Lehr- und Handbücher der Statistik

Herausgegeben von Universitätsprofessor Dr. Rainer Schlittgen

Bisher erschienene Werke:

Böhning, Allgemeine Epidemiologie
Caspary · Wichmann, Lineare Modelle
Chatterjee · Price (Übers. Lorenzen), Praxis der Regressionsanalyse, 2. Auflage
Degen · Lorscheid, Statistik-Lehrbuch
Degen · Lorscheid, Statistik-Aufgabensammlung, 3. Auflage
Hartung, Modellkatalog Varianzanalyse
Harvey (Übers. Untiedt), Ökonometrische Analyse von Zeitreihen, 2. Auflage
Harvey (Übers. Untiedt), Zeitreihenmodelle, 2. Auflage
Heiler · Michels, Deskriptive und Explorative Datenanalyse
Kockelkorn, Lineare statistische Methoden
Miller (Übers. Schlittgen), Grundlagen der Angewandten Statistik
Naeve, Stochastik für Informatik
Oerthel · Tuschl, Statistische Datenanalyse mit dem Programmpaket SAS
Pflaumer · Heine · Hartung, Statistik für Wirtschaft- und Sozialwissenschaften: Deskriptive Statistik, 2. Auflage
Pflaumer · Heine · Hartung, Statistik für Wirtschafts- und Sozialwissenschaften: Induktive Statistik
Pokropp, Lineare Regression und Varianzanalyse
Rasch · Herrendörfer u.a., Verfahrensbibliothek, Band I und Band 2
Riedwyl · Ambühl, Statistische Auswertungen mit Regressionsprogrammen
Rinne, Wirtschafts- und Bevölkerungsstatistik, 2. Auflage
Rinne, Statistische Analyse multivariater Daten - Einführung
Rüger, Induktive Statistik, 3. Auflage
Rüger, Test- und Schätztheorie, Band I: Grundlagen
Schlittgen, Statistik, 9. Auflage
Schlittgen, Statistische Inferenz
Schlittgen, GAUSS für statistische Berechnungen
Schlittgen · Streitberg, Zeitreihenanalyse, 9. Auflage
Schürger, Wahrscheinlichkeitstheorie
Tutz, Die Analyse kategorialer Daten

Fachgebiet Biometrie

Herausgegeben von Dr. Rolf Lorenz

Bisher erschienene Werke:

Bock, Bestimmung des Stichprobenumfangs
Brunner · Langer, Nichtparametrische Analyse longitudinaler Daten

Zeitreihenanalyse

Von
Prof. Dr. Rainer Schlittgen
und
Prof. Dr. Bernd H. J. Streitberg

9., unwesentlich veränderte Auflage

R. Oldenbourg Verlag München Wien

Die Deutsche Bibliothek - CIP-Einheitsaufnahme

Schlittgen, Rainer:
Zeitreihenanalyse / von Rainer Schlittgen und Bernd H. J. Streitberg. – 9., unwes. veränd. Aufl. – München ; Wien : Oldenbourg, 2001
(Lehr- und Handbücher der Statistik)
ISBN 3-486-25725-0

Rosenheimer Straße 145, D-81671 München
Telefon: (089) 45051-0
www.oldenbourg-verlag.de

Gedruckt auf säure- und chlorfreiem Papier
Druck: Tutte Druckerei GmbH, Salzweg
Bindung: R. Oldenbourg Graphische Betriebe Binderei GmbH

ISBN 3-486-25725-0

Inhaltsverzeichnis

Vorwort zur neunten Auflage

Die ‚Zeitreihenanalyse', die eine an der Datenanalyse orientierte Einführung in dieses Gebiet gibt, liegt hiermit in der 9. Auflage vor. Die intensive Nutzung des Buches drückt sich nicht nur darin aus, daß eine neue Auflage nötig wurde, sondern vor allem auch in Zuschriften von Lesern, die bei ihren detaillierten Studien auf Fehler aufmerksam wurden und mich davon in Kenntnis setzten. (Diese resultierten vor allem auf die zwischenzeitliche Umstrukturierung.) Ihnen ist es zu verdanken, wenn diese Auflage nun weitestgehend fehlerfrei ist.

Vorwort zur ersten Auflage

Wir wollen mit diesem Buch eine elementare und anwendungsorientierte Einführung in das Gebiet der Zeitreihenanalyse geben. Zeitreihen, d. h. zeitlich geordnete Folgen von Beobachtungen treten in jeder wissenschaftlichen Disziplin auf, sobald die Dynamik und zeitliche Entwicklung realer Systeme empirisch untersucht wird. Vier **Beispiele** sollen die Vielfalt der möglichen Fragestellungen illustrieren (zahlreiche weitere, i. a. prosaischere Anwendungen werden im Textteil des Buches folgen):

(1) Existiert intelligentes Leben außerhalb der Erde? Nach unserem Kenntnisstand ist eine notwendige Voraussetzung das Vorhandensein von Planetensystemen. Planeten, die fremde Fixsterne umkreisen, sind zu klein, um direkt beobachtet werden zu können. Ihre Anwesenheit kann jedoch mit zeitreihenanalytischen Methoden (Spektralanalyse) in Folgen von Lichtstärkemessungen eines Fixsterns durch den Nachweis sehr schwacher, aber regelmäßiger Schwankungen der Lichtstärke beim Durchgang des Planeten aufgedeckt werden.

(2) Effiziente Übertragung menschlicher Sprache. Technisch gesehen, ist die menschliche Sprache eine Zeitreihe von Luftdruckschwankungen. Die direkte, analoge oder digitale Übertragung dieser Zeitreihe ist unnötig aufwendig, da nur wenige verschiedene Phoneme auftreten, deren Bildung jeweils längere Zeit in Anspruch nimmt. Ein alternativer Ansatz ist die Modellierung von Teilstücken der Zeitreihe durch Modelle mit wenigen Parametern (sog. autoregressive Prozesse), die während des Telefongesprächs geschätzt werden. Anstelle die Reihe selbst zu übertragen, genügt es dann, die geschätzten Parameter zu übermitteln und am anderen Ende der Strecke die entsprechende Zeitreihe zu simulieren. Diese Technik spart erhebliche Leitungskosten und wird experimentell bereits von einer amerikanischen Telefongesellschaft erprobt.

(3) Vorhersage von Aktienkursen. Die ökonomische Theorie effizienter Märkte behauptet, daß Aktienkursreihen keinerlei für die Prognose auswertbaren Informationen mehr enthalten. Eine Detailanalyse empirischer Reihen zeigt jedoch das Vorhandensein bestimmter Regelmäßigkeiten, die u. U. eine gewisse Vorhersage des zukünftigen Verhaltens der Aktienkurse erlauben. Die Entwicklung praktikabler Prognosemethoden für Aktienkurse wäre mit Sicherheit eine lohnende Aufgabe; mit den Worten von Mosteller könnte man dann in der Baisse kaufen, in der Hausse verkaufen und sich anschließend als Philanthrop zur Ruhe setzen.

(4) Kontrolle von Marschflugkörpern. Leider sind durchaus nicht alle Anwendungen der Zeitreihenanalyse philanthropischer Natur. Moderne Marschflugkörper benutzen ein stochastisches Modell ihres Trägheitslenksystems um ein Abdriften der Flugbahn zu verhindern. Vorhersagen des Modells über die Position des Flugkörpers (mittels sog. Kalman-Filter) werden mit Radarmessungen des überflogenen Terrains verglichen, wobei Differenzen zu einer Korrektur der Flugbahn verwendet werden.

Die vier Beispiele entstammen den vier **Hauptanwendungsbereichen** der Zeitreihenanalyse:

(1) **Deskription**, d. h. Beschreibung von Zeitreihen. In der deskriptiven (explorativen) Datenanalyse gilt das Feyerabendsche Prinzip des „anything goes". Die deskriptiven Methoden, die wir im ersten, einleitenden Kapitel vorstellen, sind ver-

gleichsweise konservativ. Als vereinheitlichende Leitidee dient die Kleinstquadrateapproximation der Zeitreihe durch verschiedene lineare und nichtlineare Regressionsmodelle. Eine typische Anwendung ist die Komponentenzerlegung einer gegebenen Zeitreihe in Trendkomponente (entsprechend der langfristigen Entwicklung), zyklische Komponenten (entsprechend regelmäßigen, etwa saisonalen Schwankungen bekannter Periodizität) und einer irregulären Restkomponente. Da die Voraussetzungen des Regressionsmodells (insbesondere die Unabhängigkeitsannahme) in der Zeitreihenanalyse oft unrealistisch sind, werden diese Methoden nicht als echte Modellanpassung begriffen, sondern nur als vorläufige, heuristische Approximation der Zeitreihe mit dem Ziel der Entdeckung erster Regelmäßigkeiten.

(2) **Modellierung**, d.h. Formulierung und Anpassung stochastischer Modelle. Die beobachtete Reihe wird hier als Realisierung eines stochastischen Prozesses aufgefaßt, also einer Folge von (im allgemeinen abhängigen) Zufallsvariablen. Eine zentrale Annahme ist die der Stationarität, d.h. der zeitlichen Konstanz der stochastischen Charakteristika des Prozesses. Die Formulierung von stochastischen Prozeßmodellen wird in Kapitel 2 und 3 behandelt, wobei im Vordergrund die Theorie linearer Prozesse steht (insbesondere der von Box und Jenkins popularisierten ARMA-Prozesse). Die Schätzung und Anpassung von Modellen bildet den Hauptteil des Buches (Kapitel 5, 6 und 7). Erweiterungen auf nichtlineare Prozesse werden in einem abschließenden Kapitel 8 angesprochen.

