad (76)$$

$$= 16.6$$

Das Ergebnis weicht also erheblich von dem visuellen Eindruck ab, den die Reihe selbst vermittelt (s. oben Abb. 16). Dieser Eindruck stimmt eher mit dem Muster des Autokorrelogramms überein, das eine Periodenlänge zwischen 25 und 30 Jahren nahelegt. Das Problem liegt nicht in einer verzerrten Parameterschätzung begründet, sondern darin, daß in bestimmten Wertebereichen schon minimale Veränderungen in den Parametern erhebliche Veränderungen in der Periodenlängen implizieren. Wenn statt der in Tabelle 1 wiedergegebenen Schätzer die Koeffizienten $\phi_1^* = 1.311$ und $\phi_2^* = -0.4522$ in Gleichung (76) eingesetzt werden, ergibt sich eine um ca. 10 Jahre vergrößerte Periodenlänge von 27.9 Jahren. Wir können diese Problematik, die Box u. Jenkins (1976) in ihrem Standardwerk nicht erwähnen, hier nicht

16 Abb. 16 läßt vermuten, daß die Reihe einige Ausreißer enthält. Die Modellierung von Ausreißern ist in diesem Artikel nicht zu behandeln (s. hierzu Thome 1992 b). Die Analyse der ausreißerbereinigten Zeitreihe führt in diesem Falle nicht zu nennenswerten anderen Ergebnissen, weder im Muster der Autokorrelogramme noch in der Schätzung der Parameter.

vertiefen. Als Fazit bleibt, daß man sich bei der Bestimmung der Basis-Periode nicht allein auf die geschätzten Parameter eines autoregressiven Prozesses stützen sollte, auch dann nicht, wenn die Schätzung selbst, gemessen an den üblichen Kriterien (s. nächsten Abschnitt), nicht zu beanstanden ist. Zumindest sollte man sog. "frequenzanalytische" Verfahren (wie z. B. die Periodogramm-Analyse) ebenfalls einsetzen (deren Zuverlässigkeit allerdings auch an bestimmte Voraussetzungen gebunden ist).

9. Modellevaluierung ("Diagnose")

Box u. Jenkins (1976) schlagen für die Konstruktion stochastischer Modelle ein dreistufiges Verfahren vor: Identifikation, Schätzung und Diagnose. Da die Modellannahmen in der Realität nur zum Teil erfüllt sind und normalerweise nur wenige Realisationen (beobachtete Zeitreihenwerte) vorliegen, sind die Erkennungsinstrumente der Modellidentifikation unscharf und die Schätzung der Parameter kann verzerrt sein. Box und Jenkins sowie einige andere Autoren haben deshalb Kriterien entwickelt, nach denen die Güte eines geschätzten Modells evaluiert werden kann. Diese "diagnostic checks" bieten zwar keine Garantie, daß das "wahre" Modell (also der tatsächliche Erzeugungsprozeß) ermittelt wird, aber sie können helfen, bestimmte Mängel zu erkennen und möglicherweise ein besseres Modell zu finden. Das heißt, die Modellbildung kann die oben erwähnten drei Stufen in Form einer Schleife mehrmals durchlaufen - u. U. mit dem Ergebnis, daß überhaupt kein adäquates ARMA-Modell für die vorliegende Zeitreihe konstruierbar ist. Ich werde hier nur die gebräuchlichsten Prüfkriterien vorstellen; zunächst solche, die die Qualität der Parameterschätzer ins Visier nehmen, sodann jene, die sich auf Eigenschaften der Residuen beziehen.

Die Standard-Computerprogramme liefern mit den Parameterschätzern eine Reihe weiterer Informationen, die für eine Modellevaluierung nützlich sind. Dazu gehört der f -Wert, also der Quotient des Schätzers und seines geschätzten Standardfehlers. Mit ihm läßt sich analog zur Regressionsanalyse mit Querschnittsdaten die statistische Signifikanz der einzelnen Schätzer ermitteln (bezogen auf die Nullhypothese, daß der wahre Parameterwert Null sei). Wenn das 5-Prozent-Signifikanzniveau erreicht werden soll, muß der Parameterschätzer mindestens doppelt so groß sein wie sein Standardfehler, also $t \geq 2$. Allerdings akzeptiert man in der Zeitreihenanalyse gelegentlich auch einen Schätzer, der dieses Kriterium nicht erfüllt, dann nämlich, wenn anders die Modellresiduen nicht in "weißes Rauschen" überführt werden können (s. unten). In unserem Analysebeispiele bestehen in dieser Hinsicht keine Probleme; die geschätzten Phi-Parameter liegen erheblich über der konventionellen Signifikanzgrenze (s. oben, Tab. 1).

Die Schätzalgorithmen liefern nicht nur die Standardfehler, sondern die gesamte Varianz/Kovarianz-Matrix der Schätzer und, daraus abgeleitet, die Korrelation der Schätzer untereinander. Wenn diese Korrelation zu hoch ist, sind die Schätzer sehr instabil, d. h. sie schwanken von Stichprobe zu Stichprobe, können sich also selbst bei konstantem Erzeugungsvorgang erheblich ändern, wenn die Zeitreihe durch neugewonnene Beobachtungen verlängert wird. Die Effekte ähneln denen, die aus der Multikollinearität der Prädiktoren in der Regressionsanalyse erwachsen. Allerdings ist die Gefahrengrenze in der Zeitreihenanalyse höher anzusetzen als bei der Regressionsanalyse mit Querschnittsdaten. Pankratz (1983, S. 203) nennt als Daumenregel einen Korrelationskoeffizienten von $r > 0,9$. In unserem Beispiel haben wir es mit einer Parameterkorrelation von immerhin $r = -0,87$ zu tun. Das könnte für einen Prozeß mit stochastischen Schwingungen schon ein Problem bedeuten, da, wie wir sahen, geringfügige Abweichungen in der Größenrelation der beiden Phi-Parameter schon zu erheblich veränderten Periodenlängen führen kann. Die Zahl der Parameter zu reduzieren wäre aber in unserem Beispiel nicht sinnvoll, da das Modell dann überhaupt keine zyklische Struktur mehr abbilden könnte.

Instabile Schätzungen können sich auch ergeben, wenn die Parameter nahe an den Stationaritäts- bzw. Invertibilitätsgrenzen liegen. In unserem Beispiel sind nur die Stationaritätsvoraussetzungen zu überprüfen, da das Modell keine MA-Komponente enthält. Setzt man die Parameterwerte aus Tab. 1 in die Ungleichungen (28) ein, sieht man, daß die Stationaritätsgrenzen nicht berührt werden. Zu Demonstrationszwecken sollen hier auch die beiden Wurzeln des AR(2)-Prozesses ermittelt werden. Die charakteristische Gleichung ist $(1 - \phi_1 B - \phi_2 B^2) = 0$. Die Lösungsformel für eine solche quadratische Gleichung ist

$$\begin{aligned}
 (1 - \phi_1 B - \phi_2 B^2) &= 0 \\
 B_{1,2} &= \frac{-\phi_1 \pm \sqrt{\phi_1^2 + 4\phi_2}}{2\phi_2} \quad (77) \\
 &= 1.34985 \pm 0.53547i
 \end{aligned}$$

wobei $i = \sqrt{-1}$. Da die Wurzel negativ ist, bilden die beiden Lösungen in B konjugiert komplexe Zahlen mit der reellen Komponente $a = 1.34985$ und der komplexen Komponente $b = \pm 0.53547$. Der Absolutbetrag dieser komplexen Zahl ist deutlich größer als 1.

$$\begin{aligned}
 r &= \sqrt{a^2 + b^2} \quad (78) \\
 &= 1.45218
 \end{aligned}$$

Auch hieraus läßt sich übrigens die (geschätzte) Periodenlänge ermitteln mit Hilfe des $\arccos(a/r) = \arccos(1.34985/1.45218) = 21.638$. Daraus ergibt sich mit $360/21.638 = 16.6$ die gleiche Periodenlänge, wie sie oben bereits errechnet wurde.

Damit kommen wir zur zweiten Gruppe der Prüfkriterien, die sich mit Eigenschaften der Modellresiduen beschäftigen. Wenn das Modell korrekt identifiziert und geschätzt worden ist, sollten die Residuen einen *white noise* Prozeß repräsentieren; sie sollten also nicht miteinander korrelieren und keinerlei Systematik in ihrem zeitlichen Verlauf aufweisen. Andernfalls enthielten sie noch Informationen, die nicht in die Parameterschätzer eingegangen wären; das Modell würde dann z. B. keine optimalen Prognosen liefern. Die Residuen unseres AR(2)-Modells für die Realinvestitionen des Vereinigten Königreichs sind in Abb. (18) wiedergegeben.

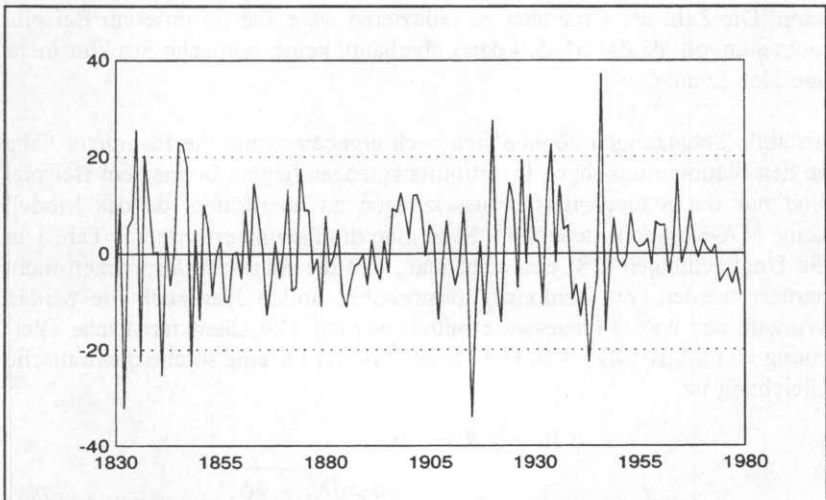


Abb. 18: Residuen des mit SPSS/PC+ geschätzten AR(2)-Modells für die Realinvestitionen des Vereinigten Königreichs

Die Berechnung der Residuen soll hier nicht allgemein erläutert werden, da sie in allen Standardprogrammen routinemäßig geliefert werden. Im Falle eines AR-Prozesses ist die Berechnung allerdings so einfach, daß sie hier ohne weiteres angegeben werden kann:

$$z_t - \hat{\phi}_1 z_{t-1} - \hat{\phi}_2 z_{t-2} = \hat{a}_t \quad (79)$$

Zu beachten ist, daß es sich dabei um Schätzgrößen handelt (deshalb das "Hütchen" über dem Buchstaben a), deren Eigenschaften mit denen eines

idealen *white noise* Prozesses nicht identisch sind. Die Autokorrelationen können nicht alle exakt gleich Null sein; die Abweichungen von dieser Nullhypothese sind also statistisch zu testen.