(3) **Prognose**, d.h. Vorhersage künftiger Werte des Prozesses aus der Beobachtung einer Zeitreihe. Die Prognose setzt wesentlich die Gültigkeit eines stochastischen Modells für den Prozeß voraus (andernfalls würde man von Wahrsagerei sprechen). Wir diskutieren in Kapitel 4 sowohl die theoretischen Grundlagen (insbesondere die Prädiktionstheorie von Wiener/Kolmogoroff und die Theorie des Kalman-Filters) als auch die daraus ableitbaren praktischen Prognoseverfahren.

(4) **Kontrolle**, d.h. die optimale Steuerung stochastischer Prozesse. Mit Kontrollproblemen befassen wir uns in diesem Buch nicht, da u.E. hier in sehr starkem Maße substanzwissenschaftliche Erfahrungen mit den zu steuernden Systemen eingehen müssen. Die (optimale) Kontrolle nichttechnischer (etwa sozialer oder ökonomischer) Systeme scheint uns aus diesen Gründen kaum realisierbar. Selbst wenn man die Möglichkeit einer solchen Kontrolle unterstellt, dürfte man doch noch Zweifel an deren Wünschbarkeit äußern (quis custodiet custodes?).

Charakteristisch für das Gebiet der Zeitreihenanalyse ist eine durchgehende Dualität zweier Betrachtungsweisen. Eine Zeitreihe (bzw. ein stochastischer Prozeß) kann zum einen im **Zeitbereich** untersucht werden, d.h. als Folge von Zahlen (bzw. Zufallsvariablen). Alternativ dazu existiert jedoch eine aequivalente Darstellung im **Frequenzbereich**, d.h. eine Darstellung als Überlagerung deterministischer oder stochastischer Schwingungen mit verschiedenen Frequenzen. Naturwissenschaftlern ist diese Darstellung aus der Beobachtung zahlreicher Naturphänomene her bekannt. So lassen sich Lichtwellen als Mischung von Wellen verschiedener Frequenzen (Farben) auffassen, die z.B. durch ein Prisma sichtbar gemacht werden können; so zerlegt das menschliche Ohr die Zeitreihe der am Trommelfell registrierten Luftdruckschwankungen in ihre Frequenzkomponenten, wodurch die Wahrnehmung von Tönen möglich wird etc. Die mathematische Technik zur Zerlegung einer gegebenen Reihe in Frequenzkomponenten ist die **Fourieranalyse**, die entspre-

chende Methode der Darstellung stochastischer Prozesse im Frequenzbereich wird als **Spektralanalyse** bezeichnet.

Erfahrungsgemäß erleben viele Nicht-Naturwissenschaftler die Fourieranalyse als wesentliche Hemmschwelle beim Erlernen der Zeitreihenanalyse. Wir wählen daher im vorliegenden Buch einen sehr sanften Zugang zu diesem Thema. Im Kapitel über deskriptive Verfahren wird die Fourieranalyse auf die bereits bekannten Ergebnisse über lineare Regressionsmodelle zurückgeführt, wodurch man eine einfache Interpretation der zentralen Begriffe erhält. In Kapitel 3 betrachten wir als theoretisches Pendant die Spektraltheorie stochastischer Prozesse, das schwierige Problem der Schätzung von Spektren wird ausführlich in Kapitel 7 diskutiert.

Eine mathematisch saubere Darstellung der Zeitreihenanalyse setzt beim Leser Kenntnisse in zahlreichen mathematischen Disziplinen voraus: Funktionentheorie, Maßtheorie, Funktionalanalysis, Algebra, Grenzwertsätze der Wahrscheinlichkeitstheorie, mathematische Statistik. Da sich das Buch vorrangig an Nichtmathematiker wenden soll, war eine bewußte Einschränkung des mathematischen Niveaus notwendig. Dabei ließen wir uns von zwei Grundsätzen leiten: Zum einen sollten keine Begriffe verwendet werden, deren Einführung die mathematische Allgemeinbildung eines Studenten mittleren Semesters wesentlich übersteigen würde. So wurde auf die elegante geometrische Interpretation eines stochastischen Prozesses als Kurve im Hilbertraum ebenso verzichtet wie auf die Anwendung der Cramèr-Darstellung, die ohne eine Theorie der stochastischen Integration nicht sauber begründbar ist. Zudem haben wir in den Anhängen die wichtigsten Grundbegriffe und Aussagen aus den Bereichen trigonometrische Funktionen, komplexe Zahlen und Wahrscheinlichkeitsrechnung zusammengestellt. Diese gehören unserer Erfahrung nach nicht unbedingt zum präsenten Wissen der potentiellen Leser.

Zum zweiten sind u. E. auch für den Anwender Beweise zumindest der wesentlichen Sätze zum wirklichen Verständnis des Gebietes erforderlich. Um den Kern der Beweisidee vor einem Gestrüpp technischer Details hervortreten zu lassen, wurde verschiedentlich auf die mögliche größere Allgemeinheit bei der Formulierung der Sätze verzichtet. Typische Einschränkungen sind etwa die Annahme der absoluten Summierbarkeit gewisser Folgen (wo quadratische Summierbarkeit reichen würde) bzw. die Unterstellung normalverteilter Prozesse bei der Ableitung asymptotischer Aussagen. Sofern uns die weiterreichenden Aussagen wichtig erschienen, wurden diese zusammen mit Literaturstellen angegeben.

Schließlich werden im wesentlichen nur univariate Prozesse behandelt, da dem Leser nach einem solchen Studium der univariaten Theorie die Übertragung auf den multivariaten Fall relativ mühelos gelingen sollte.

Für einen einsemestrigen Kurs enthält das Buch wohl trotz dieser Restriktionen noch zu viel Material. Entsprechend den beiden Schwerpunkten des Buches bieten sich jedoch zwei alternative Programme an: Ausrichtung an der **Spektralanalyse** (etwa: Kap. 1, 2, 3, Kap. 5 – evtl. kursorisch –, Kap. 7) oder Ausrichtung an der Theorie und Praxis linearer Prozesse-**„Box-Jenkins-Ansatz"** (etwa: Abschnitt 1.1–1.5, Kap. 2, Kap. 5 – evtl. kursorisch –, Kap. 6, Teile von Kap. 4, Kap. 8). Denkbar sind jedoch auch noch zahlreiche andere Auswahl- und Anordnungsmöglichkeiten des Stoffes.

Zur Numerierung: innerhalb der kleinsten Unterabschnitte des Buches werden Bei-

spiele, Definitionen und Sätze von 1 bis n durchnumeriert (also Beispiel x.y.z.1, Satz x.y.z.2 etc.), ebenso wird für Gleichungen (etwa [x.y.z.1]) und Abbildungen/Tabellen verfahren (Abb. x.y.z.1, Tab. x.y.z.2).

Die Beschäftigung mit dem faszinierenden Gebiet der Zeitreihenanalyse steht an den statistischen Instituten der Freien Universität Berlin in einer langjährigen Tradition, die von Wolfgang Wetzel begründet und von Peter Naeve u.a. fortgeführt wurde (trotz des Diktums von Hans Münzner und Peter Th. Wilrich daß niemand ohne Not mit einem so schrecklichen Gebiet sich befassen müsse). Wir hoffen, daß wir unseren direkten und indirekten Lehrern mit diesem Buch nicht allzuviel Schande bereiten und erlauben uns, ihnen sowie allen anderen ehemaligen und jetzigen Mitgliedern des Instituts für angewandte Statistik sowie des Instituts für Statistik und Versicherungsmathematik für stete freundliche Bereitschaft zum kritischen Gespräch und die offene und kollegiale Arbeitsatmosphäre an diesen Instituten zu danken. Der Dank gilt ganz besonders Herrn Herbert Buscher und Herrn Andreas Handl sowie Herrn Felix Hampe für die gründliche kritische Durchsicht des Manuskriptes.

Unser Dank gilt nicht zuletzt Herrn Martin Weigert, unserem Ansprechpartner beim Oldenbourg Verlag. Ihm danken wir nicht nur für die ausgezeichnete Betreuung und das Eingehen auf unsere Vorstellungen und Wünsche. Das Kind, das der Leser nunmehr in den Händen hält, hätte schließlich ohne den von ihm beherzt geführten Kaiserschnitt wohl niemals das Licht der Welt erblickt.

1. BESCHREIBUNG VON ZEITREIHEN

1.1 Darstellung von Zeitreihen

In vielen Gebieten der Statistik geht man von Stichproben aus, d. h. von Beobachtungen einer Größe, die unter identischen Bedingungen gewonnen wurden. Ein Beispiel ist etwa die wiederholte Messung einer physikalischen Größe wie der Lichtgeschwindigkeit. Die Reihenfolge, in der die einzelnen Werte aufgetreten sind, spielt dann keine Rolle für die anschließende statistische Analyse.

Hier betrachten wir dagegen ein Gebiet der Statistik, bei dem die Anordnung, in der die Beobachtungen gewonnen wurden, von größter Bedeutung ist. Eine (zeitlich) geordnete Folge $(x_t)_{t \in T}$ von Beobachtungen einer Größe wird als **Zeitreihe** bezeichnet. Für jeden Zeitpunkt t einer Menge T von Beobachtungszeitpunkten liegt dabei genau eine Beobachtung vor.

Zeitreihen treten in allen Bereichen der Wissenschaft auf – wir werden u. a. folgende Beispiele betrachten:

Astronomie: Messung der Lichtstärke eines pulsierenden Sterns in 600 aufeinanderfolgenden Nächten.

Chemie: Untersuchung des Zeitverhaltens eines chemischen Analyseautomatens (an Hand von Messungen im „Dauerlauf" des Automatens).