Unter der Annahme, daß das wahre Modell (mit den wahren Parametern und damit den wahren Residuen) bekannt ist, sind die geschätzten Autokorrelationen $r_{\hat{a}}(a)$ einer Menge von n a 's unkorreliert und mit einem Standardfehler von $1/\sqrt{n}$ normalverteilt um den Mittelwert Null (siehe oben, Gleichung (70)). Dieser Standardfehler ist aber nur eine grobe Approximation an den Standardfehler der Autokorrelationskoeffizienten $r_{\hat{a}}(\hat{a})$ der geschätzten Modellresiduen (Box u. Jenkins 1976: 290). Vor allem für die ersten beiden Lags können die wahren Standardfehler erheblich geringer sein, d. h., die $r_{\hat{a}}(\hat{a})$'s können signifikant sein, auch wenn sie kleiner sind als $1,96n^{-1/2}$. Pankratz (1983, S. 228) schlägt vor, statt eines t -Wertes von 1.96 für die Lags 1, 2 "and perhaps 3" einen t -Wert von 1,15 als Signifikanzgrenze anzusetzen. Das Autokorrelogramm (s. Abb. 19) kann also irreführend sein, wenn man sich an den Konfidenzgrenzen orientiert, die dort üblicherweise auf der Basis der kumulativen Bartlett-Formel (s. Gleichung (69)) eingezeichnet sind.

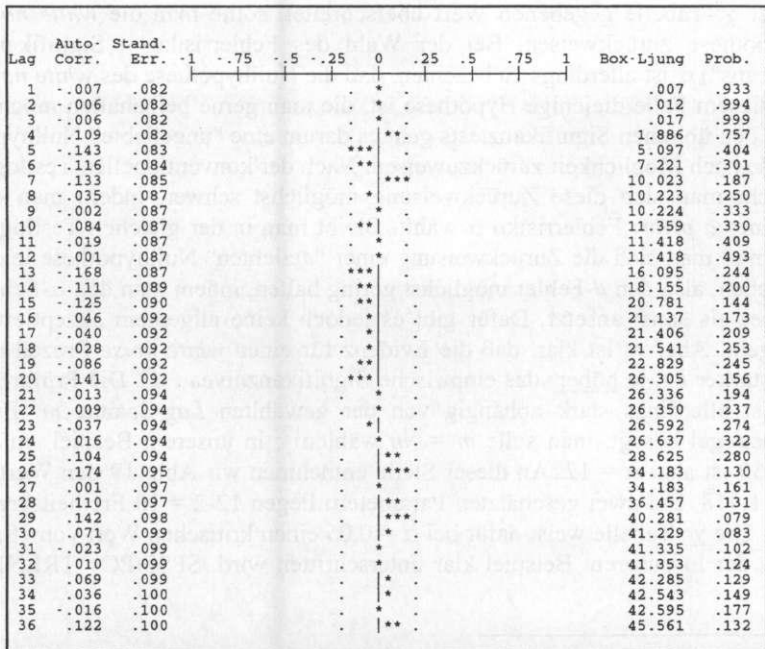


Abb. 19: AKF der Residuen des AR(2)-Modells für die Realinvestitionen

In unserem Beispiel ergeben sich aber keinerlei Probleme, da die ersten drei Koeffizienten noch erheblich unterhalb des Betrags von einem Standardfehler liegen. Auch läßt das Muster der AKF im weiteren Verlauf keine Systematik erkennen, so daß hier eine starke Evidenz für *white noise* Residuen vorliegt. Einzelne Autokorrelationskoeffizienten, die die Signifikanzgrenze überschreiten (außerhalb des Konfidenzintervalls liegen) sind tolerierbar.

Vom statistischen Standpunkt aus ist es allerdings problematisch, eine Vielzahl von Koeffizienten nacheinander an den gleichen Daten zu testen. Ljung und Box haben deshalb (im Anschluß an einen Vorschlag von Box und Pierce) eine Prüfstatistik Q entwickelt, mit der sich eine Gesamtheit von Auto-Korrelationskoeffizienten gegen die Nullhypothese des *white noise* testen läßt (sog. *Portmanteau Test*):

$$Q(m) = n(n + 2) \sum_{k=1}^m \frac{r_k^2}{n-k} \quad (80)$$

SPSS/PC+ TRENDS liefert diese Prüfstatistik automatisch mit dem Autokorrelogramm (s. Abb. 19). Sie ist χ^2 -verteilt mit $m-q-p$ Freiheitsgraden. Falls der gegebene Q -Wert bei einem bestimmten Signifikanzniveau α den in einer χ^2 -Tabelle gegebenen Wert überschreitet, sollte man die *white noise* Hypothese zurückweisen. Bei der Wahl des Fehlerrisikos ("Signifikanzniveaus") α ist allerdings zu beachten, daß die Nullhypothese des *white noise* in diesem Falle diejenige Hypothese ist, die man gerne beibehalten möchte. Bei den üblichen Signifikanztests geht es darum eine "ungeliebte" Nullhypothese nach Möglichkeit zurückzuweisen. Nach der konventionellen Testlogik macht man sich diese Zurückweisung möglichst schwer, indem man ein kleines *a priori* Fehlerrisiko α wählt. Bleibt man in der gleichen Testlogik, so muß man sich die Zurückweisung einer "geliebten" Nullhypothese leicht machen, also den β -Fehler möglichst gering halten, indem man den α -Fehler höher als sonst ansetzt. Dafür gibt es jedoch keine allgemein akzeptierten Regeln. Aber es ist klar, daß die Evidenz für einen *white noise* Prozeß um so stärker ist, je höher das empirische Signifikanzniveau α^* . Die Prüfgröße Q ist allerdings stark abhängig von der gewählten Lag-Spanne m . Eine Faustregel besagt, man solle $m = \sqrt{\alpha}$ wählen¹⁷; in unserem Beispiel mit $n = 150$ ist also $m = 12$. An dieser Stelle entnehmen wir Abb. 19 den Wert $Q = 11.418$. Bei zwei geschätzten Parametern liegen $12-2 = 10$ Freiheitsgrade vor. Die χ^2 -Tabelle weist dafür bei $\alpha = 0.05$ einen kritischen Wert von 18,31 aus, der in unserem Beispiel klar unterschritten wird. SPSS/PC+ TRENDS

17 Box u. Jenkins (1976, S. 291) empfehlen, die Lag-Spanne so zu wählen, daß die Psi-Gewichte $\hat{\psi}_j$ für alle $j \geq m$ nahe bei Null liegen.

liefert neben dem Q-Wert auch automatisch die empirische Signifikanz; in unserem Beispiel ist sie $\alpha^* = 0,493$. Wir können also davon ausgehen, daß die (geschätzten) Residuen tatsächlich einen *white noise* Prozeß repräsentieren.¹⁸

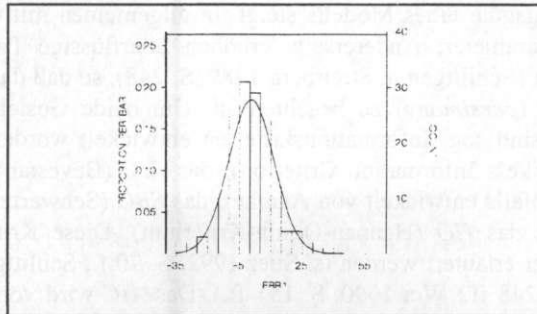


Abb. 20: Histogramm der Modellresiduen des AR(2)-Modells

Da die Parameter-Schätzungen und Test-Statistiken mit der Annahme normalverteilter Residuen arbeiten, sollte man auch diese Eigenschaft überprüfen. Üblicherweise führt man jedoch keine strengen Normalverteilungstests durch, sondern prüft lediglich anhand eines Histogramms, ob die Verteilung nicht allzu asymmetrisch ist und außerhalb des 95-Prozent-Intervalls nicht allzu viele (deutlich über 5 %) Fälle versammelt sind. In unserem Beispiel erhalten wir für die Modell-Residuen das in Abb. 20 präsentierte Histogramm. Abweichungen vom Normalverteilungsmodell sind durchaus erkennbar, aber die Asymmetrie ist nicht allzu ausgeprägt, und nur 8 % der Fälle liegen mehr als 2 Standardabweichungen vom Mittelwert entfernt.

Gelegentlich findet man sich mit dem Tatbestand konfrontiert, daß mehrere Modelle die genannten Prüfkriterien etwa gleich gut erfüllen. Als weiteres Wertungskriterium läßt sich in einem solche Falle auch noch die Anpassungsgüte der verschiedenen Modelle miteinander vergleichen. Ein Maß hierfür ist die Standardabweichung der geschätzten Residuen:

$$\hat{\sigma}_a = \sqrt{\frac{1}{n - m} \sum \hat{a}_i^2} \quad (81)$$

18 Hinzuweisen ist auf eine Aussage von Selhingen u. Streitberg (1989, S. 244), die die Ljung/Box-Statistik bei "kurzen" Reihen ($n < 100$) für "nicht empfehlenswert" halten. In der sozialwissenschaftlichen Forschungspraxis setzt man sich aber regelmäßig über diese Bedenken hinweg. Eine alternative Strategie ("Lagrange Multiplier Tests") wird von Mills (1990, S. 147 ff.) vorgestellt.

Der Buchstabe m steht dabei für die Anzahl der geschätzten Modell-Parameter. Man bezeichnet die Standardabweichung auch als "adjusted RMS" oder "adjusted RMSE" (*Root Mean Square Error*).

Die Anpassungsgüte eines Modells steigt im allgemeinen mit der Zahl der geschätzten Parameter. Andererseits erhöhen überflüssige Parameter die Prognosefehler (Schlittgen u. Streitberg 1989, S. 248), so daß das Prinzip der "Sparsamkeit" (*parsimony*) zu beachten ist. Um beide Gesichtspunkte zu kombinieren, sind sog. Informationskriterien entwickelt worden, vor allem das *AIC* (Akaike's Information Criterion), das *BIC* (Bayesian Information Criterion, ebenfalls entwickelt von Akaike), das *SBC* (Schwartz's-Bayesian-Criterion) und das *HQ* (Hannan-Quinn-Kriterium). Diese Kriterien sollen hier nicht näher erläutert werden (s. Stier 1992, S. 70 f.; Schlittgen u. Streitberg 1989, S. 248 ff.; Wei 1990, S. 151 ff.). Das *AIC* wird von Statistikern inzwischen nicht mehr als Kriterium der Modellselektion empfohlen. Die anderen eben genannten Kriterien sollen bei hinreichend großem n die wahre Modellordnung (p, q) ziemlich sicher finden lassen (Stier 1992, S. 71). Praktisch geht man dabei so vor, daß man dasjenige Modell bevorzugt, für das ein gewähltes Kriterium den niedrigsten Wert annimmt. SPSS/PC+TRENDS liefert das *AIC* und das *SBC*, die bei unseren Beispieldaten für unterschiedliche Modelle folgende Werte annehmen (s. Tab. 2).

Tab. 2
Informationskriterien verschiedener Modellschätzungen

	AIC	SBC
AR(2)	1143.91	1149.94
AR(1)	1180.20	1183.22
AR(3)	1145.92	1154.95
ARMA(2,1)	1145.92	1154.95

Nach beiden Kriterien schneidet das AR(2)-Modell (erwartungsgemäß) am besten ab. Gelegentlich ist nicht klar, ob das Modell über die gesamte Beobachtungsperiode die gleichen "wahren" Parameter aufweist. Aufgrund des Augenscheins ist durchaus denkbar, daß die Periodenlänge in der Wachstumsreihe der Realinvestitionen in der ersten Hälfte kürzer ist als in der zweiten (s. Abb. 16).

Bei ausreichend langen Zeitreihen kann man eine solche Vermutung überprüfen, indem man das Modell für beide Beobachtungsabschnitte getrennt schätzt. Wenn wir in unserem Beispiel das Jahr 1912 als Ende des 1. Ab-

Schnittes festlegen, erhalten wir folgende Parameterschätzer (Standardabweichungen in Klammern):

$$\begin{array}{l} \text{1. Abschnitt:} \\ \hat{\phi}_1 = 1.263 (0.095) \\ \hat{\phi}_2 = -.4929 (0.095) \end{array}$$

$$\begin{array}{l} \text{2. Abschnitt:} \\ \hat{\phi}_1 = 1.285 (0.109) \\ \hat{\phi}_2 = -.449 (0.109) \end{array}$$

Die Differenzen sind sowohl bei $\hat{\phi}_1$ als auch bei $\hat{\phi}_2$ so gering, daß sie statistisch nicht signifikant sind. Die Signifikanz läßt sich überprüfen, da die Varianz der Differenz gleich der Summe der beiden Einzelvarianzen (s. Tab. 1) ist. Somit erhalten wir z. B. für die Differenz von $\hat{\phi}_2$ minus $\hat{\phi}_1 = 0.022$ die Varianz $Var(\hat{\phi}_2 - \hat{\phi}_1) = 0.095^2 + 0.109^2 = 0.021$. Daraus ergibt sich durch Wurzelziehen der Standardfehler von 0.144, der wesentlich größer ist als die geschätzte Parameterdifferenz.

10. Nichtstationäre Prozesse: ARIMA-Modelle

Viele sozialwissenschaftlich relevante Zeitreihen sind nicht in dem Sinne stationär, wie wir das in den bisherigen Abschnitten vorausgesetzt haben. Ihr mittleres Niveau und/oder ihre Varianz ändern sich über Zeit. Die Nicht-Stationarität kann unterschiedliche Formen annehmen. Glücklicherweise sind nicht-stationäre Prozesse häufig in dem Sinne "homogen", daß sich ihr Verhalten in verschiedenen Perioden nur dadurch unterscheidet, daß es sich auf unterschiedlichen Niveaus (*levels*) und/oder mit lokal unterschiedlichen Trendneigungen (*slopes*) vollzieht. Die folgende Abbildung 21 gibt hierfür zwei Beispiele. Die erste Abbildung enthält eine Reihe mit zeitlich instabilem Niveau bei konstanter Trendneigung. Das Verhalten innerhalb der eingezeichneten Kästchen ist ansonsten gleich. Die untere Zeitreihe ist durch wechselnde Niveaus und wechselnde Trendneigungen gekennzeichnet. Könnte man die Unterschiedlichkeit der Niveaus und der Trendneigungen beseitigen, erhielte man offenkundig stationäre Zeitreihen, deren Verhaltensmuster über Zeit stabil wäre. Die Differenzenbildung, die schon im 1. Artikel dieser Serie behandelt wurde (s. Thome 1992 a), ist ein Mittel, genau dies zu erreichen. Somit können wir als vorläufige Definition der homogenen Nichtstationarität festhalten: Eine Zeitreihe wird als "homogen nicht-stationär" bezeichnet, wenn sie durch geeignete Differenzenbildung in eine stationäre Reihe überführt werden kann.

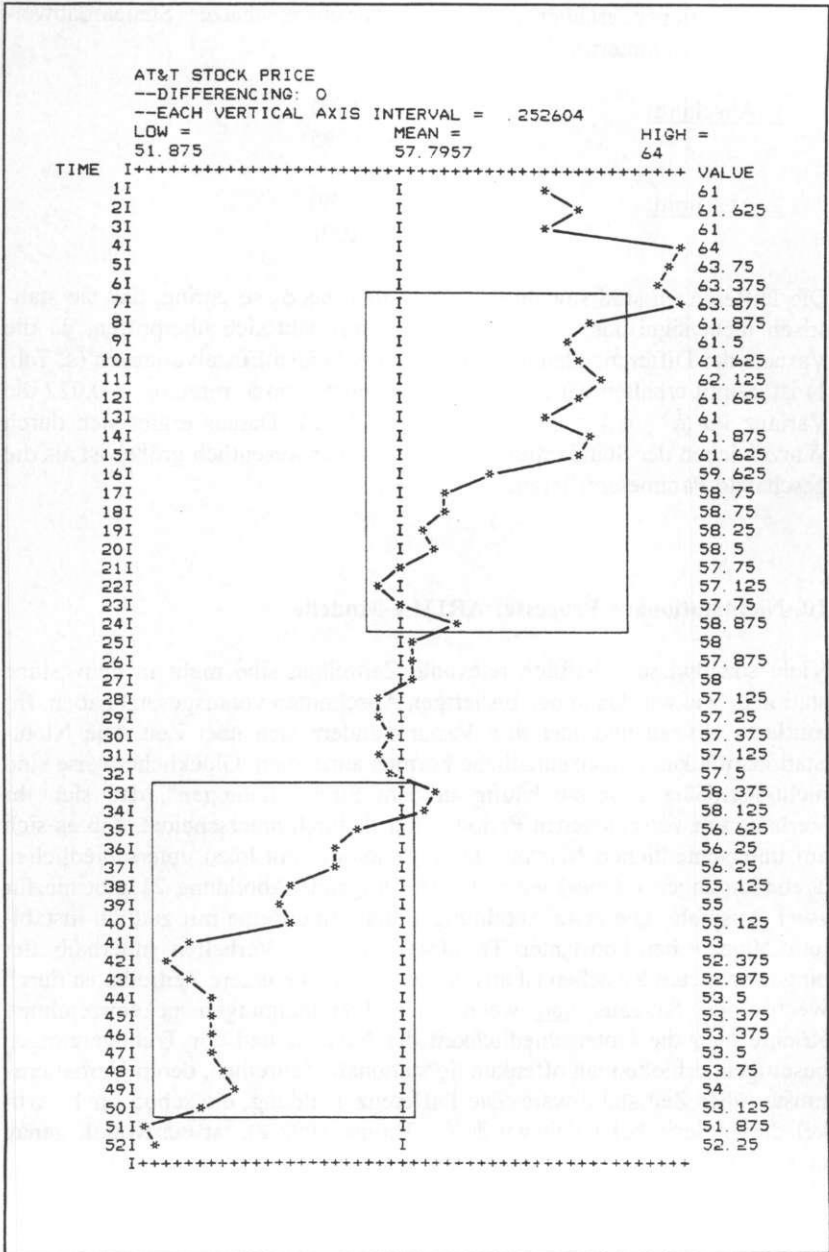


Abb. 21 a: Niveauänderungen bei stabiler Trendneigung. Quelle: Pankratz 1983, S. 161.

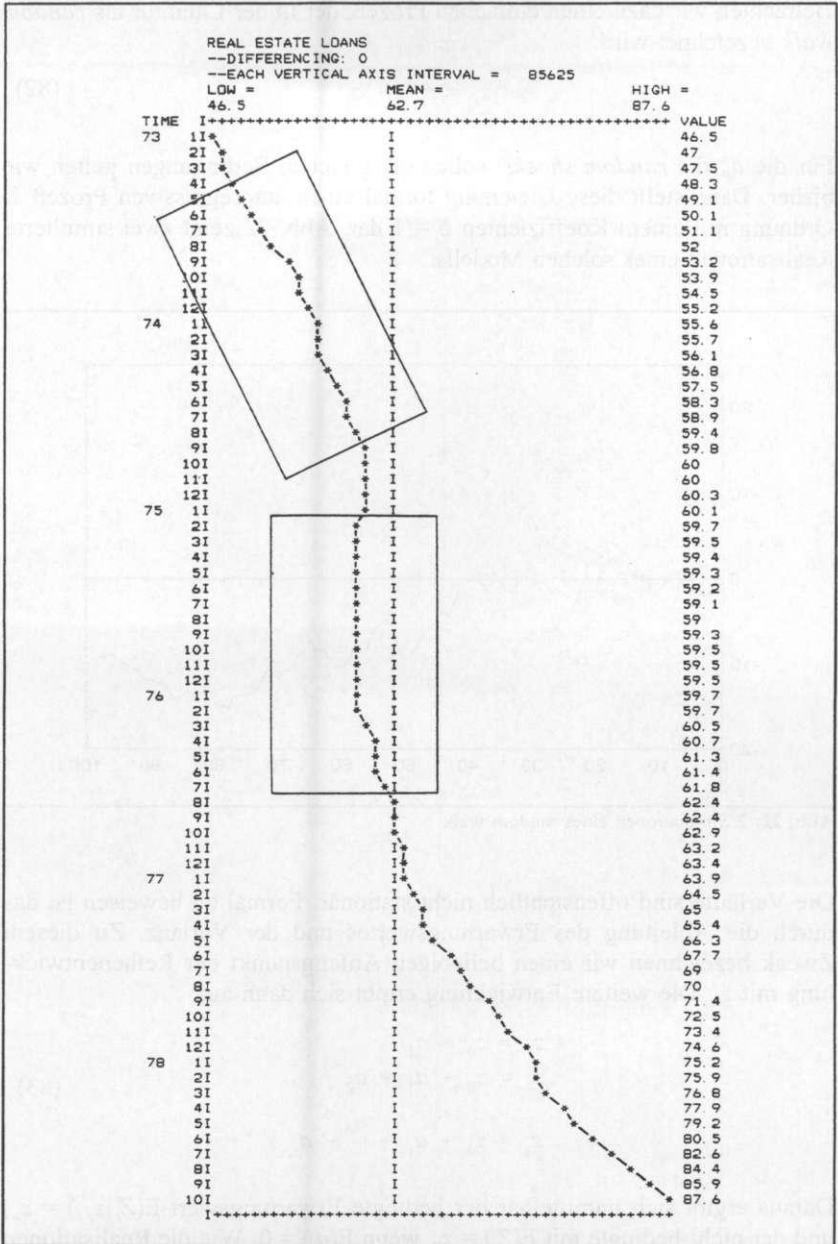


Abb. 21 b: Wechselnde Niveaus und Trendneigungen. Quelle: Pankratz 1983, S. 163.

Betrachten wir dazu einen einfachen Prozeß, der in der Literatur als *random walk* bezeichnet wird:

$$z_t = z_{t-1} + a_t \quad (82)$$

Für die a_t , die *random shocks*, sollen die gleichen Bedingungen gelten wie bisher. Dann stellt diese Gleichung formal einen autoregressiven Prozeß 1. Ordnung mit einem Koeffizienten $\phi = 1$ dar. Abb. 22 zeigt zwei simulierte Realisationen eines solchen Modells.