Medizin: Hormonausschüttungen bei Affen (viertelstündliche Messungen).

Ökonomie: Aktienkurse an aufeinanderfolgenden Börsentagen.

Ingenieurwissenschaften: Jahreshöchstlasten stromproduzierender Unternehmen.

Soziologie: Bevölkerungsentwicklung der USA (Jahreswerte von 1790–1970).

In vielen Zeitreihen ist die sogenannte **Parametermenge** T eine endliche, diskrete Menge von gleichabständigen Zeitpunkten. Man numeriert dann in der Regel die Zeitpunkte durch und setzt $T = \{1, 2, \ldots, N\}$. Zahlreiche der im folgenden dargestellten Verfahren sind jedoch auch auf den Fall übertragbar, daß die Beobachtungen zu unregelmäßigen Zeitpunkten erhoben wurden (vgl. etwa Kapitel 1.3).

Liegen kontinuierliche Messungen vor (etwa die Zeitfunktion $x(t)$ der Spannung in einem elektronischen Bauteil für ein Zeitintervall $a \leqq t \leqq b$), so wird für die Analyse i. a. eine Diskretisierung vorgenommen. Daher sind die hier dargestellten Methoden auch für solche Situationen relevant.

Für theoretische Betrachtungen ist es von Nutzen, auch unendliche Zeitreihen zuzulassen. Dann steht T für die Menge $\mathbb{N}$ der natürlichen oder für die Menge $\mathbb{Z}$ der ganzen Zahlen.

BEISPIEL 1.1.1

Die Abbildung 1.1.1 zeigt die Entwicklung der Konkurse von Unternehmen in den USA von 1867 bis 1932. Die x_t sind die prozentualen Anteile der in Konkurs gegangenen Unternehmen an den Unternehmen insgesamt.

■

Obwohl wir nur Zeitreihen mit diskreter Parametermenge betrachten, wird der Graph wie unten in der Regel kontinuierlich dargestellt.

Da bestimmte **Beispielreihen** im folgenden immer wieder aufgegriffen werden,

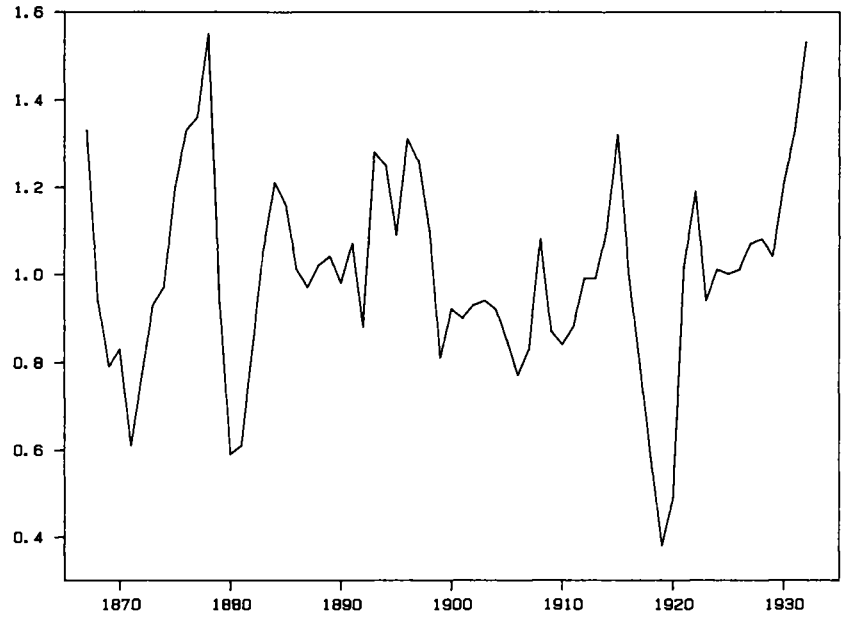


Abb. 1.1.1 *KONKURSE in den USA, (Quelle: B. Greenstein [1935])*

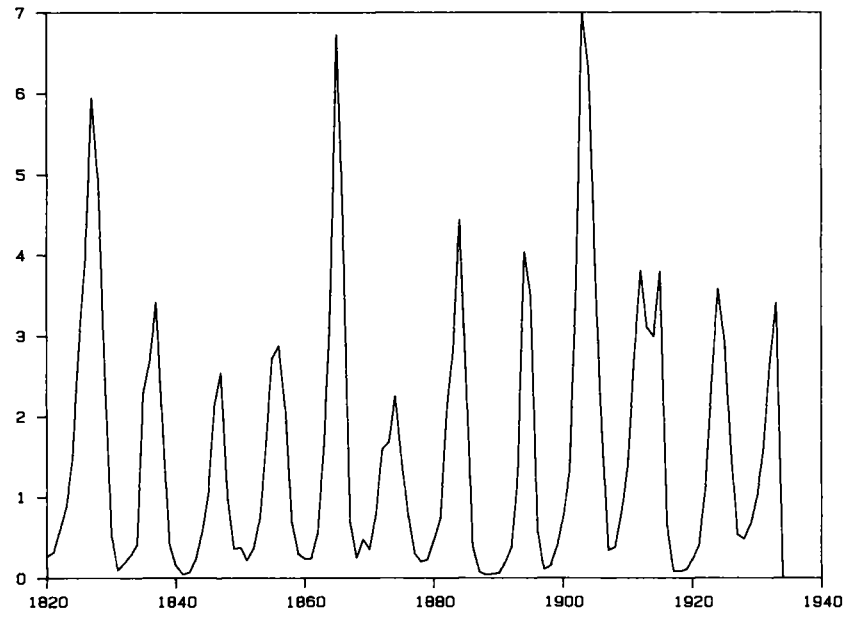


Abb. 1.1.2 *Anzahl gefangener Luchse (Quelle: M. J. Campbell and A. M. Walker [1977])*

wollen wir sie mit **Kurzbezeichnungen** versehen. Die abgebildete Reihe bezeichnen wir mit dem Kürzel KONKURSE. Alle Beispielsreihen sind im Anhang wiedergegeben, um dem Leser die Möglichkeit eigener Analysen zu verschaffen.

BEISPIEL 1.1.2

Ein klassisches Beispiel ist die Reihe LUCHS. Es handelt sich um die Anzahl der Luchse, die jährlich in einem bestimmten Gebiet Kanadas gefangen wurden (in 1000). Siehe Abb. 1.1.2.

■

Die graphische Darstellung – **der Plot** – der Zeitreihe ist in der Regel der erste Schritt bei der Analyse von Zeitreihen. Aus dem Plot lassen sich häufig die ersten Charakteristika ersehen. So erkennt man im Beispiel 1.1.2, daß etwa alle 10 Jahre regelmäßig sehr viele Luchse gefangen wurden und dazwischen jeweils recht wenige. Die Reihe im Beispiel 1.1.1 zeigt dagegen kein offensichtlich regelmäßiges Verhalten.

1.2 Empirische Momente

Nach dem Plot der Zeitreihe bietet sich als zweiter Schritt bei der Analyse an, geeignete statistische Kenngrößen zu berechnen. Die wesentlichen Kenngrößen sind die Momente erster und zweiter Ordnung. Sie bilden die Grundlage für die statistische Analyse von Zeitreihen, sind aber auch im Rahmen der Beschreibung von Interesse.

Sinnvoll ist die Berechnung dieser Größen nur für sogenannte **stationäre** Reihen. Als stationär bezeichnen wir dabei vorerst eine Reihe, die – grob gesprochen – keine systematischen Veränderungen im Gesamtbild aufweist. Stationarität beinhaltet, daß Kennziffern, die aus verschiedenen Teilreihen berechnet werden, nicht zu stark voneinander abweichen. Es ist in der Zeitreihenanalyse häufig anzuraten, die beobachtete Reihe in nicht zu kurze Segmente zu zerlegen, für die dann die jeweiligen Analyseschritte einzeln durchgeführt werden. Dies ermöglicht die praktische Überprüfung der Stationaritätsannahme. Die theoretische Bedeutung der Stationarität kann jedoch erst im zweiten Kapitel aufgezeigt werden.

Die zentrale Lage der Werte einer Zeitreihe $(x_t)_{t=1,\dots,N}$ wird durch das **arithmetische Mittel**

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

beschrieben. $\bar{x}$ ist der mittlere Wert, um den die Beobachtungen schwanken. Die Stärke der Schwankung wird gemessen durch die empirische **Varianz**,

$$s^2 = \frac{1}{N} \sum_{t=1}^{N} (x_t - \bar{x})^2,$$

oder durch die **Standardabweichung** $s = \sqrt{s^2}$, die dieselbe Maßeinheit hat wie das arithmetische Mittel.

Eine zentrale Stellung bei der Untersuchung von Zeitreihen nimmt die Frage nach den Abhängigkeiten zwischen verschiedenen Zeitpunkten ein. Unter den ver-

schiedenen Abhängigkeitsformen hat die lineare Abhängigkeit bei der Zeitreihenanalyse die weitaus größte Bedeutung erlangt. Wir gehen kurz auf die klassischen Maßzahlen zur Beschreibung des linearen Zusammenhangs ein und übertragen diese dann auf die vorliegende Situation.

Für N Beobachtungspaare (x_i, y_i) ist die empirische **Kovarianz**

$$c = \frac{1}{N} \sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y})$$

ein Maß für die Stärke des linearen Zusammenhanges zwischen den x- und den y-Werten.