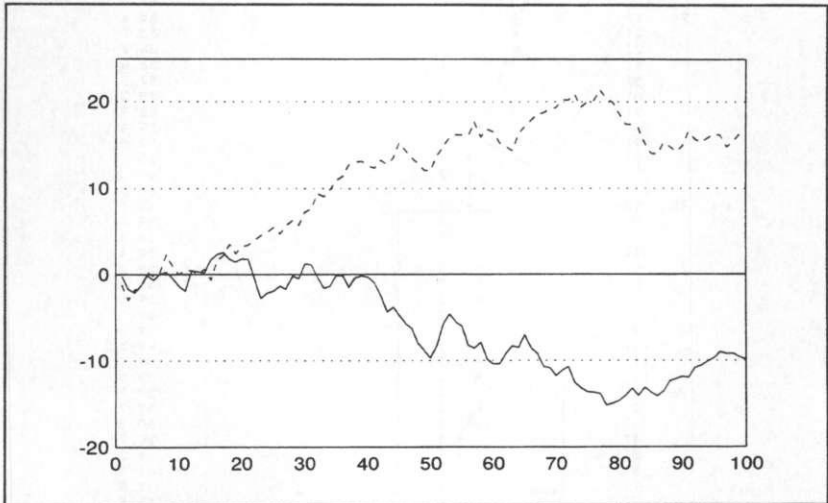


Abb. 22: 2 Simulationen eines random walk

Die Verläufe sind offensichtlich nicht stationär. Formal zu beweisen ist das durch die Ableitung des Erwartungswertes und der Varianz. Zu diesem Zweck bezeichnen wir einen beliebigen Anfangspunkt der Reihenentwicklung mit z_0 . Die weitere Entwicklung ergibt sich dann aus

$$\begin{aligned} z_1 &= z_0 + a_1 \\ z_2 &= z_0 + a_1 + a_2 \\ &\dots \\ z_t &= z_0 + a_1 + \dots + a_t \end{aligned} \quad (83)$$

Daraus ergibt sich unmittelbar der bedingte Erwartungswert $E(Z_t|z_0) = z_0$ und der nicht-bedingte mit $E(Z_t) = z_0$ wenn $E(a_t) = 0$. Wie die Realisationen in Abb. 22 zeigen, hat der *random walk* - anders als ein stationärer Prozeß - nur eine schwache Affinität zu seinem Erwartungswert. Wie die letzte Zeile

in Gleichung 83 verdeutlicht, bleibt der Einfluß jedes *random shocks* für alle Zeiten erhalten, die "Gedächtnisfunktion" ist eine Konstante. Da die *random shocks* als unabhängig voneinander vorausgesetzt werden, ergibt sich die Varianz der Summe aus der Summe der Varianzen σ^2 für die einzelnen Zufallsvariablen:

$$\text{Var}(Z_t) = \text{Var}\left(\sum_{s=1}^t a_s\right) = t \cdot \sigma^2 \quad (84)$$

Das heißt, mit $t \rightarrow \infty$ wächst die Varianz des Prozesses ins Unendliche, sie ist nicht begrenzt. Der Prozeß pendelt in langfristigen, unregelmäßig unterbrochenen Bewegungen zwischen dem positiven und dem negativen Wertebereich (oberhalb und unterhalb des Erwartungswertes) und wiederholt dabei theoretisch erst im Unendlichen irgendeinen Wert, den er zuvor schon einmal realisiert hat (Mills 1990, S. 99). Dies ist gut verträglich mit dem intuitiven Gedanken eines "reinen" Zufallsprozesses; denn irgendeine Grenze oder ein sich im Endlichen exakt wiederholendes Ablaufmuster implizierte eine deterministische Komponente. Wegen der ständigen, aber unregelmäßigen Richtungswechsel spricht man von einem stochastischen Trendverlauf.

Stationär sind aber bei diesem Prozeßtyp die Differenzenbeträge (s. Abb. 23):

$$w_t = z_t - z_{t-1} = a_t \quad (85)$$

Sie bilden in diesem Fall einen *white noise* - ARMA(0,0) - Prozeß. Entscheidend ist aber nicht, ob die Differenzen $w_t = z_t - z_{t-1}$ einer nicht-stationären Zeitreihe Z_t , einen *white noise* Prozeß repräsentieren, sondern daß sie überhaupt einen stationären und invertiblen ARMA(p,q)-Prozeß bilden. In allen diesen Fällen (einer homogenen Nicht-Stationarität in Z_t) lassen sich die Differenzen w_t nach der gleichen dreistufigen Methode modellieren, die in den vorangegangenen Abschnitten erläutert wurde. Dabei gehen keine Informationen über die Originalreihe verloren; denn wenn der erste Wert z_t der Reihe bekannt ist, ergeben sich alle folgenden Werte z_t, z_{t+1}, \dots, z_n als Summe der jeweiligen Differenzenbeträge plus z_t . Diesen Anfangsbetrag kann man sich (wie schon in Thome 1992 a erläutert) wiederum als Summe aller vorangegangenen nicht beobachteten Differenzenbeträge vorstellen, die "unendlich" weit zurückreichen mit einem fiktiven Startwert von $z_{-\infty} = 0$,

formal:

$$\begin{aligned}
 (1-B)z_t &= w_t \\
 z_t &= (1-B)^{-1}w_t \\
 z_t &= (1+B+B^2+\dots+B^\infty)w_t \\
 &= w_t + w_{t-1} + \dots + w_{t-\infty}
 \end{aligned}
 \tag{86}$$

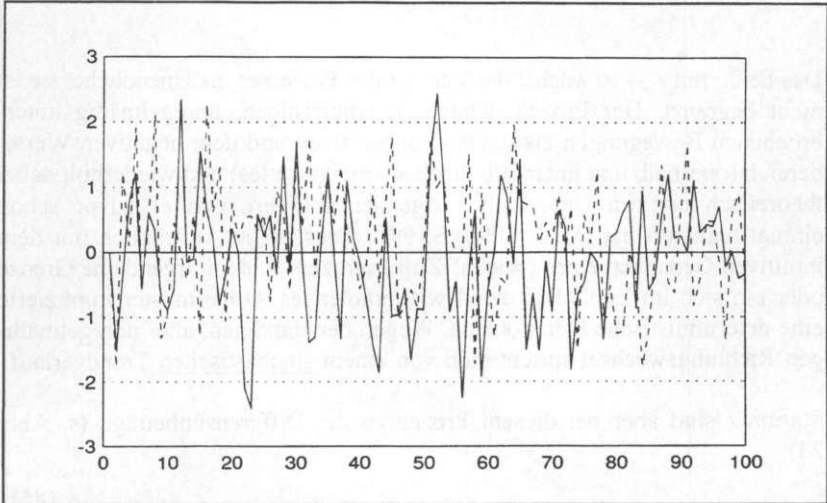


Abb. 23: 1. Differenzen der random walks aus Abb. 22

Gelegentlich reicht eine einmalige Differenzenbildung nicht aus, um eine homogen nicht-stationäre Reihe in eine stationäre Reihe zu überführen. Dann muß der "Grad" der Differenzenbildung erhöht werden, d.h., die "ersten" Differenzen müssen nochmals, evtl. mehrmals "differenziert" werden. Box u. Jenkins (1976) haben vorgeschlagen, den Grad der Differenzenbildung als eigenständigen Parameter d in die Modellstruktur aufzunehmen und die ARMA(p,q)-Modelle zu den sog. ARIMA(p,q)-Modellen zu erweitern. Dabei steht der Buchstabe I für "Integrieren", was eine etwas mißverständliche Bezeichnung ist, da die Differenzenbeträge ja nicht im Sinne der Infinitesimalrechnung integriert, sondern addiert werden. Logisch entspricht die Summenbildung bei diskreten Daten allerdings dem Integrieren bei kontinuierlichen Variablen. Der *random walk* in Gleichung (81) stellt also ein ARIMA(0,1,0)-Modell dar. In den Sozial- und Geschichtswissenschaften wird selten ein Differenzierungsgrad $d > 2$ benötigt. Wenn der Prozeß weitere AR- und/oder MA-Komponenten enthält, läßt sich das ARI-

MA(p,J,#)-Modell in folgender Gleichung ausdrücken:

$$\phi(B)(1-B)^d z_t = \theta(B)a_t \quad (87)$$

Nun stellt sich natürlich sofort die Frage, wie man erkennen kann, ob eine Reihe a) inhomogen nicht-stationär, b) homogen-nichtstationär oder c) stationär ist. Der Augenschein mag zwar einen ersten Hinweis geben, ob die Reihe stationär oder in irgendeiner Weise nicht-stationär ist, aber der Grad der notwendigen Differenzenbildung ist oft schwer zu erkennen, und eine Unterscheidung zwischen homogen und inhomogen nichtstationären Reihen ist auf diese Weise kaum zu treffen. Um die Problemlage besser zu verstehen, betrachten wir zunächst einen AR(1)-Prozeß mit $\phi > 1$, z. B.

$$z_t = 1,05z_{t-1} + a_t \quad (88)$$

Wenn man (willkürlich) einen Anfangswert $z_0 = 10$ und eine Verteilung $a_t \sim N(0,9)$ annimmt, ergibt sich daraus der Prozeßverlauf, der in Abb. 24 wiedergegeben ist.

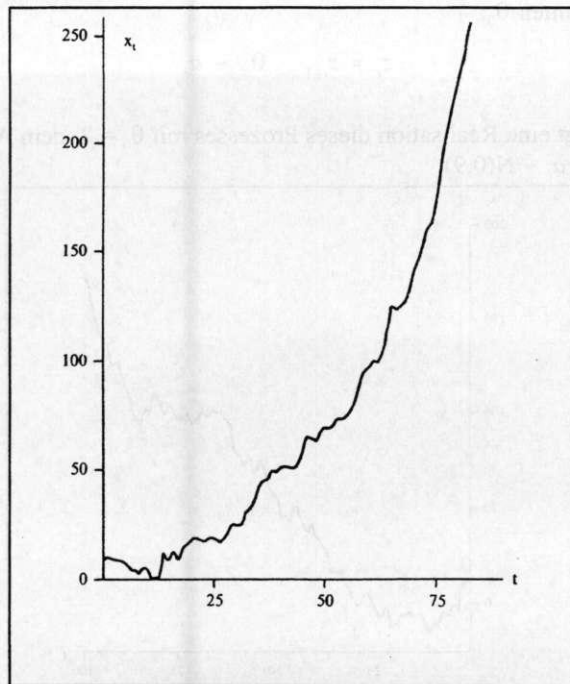


Abb. 24: AR(1)-Prozeß, $\phi=1.05$. Quelle: Mills 1990, S. 97.