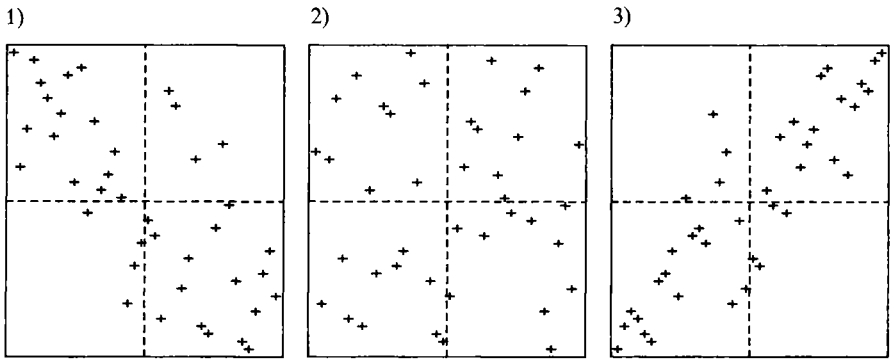


Abb. 1.2.1 *Streudiagramme und Kovarianzen von Beobachtungspaaren*
1) $r = -0.72$, $c < 0$ 2) $r = -0.01$, $c \approx 0$ 3) $r = 0.87$, $c > 0$

Bei 2) heben sich die positiven Werte des 1. und 3. Quadranten gegen die negativen des 2. und 4. Quadranten in etwa auf. c hat einen nahe bei 0 gelegenen Wert. Bei 3) überwiegen die positiven, bei 1) die negativen Summanden.

Durch die Normierung mit dem Produkt der einzelnen Standardabweichungen erhält man den **Korrelationskoeffizienten von Bravais-Pearson**:

$$r = \frac{\frac{1}{N} \sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\frac{1}{N} \sum_{i=1}^{N} (x_i - \bar{x})^2} \cdot \sqrt{\frac{1}{N} \sum_{i=1}^{N} (y_i - \bar{y})^2}}$$

Wendet man die Cauchy-Schwarz'sche Ungleichung

$$(\Sigma a_i b_i)^2 \leqq (\Sigma a_i^2)(\Sigma b_i^2) \qquad [1.2.1]$$

auf die Abweichungen $x_i - \bar{x}$ und $y_i - \bar{y}$ an, so erkennt man daß $-1 \leqq r \leqq 1$ gilt, d. h., daß der Korrelationskoeffizient betragsmäßig höchstens gleich 1 sein kann.

Werte von r, die nahe bei $+1$ (-1) liegen, deuten auf einen starken positiven (negativen) linearen Zusammenhang hin, Werte bei Null auf das Fehlen eines linearen Zusammenhangs oder Unkorreliertheit. Siehe dazu die obige Abbildung. Die Extremwerte ± 1 werden genau dann angenommen, wenn die Beobachtungspaare eine exakte lineare Beziehung erfüllen.

Um die Stärke des linearen Zusammenhanges aufeinanderfolgender Beobachtungen bei Zeitreihen zu messen, brauchen wir c bzw. r nur leicht zu modifizieren. Aus den N Beobachtungen $x_1, x_2, \ldots, x_N$ werden die $N-1$ Paare direkt aufeinanderfolgender Beobachtungen gebildet:

$$(x_1, x_2), (x_2, x_3), \ldots, (x_{N-1}, x_N).$$

Diese werden als Beobachtungspaare aufgefaßt; deren Kovarianz ist gleich:

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

Dabei sind $\bar{x}_{(1)}, \bar{x}_{(2)}$ die arithmetischen Mittel aus den jeweiligen ersten bzw. zweiten Komponenten.

Entsprechend können die Kovarianzen und Korrelationskoeffizienten für weiter auseinanderliegende Beobachtungen berechnet werden. Da sich die verschiedenen arithmetischen Mittel und Standardabweichungen wegen der Stationaritätsannahme i.a. wenig unterscheiden, setzt man generell das allgemeine Mittel $\bar{x}$ und die Standardabweichung s ein,

Damit ergeben sich die Kovarianzen

$$c_\tau = \frac{1}{N} \sum_{t=1}^{N-\tau} (x_t - \bar{x})(x_{t+\tau} - \bar{x}) \quad \tau = 0, 1, 2, \ldots, N-1.$$

c_0 ist gerade die Varianz s^2 der Zeitreihe. Die Division durch N anstelle von $N-\tau$ hat statistische Vorteile, wie wir in Kapitel 6 erkennen werden. Für große N ergeben sich keine wesentlichen Unterschiede, da c_τ meist nur für $\tau \leqq N/4$ betrachtet wird. Man beachte, daß mit wachsendem τ die Zahl der Beobachtungspaare, die in die Berechnung von c_τ eingehen, immer kleiner wird, wodurch die Aussagekraft zunehmend eingeschränkt wird.

Da die Beobachtungspaare mit dem Zeitabstand τ auch in der Form $(x_t, x_{t-\tau})_{t=\tau+1,\ldots,N}$ angegeben werden können, ist es sinnvoll, c_τ für negative Werte von τ durch $c_\tau := c_{-\tau}$ zu definieren.

Für die Korrelationskoeffizienten findet man analog

$$r_\tau = \frac{\sum_{t=1}^{N-|\tau|} (x_t - \bar{x})(x_{t+|\tau|} - \bar{x})}{\sum_{t=1}^{N} (x_t - \bar{x})^2} = \frac{c_\tau}{c_0}.$$

Auch für den so definierten Korrelationskoeffizienten gilt

$$-1 \leqq r_\tau \leqq 1,$$

wie man sich leicht mit Hilfe der Cauchy-Schwarz'schen Ungleichung [1.2.1] klarmacht.

BEISPIEL 1.2.1

Wir illustrieren die Berechnung von c_τ und r_τ an einem überschaubaren Zahlenbeispiel. Sei eine hypothetische Zeitreihe BEI10 mit $N = 10$ Beobachtungen gegeben.

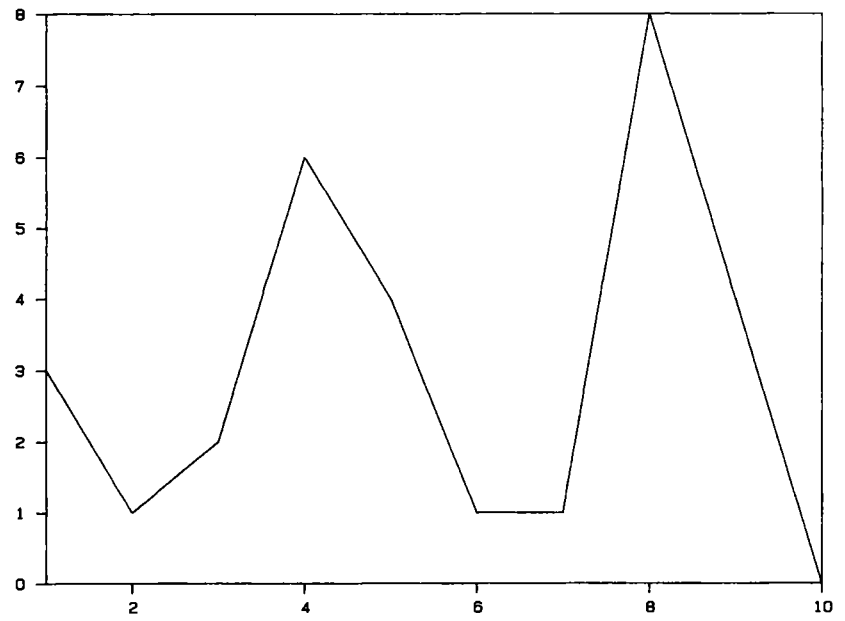


Abb. 1.2.2 *Hypothetische Zeitreihe BEI10*

Um etwa c_2 zu berechnen, wird die Reihe der Abweichungen $(x_t - \bar{x})$ um zwei Zeitpunkte verschoben und die Summe der Produkte untereinanderstehender Werte berechnet:

t	1	2	3	4	5	6	7	8	9	10	Σ
$x_t - \bar{x}$	0	−2	−1	3	1	−2	−2	5	1	−3	–
$x_{t+2} - \bar{x}$	−1	3	1	−2	−2	5	1	−3	–	–	–
$(x_t - \bar{x})(x_{t+2} - \bar{x})$	0	−6	−1	−6	−2	−10	−2	−15	–	–	−42

Man erhält $c_2 = -42/10 = -4.2$.

Der Leser möge die Berechnung für andere Zeitabstände τ analog nachvollziehen:

τ	0	1	2	3	4	5	6	7	8	9
c_τ	5.80	−0.40	−4.20	0.30	2.80	−0.10	−2.00	0.10	0.60	0.00
r_τ	1.00	−0.07	−0.72	0.05	0.48	−0.02	−0.34	0.02	0.10	0.00

■

Um lineare Abhängigkeiten von Werten zwischen den Zeitpunkten innerhalb einer Zeitreihe von den linearen Beziehungen zwischen verschiedenen Zeitreihen sprachlich zu differenzieren, verwendet man die Bezeichnungen Autokovarianzen und -korrelationen bzw. Kreuzkovarianzen und -korrelationen.

DEFINITION 1.2.2

Die **Autokovarianzfunktion** (c_τ) einer Zeitreihe ist die durch

$$c_\tau = \frac{1}{N} \sum_t (x_t - \bar{x})(x_{t+\tau} - \bar{x})$$

definierte Funktion des Zeitabstands oder **Lags** $\tau = -(N-1), \ldots, -1, 0, 1, \ldots, (N-1)$. Die **Autokorrelationsfunktion** (r_τ) ist durch $r_\tau = c_\tau / c_0$ gegeben. Der Graph dieser Funktion heißt **Korrelogramm**. ■

Das Korrelogramm ist von erheblicher Bedeutung, da es die wesentlichen Informationen über zeitliche Abhängigkeiten in der beobachteten Reihe enthält. Es wird daher im allgemeinen als zweites nach der Zeitreihe gezeichnet. Da sich r_τ und c_τ nur um den konstanten Faktor c_0 unterscheiden, genügt es, eine der beiden Funktionen zu zeichnen. Die beiden folgenden Beispiele illustrieren die üblichen Darstellungsweisen dieser Funktionen.

BEISPIEL 1.2.3

Das Korrelogramm der Reihe der KONKURSE läßt vermuten, daß besonders ausgeprägte Abhängigkeiten im Rhythmus von etwa 4 bis 5 Jahren auftreten. Denn in diesem Abstand nimmt die Autokorrelationsfunktion (absolut gesehen) besonders große Werte an.

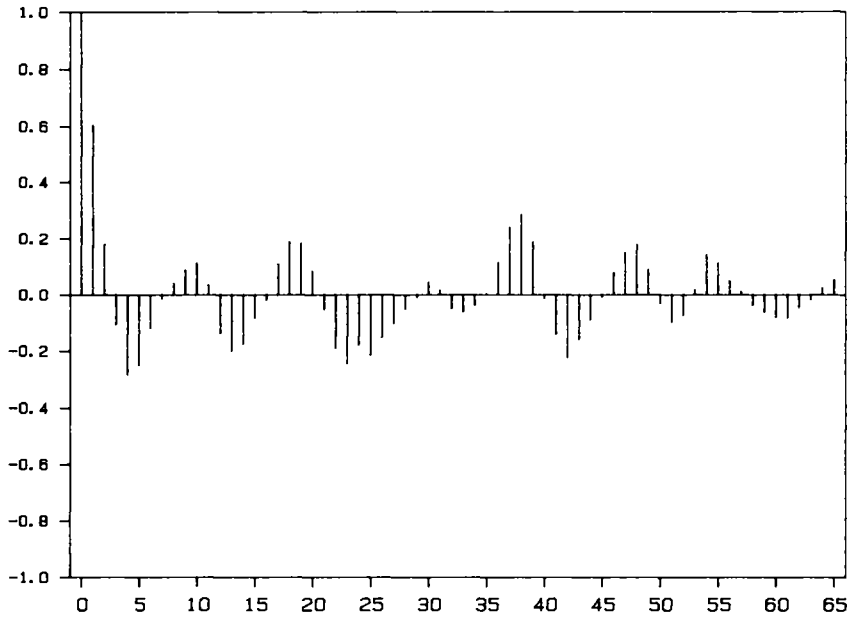


Abb. 1.2.3 *Autokorrelationsfunktion der Reihe KONKURSE*

■

BEISPIEL 1.2.4

Die Autokorrelationsfunktion der Reihe LUCHS zeigt starke positive Abhängigkeiten für Lags um $\tau = 10, 20, \ldots$ an. Da die Zeiträume zwischen guten und schlechten Fangerträgen jeweils etwa 5 Jahre auseinanderliegen, ist die Autokorrelationsfunktion negativ für Lags um $\tau = 5, 15, \ldots$.

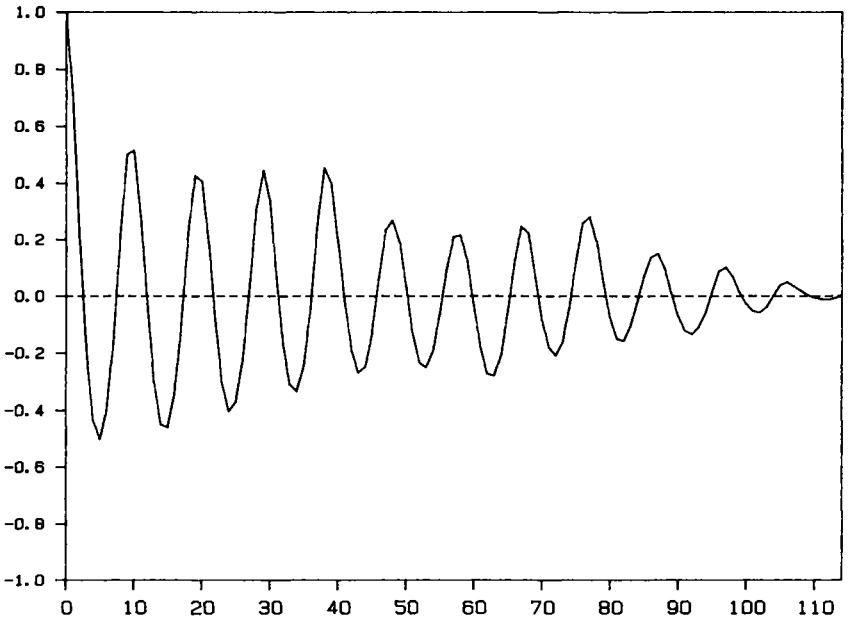


Abb. 1.2.4 *Autokorrelationsfunktion der Reihe LUCHS*

■

Abschließend sei vor der unkritischen Verwendung der angegebenen Maßzahlen gewarnt. In Größen $\bar{x}$, s^2, c_τ und r_τ geht ein einzelner Beobachtungswert x_t mit großem Gewicht ein, wenn er weit vom Durchschnitt der anderen Beobachtungen abweicht. Diese Maßzahlen sind daher recht anfällig („nichtrobust") gegenüber dem Auftreten von Ausreißern oder untypischen Werten in der Zeitreihe. In der Praxis werden verschiedene heuristische Vorgehensweisen zur **Ausreißerbeseitigung** angewendet: Ersetzen eines Ausreißers durch das arithmetische Mittel aller oder der beiden benachbarten Werte bzw. ‚Heranschieben' des Ausreißers an die übrigen Werte (‚Winsorisieren'). Dadurch kann die Analyse einer Zeitreihe jedoch wesentlich und in unkontrollierbarer Weise beeinflußt werden.

In neuerer Zeit wird an der Entwicklung robuster statistischer Verfahren gearbeitet. Im Bereich der Zeitreihenanalyse steckt diese Entwicklung jedoch noch in den Anfängen (vgl. Martin [1980]).

Darüber hinaus wird die Interpretation speziell des Korrelogramms dadurch erschwert, daß die Größen r_τ nicht unabhängig voneinander sind, da jeweils die gleichen Beobachtungen in die Berechnung eingehen. Dramatisch drückt dies ein Zitat von John W. Tukey [1967] aus: „*I have frequently said that the tragic accident*

that killed H. R. Sewell and his family destroyed the only man, who could usefully look at a plot of autocorrelation against lag. Factually such a statement cannot, of course, be true, yet the impression it gives about the usefulness of autocorrelation curves is quite precise and very helpful".

1.3 Das klassische Komponentenmodell

Im vorigen Abschnitt wurde unterstellt, daß die betrachteten Zeitreihen „stationär" sind. Diese Forderung ist häufig nicht erfüllt. Vor allem bei ökonomischen Reihen findet man langfristige Veränderungen oder eine sich jährlich wiederholende Saisonfigur. So ist die Zahl der Arbeitslosen regelmäßig im Winter größer als im Sommer, ebenso hängt etwa der Stromverbrauch von der Jahreszeit ab. Für ökonomische Zeitreihen wurden demgemäß Modelle vorgeschlagen, die diesen Gegebenheiten Rechnung tragen. Sie gehen von folgenden vier Komponenten aus:

(i) dem **Trend**, das ist eine langfristige systematische Veränderung des mittleren Niveaus der Zeitreihe;

(ii) einer **Konjunkturkomponente**, die eine mehrjährige, nicht notwendig regelmäßige Schwankung darstellt;

(iii) der **Saison**, das ist eine jahreszeitlich bedingte Schwankungskomponente, die sich relativ unverändert jedes Jahr wiederholt;

(iv) der **Restkomponente**, die die nicht zu erklärenden Einflüsse oder Störungen zusammenfaßt.

Die ersten beiden Komponenten werden bisweilen zu einer einzigen, der sogenannten **glatten Komponente** zusammengefaßt. Alternativ faßt man manchmal die Komponenten (ii) und (iii) zur sogenannten **zyklischen Komponente** zusammen.

BEISPIEL 1.3.1

Die Monatswerte der Arbeitslosen in Westberlin vom Januar 1967 bis Oktober 1979 weisen sowohl einen Trend auf als auch den deutlichen Einfluß der Jahreszeit in Form der Saison, vgl. Abb. 1.3.1. Offensichtlich ist hier die Trennung der Trendkomponente und der Konjunkturkomponente schon problematisch. Dies zeigt sich noch deutlicher, wenn die Arbeitslosigkeit über einen längeren Zeitraum hinweg betrachtet wird. Aussagen über die Restkomponente sind erst möglich, wenn die Bestimmung der glatten und Saisonkomponente erfolgt ist.