Der Verlauf ist "explosiv", mit $t \rightarrow \infty$ übersteigt der Betrag z_t jede Grenze:

$$z_t = z_0 \phi^t + \sum_{i=0}^t \phi^i a_{t-i} \quad (89)$$

Der Einfluß der *random shocks* wird im Vergleich zu dem erreichten Niveau mit fortschreitender Zeit zunehmend geringer, der Prozeß entwickelt sich nahezu deterministisch in einer exponentiellen Wachstumskurve (Nelson 1973, S. 57). Solange die Stationaritätsbedingung nicht erfüllt ist, änderte sich an diesem Verlauf nichts Wesentliches, wenn weitere AR- oder MA-Komponenten hinzugefügt würden (Mills 1990, S. 96). Eine solche Zeitreihe ist durch Differenzenbildung nicht in einen stationären Prozeß überführbar. Formal ist das dadurch erkennbar, daß die Lösung der charakteristischen Gleichung in B kleiner als 1 ist:

$$(1 - 1,05B) = 0 \Rightarrow B = 0,9524 \quad (90)$$

Im Vergleich hierzu betrachten wir als nächstes ein *random walk* Modell mit einer Konstanten θ_0 :

$$z_t = z_{t-1} + \theta_0 + a_t \quad (91)$$

Abb. 25 zeigt eine Realisation dieses Prozesses mit $\theta_0 = 2$, dem Anfangswert $z_0 = 10$ und $a_t \sim N(0,9)$.

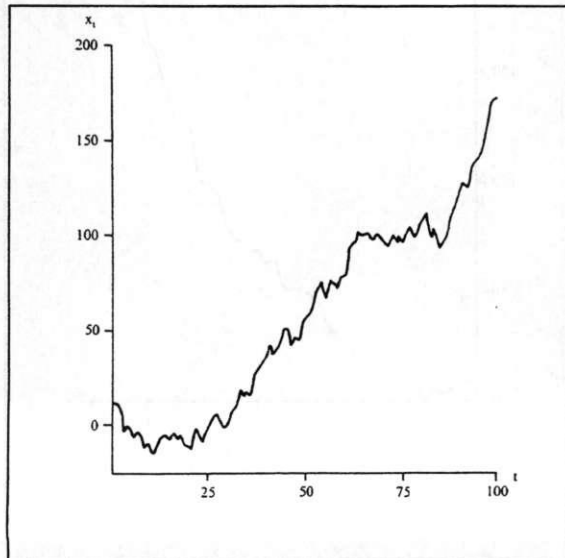


Abb. 25: Random walk mit drift. Quelle: Mills 1990, S. 98.

Wenn wir erneut einen willkürlichen Startwert mit z_0 bezeichnen, erhalten wir den Erwartungswert

$$E(Z_t) = z_0 + t \cdot \theta_0 \quad (92)$$

Die Varianz ist weiterhin $t\sigma^2$. Die Konstante θ_0 bezeichnet man als *drift* Parameter, denn sie führt eine deterministische Komponente in Form eines linearen Trends (mit θ_0 als Steigungskoeffizient) in den Prozeß ein: die Zeitreihe entfernt sich - mit stochastischen Unterbrechungen - zunehmend von ihrem Ursprungswert. Aber auch dieser *random walk with drift* läßt sich durch (einfache) Differenzenbildung in eine stationäre Reihe transformieren:

$$(1-B)z_t = \theta_0 + a_t \quad (93)$$

Wir modifizieren das *random walk* Modell noch einmal, indem wir annehmen, es läge ein zweifach integrierter Prozeß vor, ein Prozeß also, der nur durch zweimalige Differenzenbildung in einen stationären Prozeß überführt werden kann:

$$(1-B)^2 z_t = \theta_0 + a_t \quad (94)$$

Die folgenden Abbildungen 26 a und b zeigen zwei Realisationen eines solchen Prozesses, einmal mit $\theta_0 = 0$ und zum anderen mit $\theta_0 = 2$.

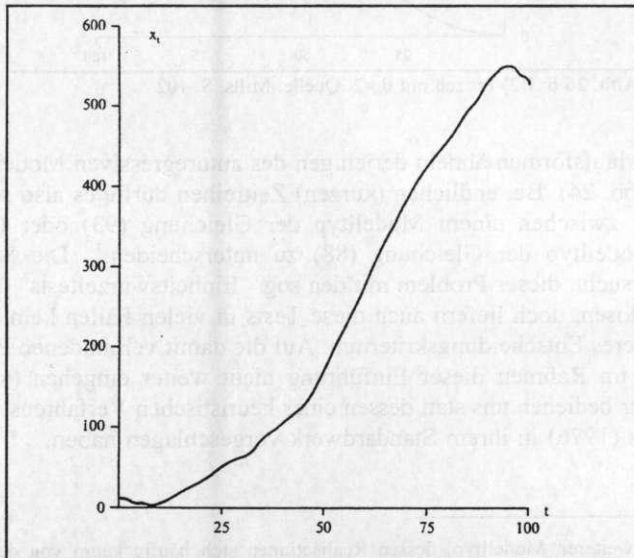


Abb. 26 a: I(2)-Prozeß mit Konstante $\theta_0=0$. Quelle: Mills, S. 101

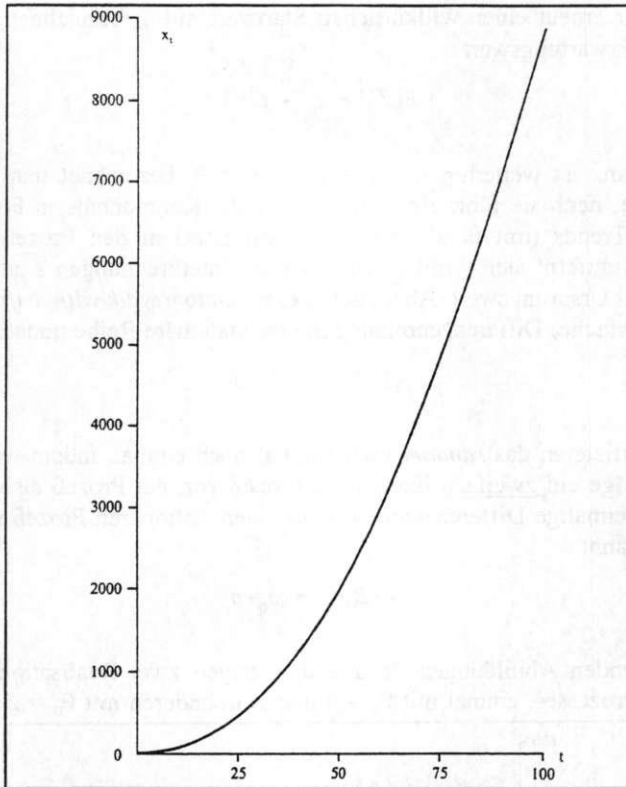


Abb. 26 b: I(2)-Prozeß mit $\theta_0=2$. Quelle: Mills, S. 102

Beide Verlaufsformen ähneln derjenigen des autoregressiven Modells mit $\phi > 1$ (s. Abb. 24). Bei endlichen (kurzen) Zeitreihen dürfte es also schwierig sein, klar zwischen einem Modelltyp der Gleichung (93) oder (94) und einem Modelltyp der Gleichung (88) zu unterscheiden.¹⁹ Die Statistiker haben versucht, dieses Problem mit den sog. "Einheitswurzeltests" (*unit root tests*) zu lösen, doch liefern auch diese Tests in vielen Fällen kein befriedigend sicheres Entscheidungskriterium. Auf die damit verbundenen Probleme kann ich im Rahmen dieser Einführung nicht weiter eingehen (s. Haßler 1994). Wir bedienen uns statt dessen eines heuristischen Verfahrens, das Box u. Jenkins (1976) in ihrem Standardwerk vorgeschlagen haben.

19 Einen weiteren Modelltyp, dessen Realisationen sich häufig kaum von denen eines "integrierten" Prozesses unterscheiden lassen, stellt die Gleichung $z_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \dots + \beta_d t^d + a_t$ dar.

Solange keine substanztheoretischen Gründe dagegensprechen, arbeitet man mit der Vermutung, daß eine erkennbare Nicht-Stationarität homogen sei. Außerdem macht man sich die Tatsache zunutze, daß das Autokorrelogramm einer nicht-stationären Reihe nur allmählich in nahezu linearer Form bei zunehmendem *Lag* *k* gegen Null geht. Ein Beispiel liefert in Abb. 27 die SAKF der schwedischen Kriminalitätsdaten, die in Abschnitt 1 (s. Abb. 2) schon präsentiert wurden.²⁰

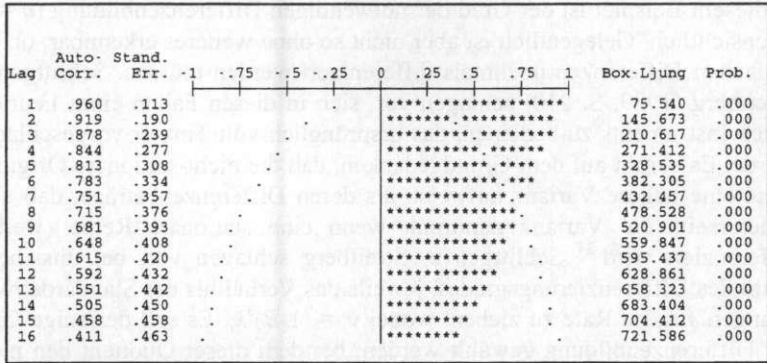


Abb. 27 a: SAKF der schwedischen Kriminalitätsrate aus Abb. 2

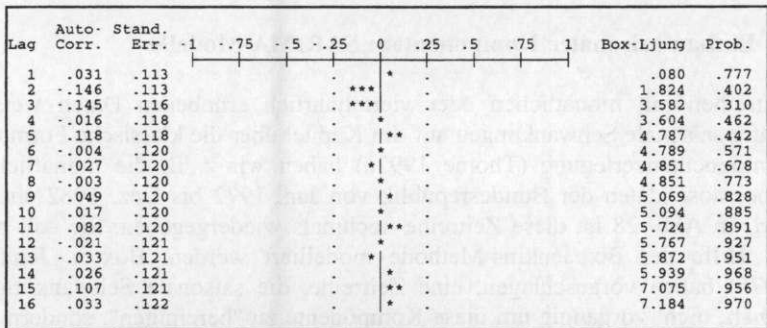


Abb. 27 b: SAKF der einfach differenzierten schwedischen Kriminalitätsrate

Die SAKF in Abb. 27a zeigt den typischen Verlauf der Autokorrelationsfunktion nicht-stationärer Reihen. In den meisten Fällen nimmt sie bei *Lag*

20 Es sei darauf hingewiesen, daß Ausreißer die SAKF erheblich verzerren können, so daß bei einer ernsthaften Analyse die Reihe vor Erstellung des Autokorrelogramms entsprechend adjustiert werden sollte.

$k = 1$ einen Wert in der Nähe von 1 an; gelegentlich kann der Wert aber auch deutlich darunter liegen (s. Box u. Jenkins 1976, S. 200). Da der Trendverlauf die AKF dominiert, sind eventuelle AR- oder MA-Komponenten auf dieser Basis nicht identifizierbar. Das wird anders, wenn wir das Autokorrelogramm der einfach differenzierten Reihe in Abb. 27b betrachten, das auf einen white noise-Prozeß hinweist. Für die Originalreihe kann ein ARIMA(0,1,0)-Modell angenommen werden.