■

Häufig wird unterstellt, daß sich die Komponenten additiv überlagern. Dann erhält man die folgende formale Beschreibung des **additiven Modells**:

$$\text{Reihe} = \underbrace{\text{Trend}}_{\text{glatte Komponente}} + \overbrace{\underbrace{\text{Konjunktur}}_{\text{zyklische Komponente}} + \text{Saison}} + \text{Rest}$$

$$x_t = \underbrace{m_t}_{g_t} + \overbrace{\underbrace{k_t}_{z_t} + s_t} + u_t .$$

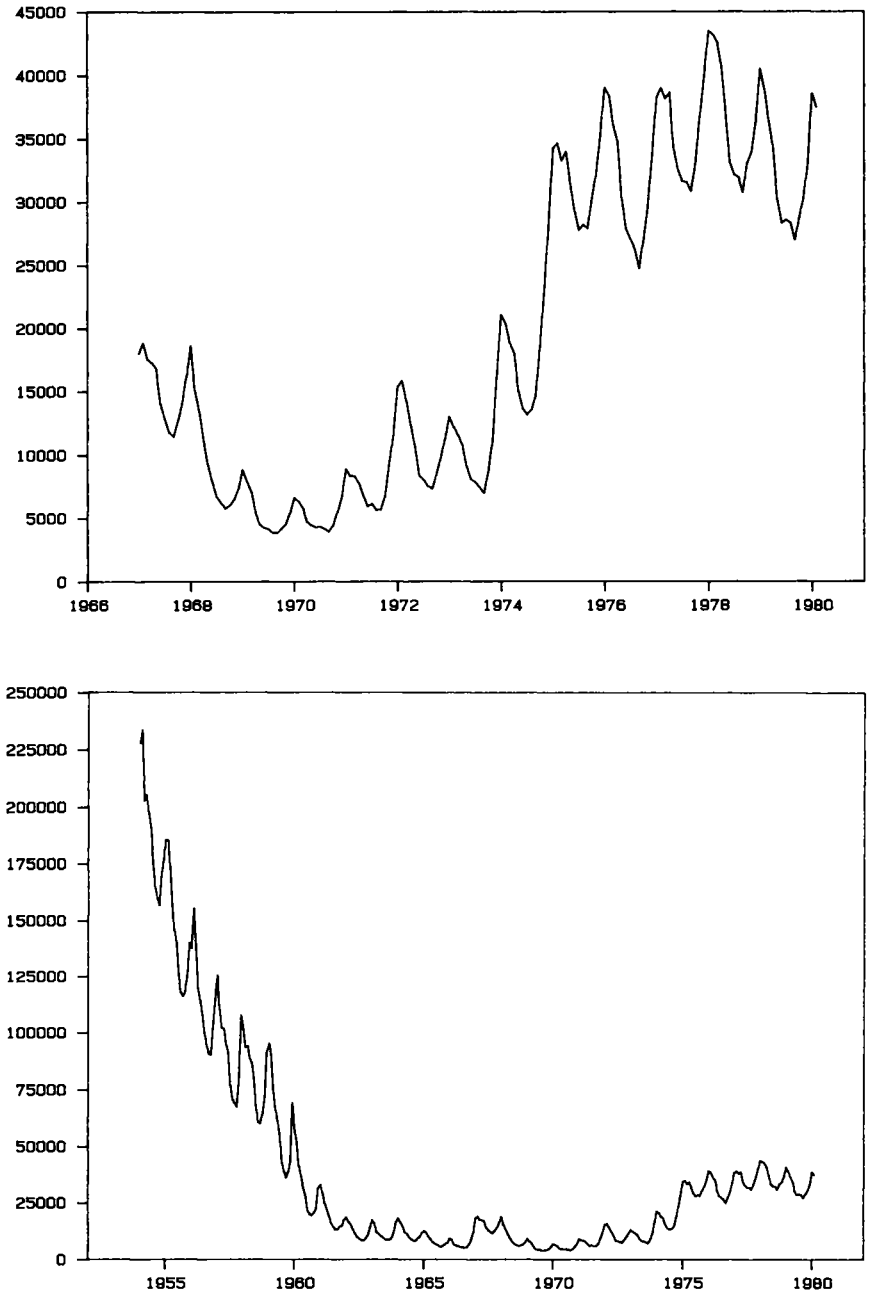


Abb. 1.3.1 a) *Reihe ALBERLIN der Arbeitslosen in Westberlin 1967–1980*
b) *ALBERLIN 1954–1980*

Quelle: Berliner Statistik, verschiedene Jahrgänge

Man findet in empirischen Reihen nicht selten, daß mit der glatten Komponente auch die Streuung der Werte um die glatte Komponente ansteigt. Bei solchen Reihen wird oft ein **multiplikatives Modell** unterstellt:

$$x_t = g_t \cdot s_t \cdot u_t .$$

Dieses kann durch die Bildung von Logarithmen auf ein additives zurückgeführt werden:

$$\begin{aligned} \log(x_t) &= \log(g_t \cdot s_t \cdot u_t) \\ &= \log(g_t) + \log(s_t) + \log(u_t) . \end{aligned}$$

BEISPIEL 1.3.2

Die Zeitreihe STROM des monatlich in der BRD produzierten Stromes zeigt neben einem deutlichen Trend auch das Vorhandensein einer Saisonkomponente. Zum einen spielt hier die Jahreszeit eine wesentliche Rolle. Im Winterhalbjahr wird mehr Strom verbraucht und dementsprechend produziert. Jedoch trägt auch die unterschiedliche Länge der Kalendermonate zu dem Bild der Saisonfigur bei. Im Februar wird stets weniger Strom produziert als in den beiden Nachbarmonaten. Schließlich zeigt der Plot der Reihe, daß hier ein multiplikatives Modell angebracht ist.

Das Diagramm 1.3.3 illustriert das Resultat einer Komponentenzerlegung.

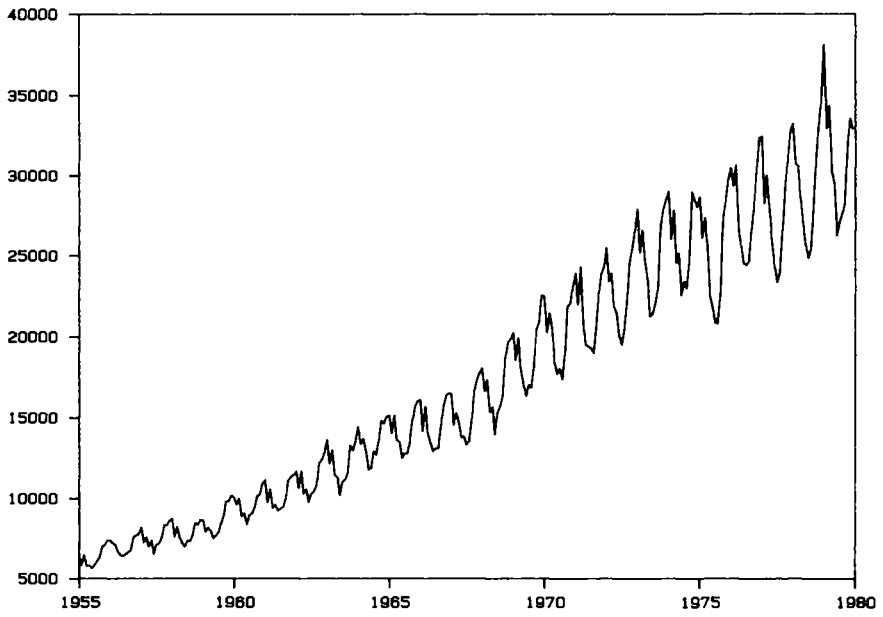


Abb. 1.3.2 *Produzierter STROM in der BRD (Mio kWh)*
(Quelle: Wirtschaft und Statistik, verschiedene Jahrgänge)

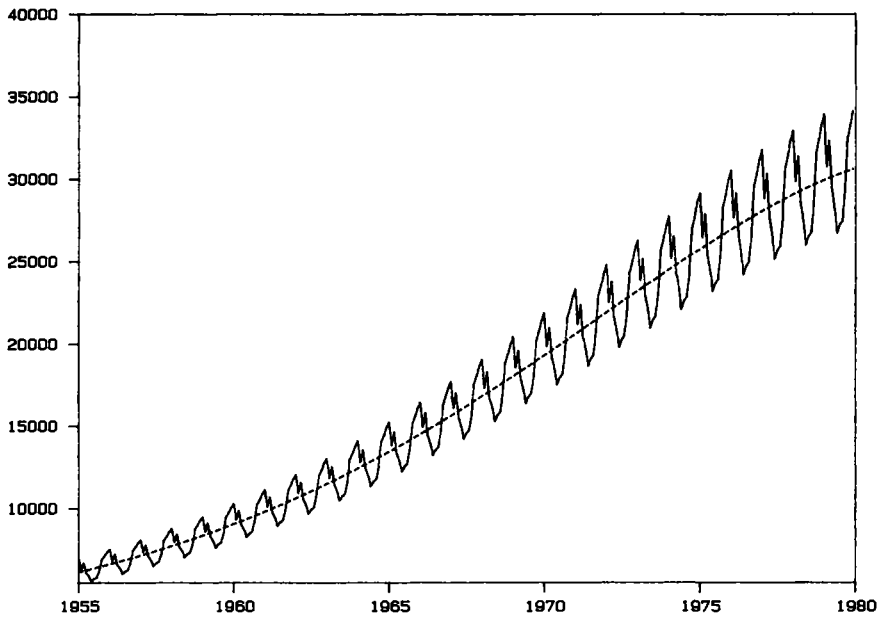


Abb. 1.3.3 *Geschätzte glatte Komponente* ($\hat{g}_t$) *sowie Schätzung von Trend* × *Saison* ($\hat{g}_t \cdot \hat{s}_t$)

■

Das Komponentenmodell ist unbestimmt, solange nicht die einzelnen Komponenten präziser durch Modelle spezifiziert werden. Verschiedene Modellansätze können dabei sehr unterschiedliche Ergebnisse liefern, so daß bei der Interpretation veröffentlichter Reihenkomponenten (etwa saisonbereinigter Arbeitslosenziffern) stets nach dem jeweils benutzten Verfahren zu fragen ist. Die klassischen Verfahren basieren im wesentlichen auf zwei unterschiedlichen Ansätzen:

(1) **globalen Komponentenmodellen**, die für die gesamte Reihe ein lineares oder nichtlineares **Regressionsmodell** unterstellen, dessen Parameter mit Hilfe der Kleinstquadratemethode geschätzt werden;

(2) **lokalen Komponentenmodellen**, die derartige Modelle nur stückweise unterstellen, wobei die Komponenten im allgemeinen durch sogenannte **Filter** extrahiert werden.

Beide Methoden werden zunächst im Rahmen der Bestimmung von Trendkomponenten illustriert.

1.4 Trendbestimmung

Verschiedene Fragestellungen führen auf das Problem der Trendbestimmung. Bei der Analyse von zeitlichen Verläufen ökonomischer Größen oder bei Prognosen sucht man mit dem Trend die langfristige Entwicklung wiederzugeben. Weiter ist z. B. für die Untersuchung der Abhängigkeiten zwischen den Zeitpunkten einer Zeitreihe gegebenenfalls die Trendkomponente zu eliminieren, oder wie man sagt,

eine **Trendbereinigung** vorzunehmen. Bei Zeitreihen, die einen Trend aufweisen, ist die Autokorrelationsfunktion nicht geeignet, die Abhängigkeiten zwischen den Zeitpunkten zu beschreiben. Bei der Herleitung wurde ja wesentlich benutzt, daß die Mittel aus den ersten bzw. letzten Daten in etwa dem Gesamtmittel entsprechen.

Für den hier zu diskutierenden Ansatz unterstellen wir, daß die Zeitreihe sich im wesentlichen durch eine Funktion m_t der Zeit beschreiben läßt und daß die Beobachtung x_t nur durch Zufallseinflüsse oder Meßfehler u_t von m_t abweichen. Dabei wird angenommen, daß x_t sich additiv aus m_t und u_t zusammensetzt:

$$x_t = m_t + u_t, \quad t = 1, 2, \ldots, N.$$

Für die Zufallseinflüsse u_t wird unterstellt, daß es sich um Realisierungen unabhängiger Zufallsvariablen U_t handelt, deren Erwartungswerte gleich 0 sind und deren Varianz im Zeitablauf konstant bleibt (vgl. zur Erklärung der Begriffe den Anhang *C* über Wahrscheinlichkeitstheorie). Eine kritische Diskussion dieser Annahmen findet sich im Abschnitt 1.4.3.

Im Einzelfall stellt sich die Frage, welche Funktion den Trend am besten beschreibt. Hat man aus substanzwissenschaftlichen Gründen heraus eine Vorstellung von der Gestalt der Funktion, so sind eventuell unbekannte Parameter zu ermitteln. Andernfalls ist der Trend durch eine geeignete Funktion zu approximieren. Wir behandeln in den folgenden Unterabschnitten die Aufgabe, bei speziellen Funktionsklassen die geeignete Trendfunktion auszuwählen.

1.4.1 Lineare Trends

Die einfachste Trendfunktion ist eine Gerade:

$$m_t = \beta_1 + \beta_2 t.$$

Will man die geeignete Trendgerade für eine gegebene Zeitreihe auswählen, so benötigt man ein Auswahlkriterium. Dazu bietet sich das **Kleinst-Quadrate-Kriterium** an:

Ausgehend von den Differenzen

$$u_t = x_t - m_t = x_t - \beta_1 - \beta_2 t, \quad t = 1, 2, \ldots, N$$

wird die Gerade gewählt, welche die Summe Q der quadrierten Differenzen minimal werden läßt. Formal können wir dies beschreiben durch

$$Q = \sum_{t=1}^{N} (x_t - m_t)^2 \stackrel{!}{=} \min$$

bzw.

$$Q = \sum_{t=1}^{N} (x_t - \beta_1 - \beta_2 t)^2 \stackrel{!}{=} \min_{\beta_1, \beta_2}.$$

Zur Bestimmung der Schätzungen $\hat{\beta}_1$ und $\hat{\beta}_2$ für die Parameter β_1 und β_2 wird Q als Funktion in β_1 und β_2 aufgefaßt. Notwendig dafür, daß Q in $\hat{\beta}_1$, $\hat{\beta}_2$ ihr Minimum annimmt, ist das Verschwinden der ersten partiellen Ableitungen

$$\frac{\partial}{\partial \beta_1} \sum_{t=1}^{N} (x_t - \beta_1 - \beta_2 t)^2, \quad \frac{\partial}{\partial \beta_2} \sum_{t=1}^{N} (x_t - \beta_1 - \beta_2 t)^2$$

an der Stelle $\hat{\beta}_1$, $\hat{\beta}_2$.

Durch Ausführen der Differentiation und Nullsetzen der Ableitungen erhalten wir die Normalgleichungen

$$\sum_{t=1}^{N} (x_t - \hat{\beta}_1 - \hat{\beta}_2 t) = 0$$
$$\sum_{t=1}^{N} t(x_t - \hat{\beta}_1 - \hat{\beta}_2 t) = 0 .$$

Die Auflösung der Gleichungen ergibt

$$\hat{\beta}_2 = \frac{\overline{tx} - \bar{t} \cdot \bar{x}}{\overline{t^2} - (\bar{t})^2}, \quad \hat{\beta}_1 = \bar{x} - \hat{\beta}_2 \bar{t}. \qquad [1.4.1.1]$$

Dabei sind $\overline{tx} = \frac{1}{N} \sum_{t=1}^{N} t x_t$, $\bar{t} = \frac{1}{N} \sum_{t=1}^{N} t$ und $\overline{t^2} = \frac{1}{N} \sum_{t=1}^{N} t^2$.

Wie man zeigen kann, nimmt die Funktion $\sum (x_t - \beta_1 - \beta_2 t)^2$ an dieser Stelle tatsächlich ihr Minimum an. Die gesuchte Trendgerade ist daher gegeben durch

$$\hat{m}_t = \hat{\beta}_1 + \hat{\beta}_2 t ,$$

wobei $\hat{\beta}_1$ und $\hat{\beta}_2$ gemäß [1.4.1.1] bestimmt werden.

BEISPIEL 1.4.1.1

Die bestimmende Größe für die Kapazität von Kraftwerken ist der jährliche Stromspitzenbedarf, auch Jahreshoechstlast genannt. Dies ist die höchste Leistungsinanspruchnahme sämtlicher Stromverbraucher, welche zu einem bestimmten Zeitpunkt während eines Jahres anfällt. Die Tabelle im Anhang gibt die Reihe JLAST

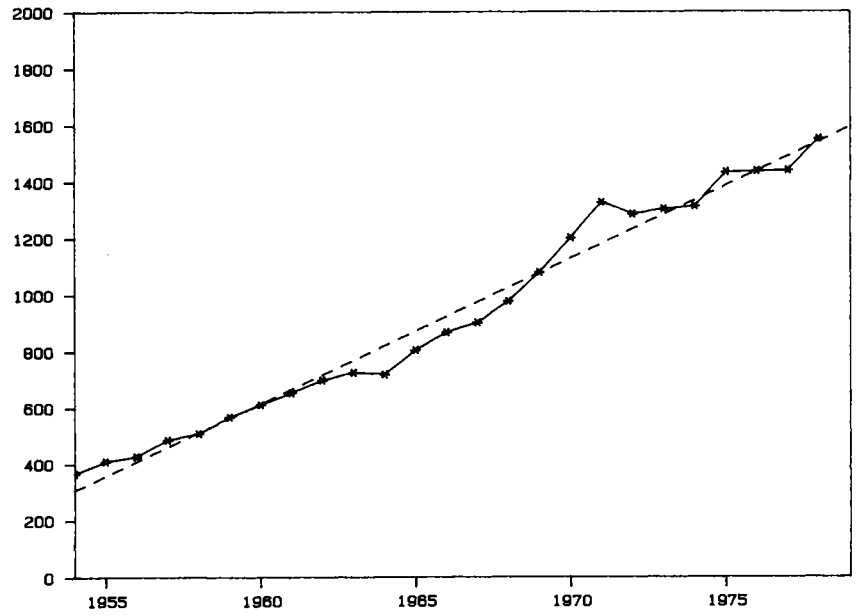


Abb. 1.4.1.1 *Jahreshöchstlasten und angepaßte Trendgerade*

der Jahreshoechstlasten in Berlin(West) für die Jahre 1954/55 bis 1978/79 wieder. (Jeweils von Juli bis Juni, entsprechend dem Geschäftsjahr der BEWAG.)

Dieser Zeitreihe soll ein linearer Trend angepaßt werden.

Werden die Jahre von $t = 1$ bis 25 durchnumeriert, so erhalten wir aus den angegebenen Werten:

$$\bar{t} = \frac{1}{N}\sum_{t=1}^{N} t = \frac{1}{N}\,\frac{N(N+1)}{2} = 13,$$

$$\overline{t^2} = \frac{1}{N}\cdot\sum_{t=1}^{N} t^2 = \frac{1}{N}\,\frac{N(N+1)(2N+1)}{6} = 221,$$

$$\bar{x} = 924.04, \quad \overline{tx} = 14688.12.$$

Damit folgt

$$\hat{\beta}_1 = 255.14, \quad \hat{\beta}_2 = 51.45.$$

Die vorstehende Abbildung gibt die Zeitreihe (x_t) und die geschätzte Trendgerade wieder:

$$m_t = 255.14 + 51.45 \cdot t.$$

■

1.4.2 Lineare Regressionsmodelle zur Trendbestimmung

Die eben beschriebene Kleinstquadratemethode zur Anpassung einer Trendgeraden kann ohne Mühe auf weit allgemeinere Trendmodelle übertragen werden. Das sogenannte lineare **Regressionsmodell** geht von k bekannten Funktionen $m_1(t), m_2(t), \ldots, m_k(t)$ der Zeit aus und unterstellt als Trendmodell eine Linearkombination

$$m(t) = \beta_1 m_1(t) + \beta_2 m_2(t) + \ldots + \beta_k m_k(t) \qquad [1.4.2.1]$$

dieser Funktionen, wobei die Koeffizienten $\beta_1, \beta_2, \ldots, \beta_k$ so gewählt werden sollen, daß sich eine möglichst gute Anpassung an die beobachtete Reihe ergibt.