In diesem Beispiel ist der Grad der notwendigen Differenzenbildung ($d \neq 1$) offensichtlich. Gelegentlich ist aber nicht so ohne weiteres erkennbar, ob die einfachen Differenzen nochmals differenziert werden müssen. Schlittgen u. Streitberg (1989, S. 210) schlagen vor, sich in diesen Fällen eines heuristischen Instruments zu bedienen, das ursprünglich von Tintner vorgeschlagen wurde. Es beruht auf dem Grundgedanken, daß die nicht-stationäre Originalreihe eine höhere Varianz aufweist, als deren Differenzenbeträge, daß aber andererseits die Varianz zunimmt, wenn eine stationäre Reihe (weiter) differenziert wird.²¹ Schlittgen u. Streitberg schlagen vor, bei unsicherer Wahl des Differenzierungsgrades, jeweils das Verhältnis der Standardabweichungen s/s_x zu Rate zu ziehen, wobei $y_i = (I-B)^d x_i$. Es soll derjenige Grad der Differenzenbildung gewählt werden, bei dem dieser Quotient den niedrigsten Wert annimmt. Der nächste Abschnitt wird hierzu ein Beispiel liefern.

11. Einbau saisonaler Komponenten: SARIMA-Modelle

Zeitreihen mit monatlichen oder vierteljährlich erhobenen Daten weisen häufig saisonale Schwankungen auf. Im Kapitel über die klassische Form der Komponentenzerlegung (Thome 1992a) haben wir z. B. die monatlichen Arbeitslosendaten der Bundesrepublik von Jan. 1972 bis Dez. 1982 analysiert. In Abb. 28 ist diese Zeitreihe nochmals wiedergegeben. Sie soll nun mit Hilfe der Box/Jenkins-Methode modelliert werden. Box u. Jenkins (1976) haben vorgeschlagen, eine Zeitreihe, die saisonale Schwankungen enthält, nicht vorgängig um diese Komponente zu "bereinigen", sondern sie simultan mit den anderen Komponenten der Reihe zu modellieren.

Saisonale Schwankungen sind (meist stochastische, nicht deterministische) Zyklen mit einer Periode von 1 Jahr. Wir erwarten also hohe Autokorrelationen zwischen Zeitreihenwerten, die 1 Jahr, 2 Jahre, 3 Jahre usw. auseinanderliegen.

21 Dieser Effekt war bei der Behandlung der Komponentenzerlegung bereits verdeutlicht worden (s. Thome 1992 a).

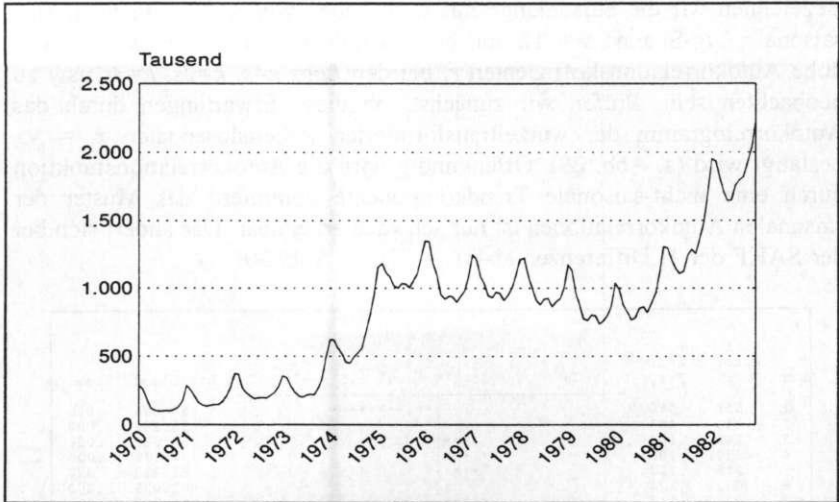


Abb. 28: Anzahl der Arbeitslosen in der BRD von 1972 bis 1982.

Lag	Auto-Corr.	Stand. Err.							Box-Ljung	Prob.
1	.971	.080							149.986	.000
2	.933	.136							289.205	.000
3	.892	.172							417.421	.000
4	.854	.200							535.589	.000
5	.822	.222							645.760	.000
6	.794	.241							749.319	.000
7	.768	.257							847.001	.000
8	.748	.271							940.105	.000
9	.733	.284							1030.182	.000
10	.721	.296							1117.828	.000
11	.705	.307							1202.356	.000
12	.680	.317							1281.544	.000
13	.645	.326							1353.168	.000
14	.602	.334							1416.147	.000
15	.560	.341							1471.026	.000
16	.522	.347							1519.082	.000
17	.490	.352							1561.671	.000
18	.463	.357							1599.952	.000
19	.440	.360							1634.732	.000
20	.422	.364							1667.022	.000
21	.411	.367							1697.913	.000
22	.404	.370							1727.937	.000
23	.395	.373							1756.917	.000
24	.379	.375							1783.738	.000
25	.354	.378							1807.270	.000
26	.323	.380							1827.091	.000
27	.292	.382							1843.396	.000
28	.265	.383							1856.922	.000
29	.243	.384							1868.376	.000
30	.225	.385							1878.306	.000
31	.211	.386							1887.082	.000
32	.202	.387							1895.199	.000
33	.199	.388							1903.106	.000
34	.198	.388							1911.022	.000
35	.194	.389							1918.717	.000
36	.182	.389							1925.543	.000
37	.162	.390							1930.994	.000
38	.136	.390							1934.862	.000
39	.109	.391							1937.366	.000
40	.085	.391							1938.885	.000

Abb. 29: SAKF der Arbeitslosendaten

Bezeichnen wir die Saisonlänge mit s , so haben wir bei Monatsdaten eine saisonale Lag-Spanne $s = 12$ und bei Quartalsdaten $s = 4$. Es sollten also hohe Autokorrelationskoeffizienten r_k bei den Lags $k=s, k=2s, k=3s$ usw zu beobachten sein. Prüfen wir zunächst, ob diese Erwartungen durch das Autokorrelogramm der wurzeltransformierten Arbeitslosendaten $x_t = \sqrt{z_t}$ bestätigt wird (s. Abb. 29). Offenkundig wird die Autokorrelationsfunktion durch eine nicht-saisonale Trendkomponente dominiert; das Muster der saisonalen Autokorrelationen ist nur schwach erkennbar. Das ändert sich bei der SAKF der 1. Differenzen $(1-B)x_t = \nabla x_t$ in Abb. 30.

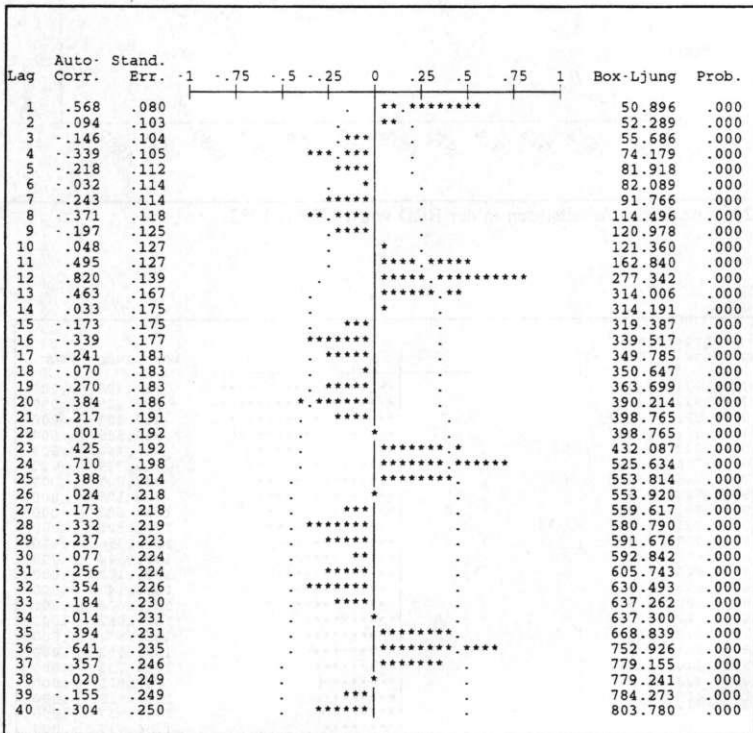


Abb. 30: SAKF der ersten Differenzen der Arbeitslosendaten

Jetzt sind die "Peaks" bei den saisonalen Lags $k = 12, 24, 36$ deutlich erkennbar, außerdem relativ hohe Autokorrelationskoeffizienten in den unmittelbar benachbarten Lags. Das Muster der r_k 's für die Lags $k= 12, 24$ usw. (wir wollen in Zukunft von "saisonalen Lags" und "saisonalen Autokorrelationskoeffizienten" sprechen) ist in der gleichen Weise zu interpretieren wie das Muster der SAKF nicht-saisionaler Zeitreihen. Als erstes ist festzustellen, daß die saisonalen Autokorrelationskoeffizienten nur ganz allmählich

nach Null abfallen, also saisonale Nicht-Stationarität indizieren. Folglich muß auch eine saisonale Differenzenbildung vorgenommen werden (s. Thome 1992a):

$$(1 - B)(1 - B^{12})x_t = \nabla_1 \nabla_{12} x_t \quad (95)$$

Wird ein Filter (in diesem Falle ein Differenzenfilter) auf eine bereits gefilterte Reihe angewandt, erzielt man den gleichen Effekt wie bei der Anwendung der multiplizierten Filter auf die Originalreihe. Inhaltlich bedeutet das "Hintereinanderschalten" von einfachem und saisonalem Filter, daß die Differenzen in den Werten benachbarter Monate um die Wertdifferenz "bereinigt" werden, die für die gleichen Monate ein Jahr zuvor beobachtet wurden:

$$(1 - B - B^{12} + B^{13})x_t = (x_t - x_{t-1}) - (x_{t-12} - x_{t-13}) \quad (96)$$

Für die einfach und saisonal differenzierte Zeitreihe erhalten wir die in den Abbildungen 31 a und b wiedergegebenen Autokorrelogramme:

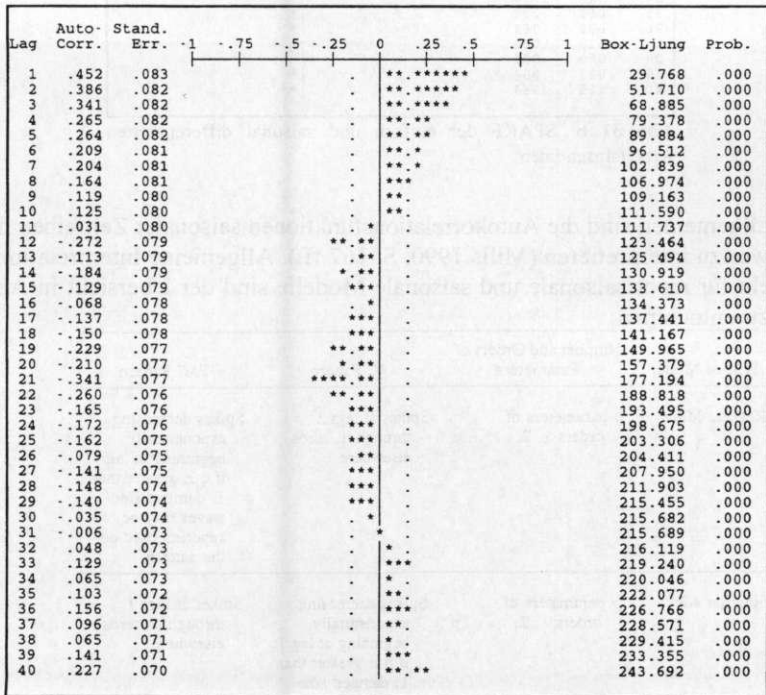


Abb. 31 a: SAKF der einfach und saisonal differenzierten Arbeitslosenraten

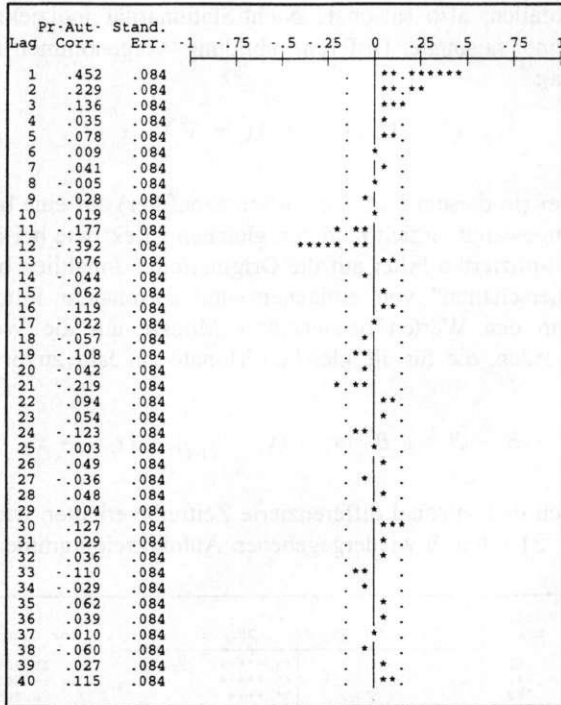


Abb. 31 b: SPAKF der einfach und saisonal differenzierten Arbeitslosenraten

Im allgemeinen sind die Autokorrelationsfunktionen saisonaler Zeitreihen nur schwer zu interpretieren (Mills 1990, S. 167 ff.). Allgemeine Interpretationsregeln für nicht-saisonale und saisonale Modelle sind der Übersicht in Abb. 32 zu entnehmen.

Type of Model	Number and Orders of Parameters	AC Pattern	PAC Pattern
Regular MA	q parameters of orders 1, 2, . . . , q	Spikes at lags 1 through q , zeros elsewhere	Spikes decreasing exponentially beginning at lag 1; if q is greater than 1, damped sine waves may be superimposed on the pattern
Regular AR	p parameters of orders 1, 2, . . . , p	Spikes decreasing exponentially beginning at lag 1; if p is greater than 1, damped sine waves may be superimposed on the pattern	Spikes at lags 1 through p , zeros elsewhere

Regular ARMA	q MA and p AR parameters of orders 1, 2, . . . , q and 1, 2, . . . , p , respectively	Irregular pattern of spikes at lags 1 through q ; remaining autocorrelations as in regular AR	Same as AC pattern, with irregular pattern of spikes at lags 1 through p
Purely seasonal MA	Q parameters of orders $s, 2s, \dots, Qs$	Spikes at lags $s, 2s, \dots, Qs$, zeros elsewhere	Spikes decreasing exponentially at lags that are multiples of s , beginning at lag s
Purely seasonal AR	P parameters of orders $s, 2s, \dots, Ps$	Spikes decreasing exponentially at lags that are multiples of s , beginning at lag s	Spikes at lags $s, 2s, \dots, Ps$, zeros elsewhere
Purely seasonal ARMA	Q MA and P AR parameters of orders $s, 2s, \dots, Ps$ and $s, 2s, \dots, Qs$, respectively	Irregular pattern of spikes at lags 1 through Q ; remaining autocorrelations as in seasonal AR	Same as AC pattern, with irregular pattern of spikes at lags 1 through P

Abb. 32: Interpretationsregeln für verschiedene Modelle. Quelle: Hoff 1983, S. 191.

Betrachten wir nur die saisonalen *Lags*, so ist in der SAKF lediglich bei $k = 12$ (nicht bei $k = 24$ oder $k = 36$) ein signifikanter Autokorrelationskoeffizient festzustellen. Das spräche für eine MA(1)-Komponente in diesem Modellteil. (Allerdings ist ein geometrischer Abfall im SPAKF-Muster bei den *Lags* $k = 12, 24, 36$ nur mit einigem guten Willen zu erkennen.) Falls die Saisonfigur völlig starr wäre, d. h. sich von Jahr zu Jahr die gleichen Monatswerte wiederholten, wäre diese Figur durch den saisonalen Differenzenfilter $(1-B^{12})$ völlig eliminiert. Die in unserem Beispiel nach der saisonalen Differenzenbildung weiterhin bestehende jährliche Abhängigkeitsbeziehung zum *Lag* $k=12$ kann mit der gleichen Modellogik formalisiert werden, mit der wir bisher schon die monatlichen (nicht-saisonalen) Abhängigkeiten einer Zeitreihe modelliert haben. Ein ARIMA(0,1,1)₁₂-Modell nur für die Saisonkomponente läßt sich also wie folgt schreiben:

$$\Phi(B^s)\nabla_s^D x_t = \Theta(B^s)\alpha_t \tag{97}$$

Der Exponent "D" gibt den Grad der saisonalen Differenzenbildung an (meistens ist $D = 1$). Die AR- und MA-Polynome sind nun wie folgt zu interpretieren:

$$\begin{aligned} \Phi(B^s) &= 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \dots - \Phi_p B^{Ps} \\ \Theta(B^s) &= 1 - \Theta_1 B^s - \Theta_2 B^{2s} - \dots - \Theta_Q B^{Qs} \end{aligned} \tag{98}$$

Der Grad der autoregressiven Saison-Komponente wird mit (groß) P , der Grad der *moving average* Saison-Komponente mit (groß) Q angegeben. Für die saisonalen Polynome gelten die gleichen Stationaritäts- und Invertibilitätsbedingungen wie für die nicht-saisonalen. Die Fehlerkomponente in Gleichung (97) ist nun mit a_t bezeichnet. Die "Fehler" a_t , die zu einem bestimmten Monat (z. B. "März") gehören, sind bei korrekter Bestimmung der saisonalen Komponente von Jahr zu Jahr unkorreliert. Korreliert sein können aber noch die Fehler der Monate, die näher benachbart sind, vor allem die Fehler der unmittelbar aufeinander folgenden Monate. Diese verbleibende Abhängigkeitsstruktur in der Reihe a_t kann nun ihrerseits durch ein zweites, ein ARIMA($p, d, <7$)-Modell, repräsentiert werden:

$$\phi(B)\nabla^d \alpha_t = \theta(B)a_t \quad (99)$$

Substituiert man diese Gleichung in Gleichung (97), erhält man das allgemeine multiplikativ- saisonale ARIMA-Modell, das auch als SARIMA-Modell bezeichnet wird:

$$\phi_p(B)\Phi_p(B^s)\nabla^d \nabla_s^D x_t = \theta_q(B)\Theta_q(B^s)a_t \quad (100)$$

Falls die Reihe auch noch eine deterministische Trendkomponente aufweist, kann dies durch eine Konstante θ_0 auf der rechten Gleichungsseite berücksichtigt werden.

Nachdem wir in unserem Beispiel versuchsweise schon ein ARIMA(0,1,1)₁₂-Modell für die Saison-Komponente identifiziert haben, müssen wir nun für die Restreihe a_t , ebenfalls ein entsprechendes (Teil-)Modell identifizieren. Dabei stützen wir uns vor allem auf das Muster der (partiellen) Autokorrelationskoeffizienten der Lags $k = 1$ bis $k = 11$. Die SPAKF der einfach und saisonal differenzierten Reihe (s. oben Abb. 31b) zeigt nur zwei signifikante Koeffizienten bei $k = 1$ und $k = 2$. Dies spricht für ein AR(2)-Modell in der nicht-saisonalen Komponente. Werden saisonale und nicht-saisonale Komponente gemäß Gleichung (100) zusammengefügt, resultiert daraus ein SARIMA(2,1,0)(0,1,1)₁₂-Modell der Arbeitslosendaten:

$$(1 - \phi_1 B - \phi_2 B^2)(1 - B)(1 - B^{12})x_t = (1 - \theta B^{12})a_t \quad (101)$$

Die Interpretation der SAKF in Abb. 31a ist aber unsicher, weil auch nach einfacher und saisonaler Differenzierung der Reihe die empirische Autokorrelationsfunktion immer noch recht langsam nach Null abfällt. Möglicherweise wäre hier ein nochmaliges Differenzieren (also $d = 2$) angebracht. Abb. 33 stellt die differenzierte Reihe $y_t = \nabla \nabla_{12} x_t$ der Reihe $y_t = \nabla^2 \nabla_{12} x_t$ gegenüber.

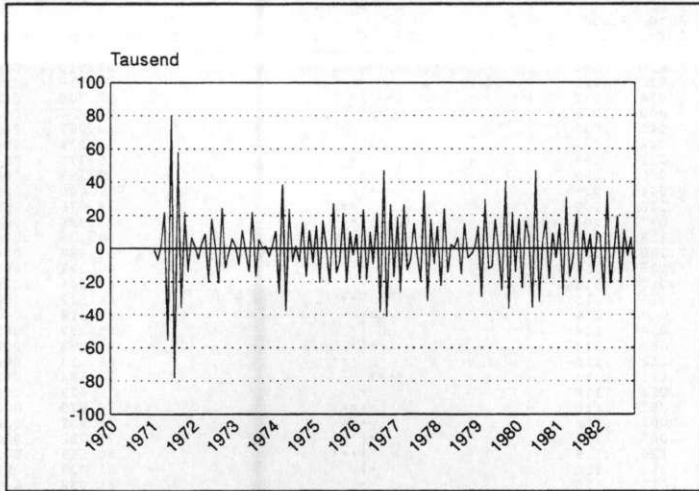


Abb. 33 a: Differenzierte Arbeitslosenreihe ($d=1, D=1$)

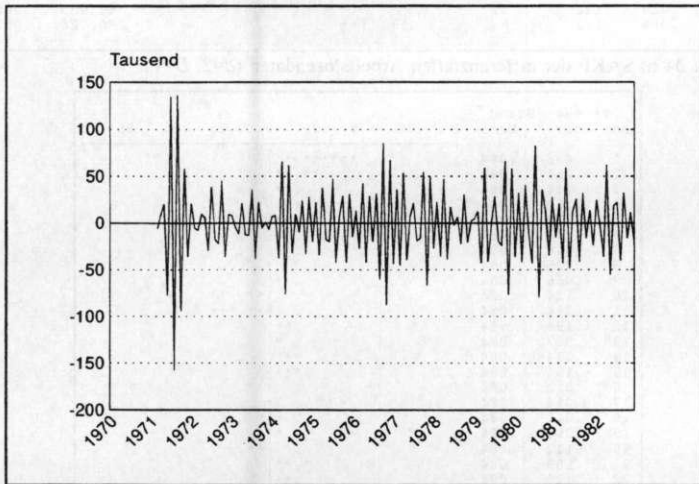


Abb. 33 b: Differenzierte Arbeitslosenreihe ($d=2, D=1$)

Die SAKF und die SPAKF von der mit $d = 2$ differenzierten Reihe sind Abb. 34 a und b zu entnehmen.

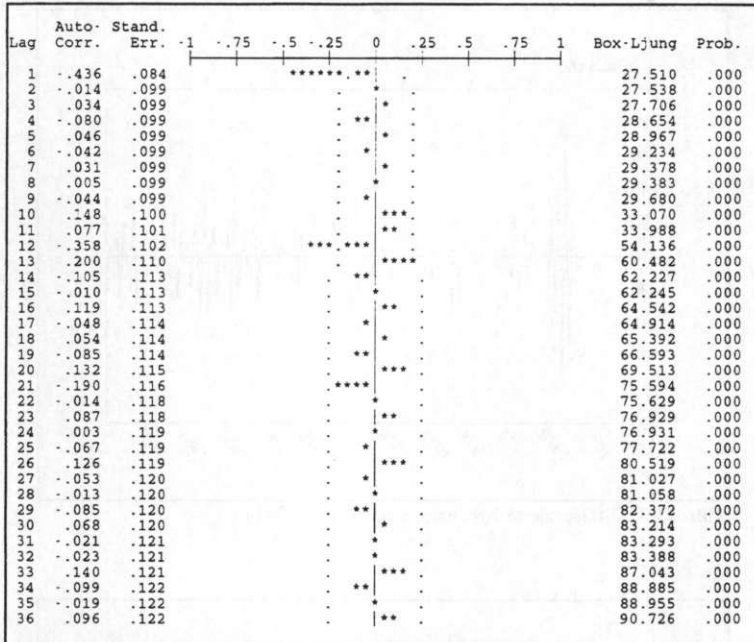


Abb. 34 a: SAKF der differenzierten Arbeitslosendaten ($d=2, D=1$)

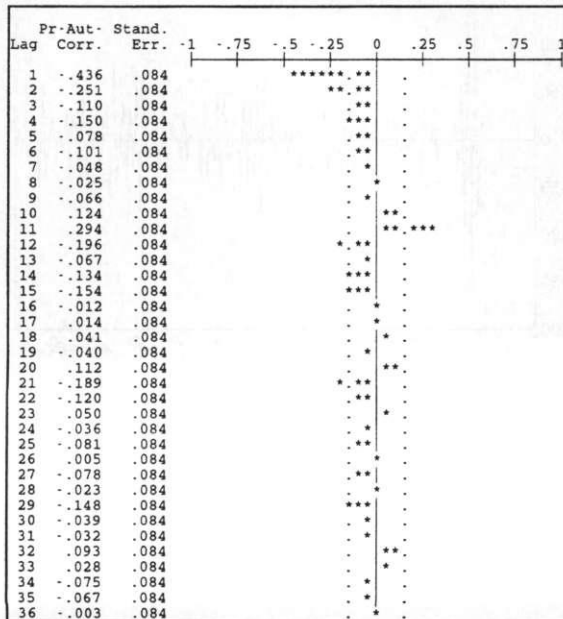


Abb. 34 b: SPAKF der diff. Arbeitslosendaten ($d=2, D=1$)

Die dort erkennbaren Muster legen recht deutlich ein $SARIMA(0,2,1)(0,1,1)$ -Modell nahe: $(1-B)^2(1-B^{12})x_t = (1-\theta B^{12})a_t$. Beim Übergang von $d = 1$ zu $d = 2$ nimmt allerdings die Varianz leicht zu (s. Tabelle 3).

Tabelle 3:
Quotient der Standardabweichungen s_t/s_t , bei
verschiedenen Graden der Differenzenbildung

	$(1-B^{12})^0$	$(1-B^{12})^1$
$(1-B)^0$	1	-
$(1-B)^1$	0,144	0,0624
$(1-B)^2$	-	0,0652

Ein so geringer Zuwachs des Varianzquotienten ist allerdings akzeptabel, wenn die Diagnoseergebnisse (gemäß der in Abschnitt 9 erläuterten Kriterien) des Modells mit höherem Grad der Differenzenbildung "besser" sind als diejenigen des Modells mit niedrigerem Grad der Differenzenbildung. Die beiden Modelle werden mit folgenden Parametern geschätzt (Standardfehler in Klammern):

$SARIMA(2.1.0)(0JJ)$:

$$\hat{\phi}_1 = 0.35 \quad \hat{\phi}_2 = 0.28 \quad \hat{\theta}_{12} = 0.53$$

(0.08) (0.08) (0.08)

$SARIMA(0.2.1)(0JJ)$:

$$\theta_1 = 0.69 \quad \theta_{12} = 0.57$$

(0.06) (0.08)

Wir ersparen uns hier eine ausführliche Darstellung der Diagnoseergebnisse, die für beide Modelle akzeptabel sind. Nach den *AIC*- und *SBC*-Informationskriterien ist allerdings das zweite (höher differenzierte) Modell vorzuziehen. Ein weiterer Gütetest wäre zu gewinnen, wenn wir beide Modelle zur Prognose der Arbeitslosenzahlen ab Januar 1983 benutzen würden. Das Prognosethema kann im Rahmen dieses Artikels allerdings nicht mehr behandelt werden. Zum Schluß sei der Gang der Box/Jenkins-Modellbildung in einem Ablaufschema zusammengefaßt (s. Abb. 35)

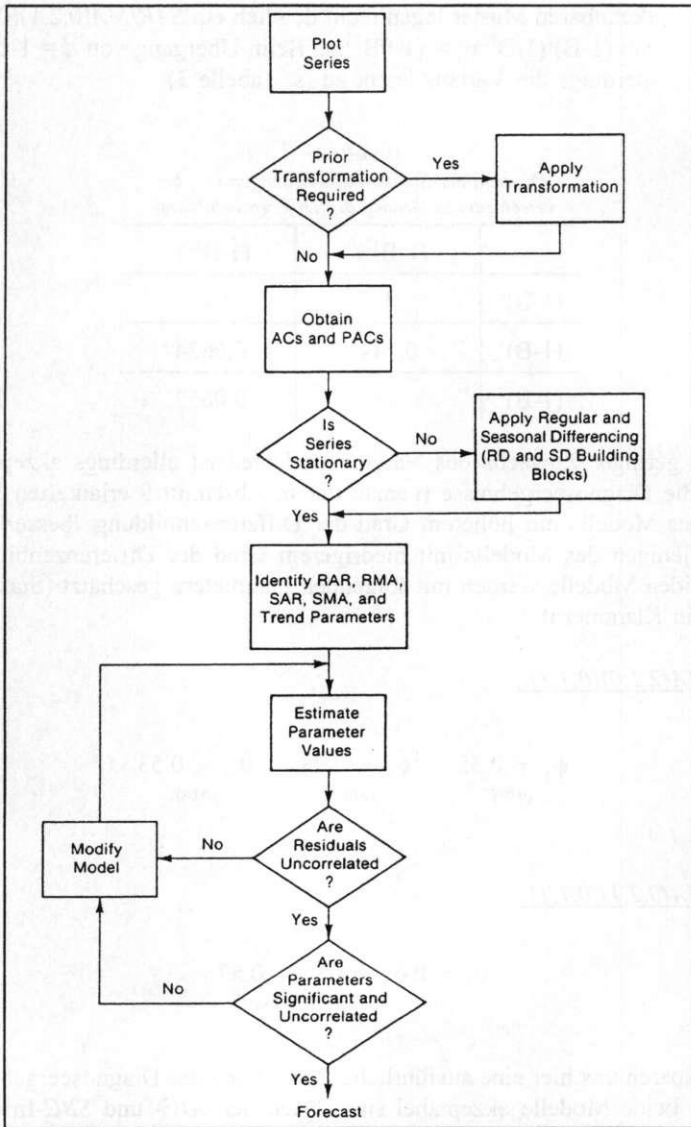


Abb. 35: Ablaufschema der Box/Jenkins-Modellbildung. Quelle: Hoff 1983, S. 240.

Der Buchstabe "R" steht hier für "regulär", also für die nicht-saisonale Modellkomponente. Die "prior transformations" zu Beginn des Analyseprozesses beziehen sich vor allem auf Datentransformationen zur Varianzstabilisierung.

Literatur

- ANDERSON, O. D., 1976: Time Series Analysis and Forecasting, the Box-Jenkins Approach; London: Butterworth.
- BOX, G. E. P. u. JENKINS, G. M., 1976: Time Series Analysis, Forecasting and Control (2nd Ed.); San Francisco: Holden Day.
- HABLER, U., 1994: Einheitswurzeltests - Ein Überblick. Allgemeines Statistisches Archiv 78: 207-228.
- HOFF, J. C., 1983: A Practical Guide to Box-Jenkins Forecasting; Belmont: Wadsworth.
- LIU, L.-M. u. HUDAK, G. B. in collaboration with G. E. P. Box, M. E. Muller, G. C. Tiao, 1992: Forecasting and Time Series Analysis Using the SCA Statistical System. Volume 1. De Kalb, Illinois: Scientific Computing Associates.
- METZ, R. u. STIER, W., 1992: Modelling the Long Wave-Phenomena. Historical Social Research / Historische Sozialforschung 17/3: 43-62.
- MILLS, T. C., 1990: Time Series Techniques for Economists; Cambridge etc.: Cambridge University Press.
- PANKRATZ, A., 1983: Forecasting With Univariate Box-Jenkins Models. Concepts and Cases; New York etc.: John Wiley & Sons.
- PINDYCK, R. S. u. RUBINFELD, D. L., 1981: Econometric Models and Economic Forecasts (2nd Ed.); New York etc.: McGraw-Hill.
- SCHLITGEN, R. u. STREITBERG, B. H. J., 1989: Zeitreihenanalyse (3. A.); München / Wien: Oldenbourg.
- STIER, W., 1992: Multivariate Zeitreihenanalyse. Skript zum Frühjahrsseminar 1992 des Zentralarchivs für Empirische Sozialforschung.
- THOME, H., 1990: Grundkurs Statistik für Historiker. Teil II: Induktive Statistik und Regressionsanalyse. Historical Social Research / Historische Sozialforschung, Supplement 3.
- THOME, H., 1992a: Beschreibende Zeitreihenanalyse: Komponentenerlegung mit Hilfe gleitender Mittelwerte. Historical Social Research / Historische Sozialforschung 17/3: 63-105.
- THOME, H., 1992b: Ausreißer und fehlende Werte in der Zeitreihenanalyse: Ihre Modellierung im Rahmen des Box/Jenkins-Ansatzes. ZA-Information 31: 37-69.
- WEI, W. W. S., 1990: Time Series Analysis. Univariate and Multivariate Methods; Redwood City etc.: Addison-Wesley.