Aus dem allgemeinen Modell entsteht z. B. das Modell eines linearen Trends durch Setzen von $m_1(t) = 1, m_2(t) = t$. Falls kein einfacher linearer Trend unterstellt werden kann, bieten sich als nächsteinfache Funktionen **Polynome** an, da sich genügend glatte Funktionen in einem endlichen Zeitintervall gut durch Polynome approximieren lassen.

Ein polynomialer Trend $(k-1)$'ter Ordnung

$$m(t) = \beta_1 + \beta_2 t + \beta_3 t^2 + \ldots + \beta_k t^{k-1}$$

entsteht aus dem linearen Regressionsmodell durch die Wahl

$$m_1(t) = 1, m_2(t) = t, \ldots, m_k(t) = t^{k-1}.$$

Während ein solches Trendmodell oft gut zur Beschreibung der beobachteten Reihe geeignet ist, muß erfahrungsgemäß jedoch vor einer Trendextrapolation (Prognose) mittels polynomialer Funktionen gewarnt werden, da Polynome außerhalb des Anpassungsbereichs rasch nach $\pm\infty$ gehen.

Zur Schätzung der Parameter benutzen wir wieder das Kleinstquadratekriterium, d.h. wir suchen $\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_k$ so, daß die Summe der quadrierten Abweichungen von Daten und Trend

$$Q = \sum_{t=1}^{N} (x_t - m(t))^2 = \sum_{t=1}^{N} (x_t - \beta_1 m_1(t) - \beta_2 m_2(t) - \ldots - \beta_k m_k(t))^2 \quad [1.4.2.2]$$

möglichst klein wird. Bildet man die partiellen Ableitungen von Q nach den Koeffizienten $\beta_1, \beta_2, \ldots, \beta_k$ und setzt diese Null, so entsteht ein lineares Gleichungssystem mit k Gleichungen, den **Normalgleichungen**, in den gesuchten Schätzwerten $\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_k$:

$$\begin{aligned}
c_{11}\hat{\beta}_1 + c_{12}\hat{\beta}_2 + \ldots + c_{1k}\hat{\beta}_k &= \sum_{t=1}^{N} x_t \cdot m_1(t) \\
c_{21}\hat{\beta}_1 + c_{22}\hat{\beta}_2 + \ldots + c_{2k}\hat{\beta}_k &= \sum_{t=1}^{N} x_t \cdot m_2(t) \\
&\vdots \\
c_{k1}\hat{\beta}_1 + c_{k2}\hat{\beta}_2 + \ldots + c_{kk}\hat{\beta}_k &= \sum_{t=1}^{N} x_t \cdot m_k(t)
\end{aligned} \quad [1.4.2.3]$$

mit

$$c_{ij} = \sum_{t=1}^{N} m_i(t) m_j(t).$$

Sofern die $m_i(t)$ linear unabhängig sind, hat das Gleichungssystem eine eindeutige Lösung $\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_k$, und diese Lösung liefert in der Tat ein Minimum von Q.

■

BEISPIEL 1.4.2.1

Wir illustrieren die Anpassung eines Polynoms an die Reihe STROM. Da die Schwankungen der Reihe im Zeitablauf steigen, gehen wir zu den Logarithmen über und passen ein Polynom 3. Grades an die Reihe $(x_t) = \log_{10}\text{STROM}$ an. Der Ansatz

$$Q = \sum_{t=1}^{N} (x_t - \beta_1 - \beta_2 t - \beta_3 t^2 - \beta_4 t^3)^2 \stackrel{!}{=} \min$$

führt auf die Normalgleichungen

$$\begin{aligned}
N\hat{\beta}_1 + (\Sigma t)\hat{\beta}_2 + (\Sigma t^2)\hat{\beta}_3 + (\Sigma t^3)\hat{\beta}_4 &= (\Sigma x_t) \\
(\Sigma t)\hat{\beta}_1 + (\Sigma t^2)\hat{\beta}_2 + (\Sigma t^3)\hat{\beta}_3 + (\Sigma t^4)\hat{\beta}_4 &= (\Sigma x_t \cdot t) \\
(\Sigma t^2)\hat{\beta}_1 + (\Sigma t^3)\hat{\beta}_2 + (\Sigma t^4)\hat{\beta}_3 + (\Sigma t^5)\hat{\beta}_4 &= (\Sigma x_t \cdot t^2) \\
(\Sigma t^3)\hat{\beta}_1 + (\Sigma t^4)\hat{\beta}_2 + (\Sigma t^5)\hat{\beta}_3 + (\Sigma t^6)\hat{\beta}_4 &= (\Sigma x_t \cdot t^3)
\end{aligned}$$

Dies ergibt die Schätzwerte

$$\begin{aligned}
\hat{\beta}_1 &= 3.7887 \\
\hat{\beta}_2 &= 0.0025717 \\
\hat{\beta}_3 &= 0.000003929 \\
\hat{\beta}_4 &= -0.00000001624.
\end{aligned}$$

Für die Originalreihe STROM entspricht das eben angepaßte Polynom einem exponentiellen Modell der Form $10^{m(t)}$, vgl. Abb. 1.3.3.

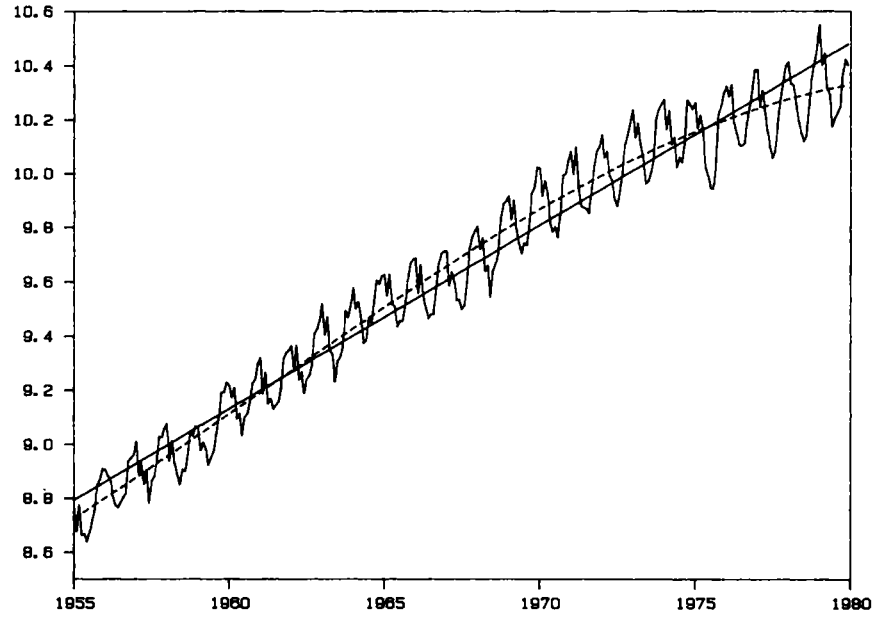


Abb. 1.4.2.2 *log_{10} STROM und lineare sowie kubische Trendkurve*

■

Ein Maß für die Güte der Anpassung eines linearen Trendmodells ist offenbar der quadrierte Abstand der Reihe vom geschätzten Trend $\hat{m}(t)$:

$$Q = \sum_{t=1}^{N} (x_t - \hat{m}(t))^2 . \qquad [1.4.2.4]$$

Um den Grad des Trendpolynoms zu bestimmen, bietet sich die folgende Vorgehensweise an. Es werden sukzessive Polynome 0'ten, 1'ten, 2'ten, ... Grades angepaßt und das Verhalten der Werte von Q beobachtet. Idealerweise sollte Q mit wachsendem Polynomgrad zunächst stärker und dann wesentlich langsamer abnehmen. Dann ist der geeignete Grad des Polynoms derjenige, von dem ab keine wesentliche Verringerung von Q mehr erreicht wird. In der Praxis ist diese Methode schwer anwendbar, da Q meist recht irregulär sinkt. Ein exaktes Auswahlverfahren gibt z. B. Anderson [1971] S. 34ff. an, einen heuristischen Ansatz skizzieren wir im Abschnitt 1.5.3.

1.4.3. Residuen-Analyse

Unter bestimmten Annahmen über die Störungen U_t haben die geschätzten Koeffizienten $\hat{\beta}_j$ Optimalitätseigenschaften. Sofern die U_t unabhängig normalverteilt mit Erwartungswert 0 und Varianz σ^2 sind, handelt es sich bei Schätzern, die mit dem Kleinstquadrateprinzip gewonnen wurden, um beste unverzerrte Schätzer. Werden diese Annahmen allerdings verletzt, so besitzen die Kleinstquadratschätzer keine oder nur noch unrealistisch eingeschränkte Optimalitätseigenschaften (wie „beste lineare Unverzerrtheit"). Es ist daher wichtig, die Gültigkeit der Annahmen jeweils kritisch zu überprüfen.

Grundlegend dabei sind die sogenannten **Residuen**, d.h. die Abweichungen von Daten x_t und geschätztem Trend $\hat{m}_t$:

$$\hat{u}_t = x_t - \hat{m}_t .$$

Unter Normalverteilungsannahmen sollte das Histogramm der Residuen näherungsweise einer Normalverteilung folgen – insbesondere sollten keine Ausreißer auftreten. Kleinstquadrateschätzer werden von untypischen Werten sehr stark beeinflußt. In eine quadratische Zielfunktion geht z. B. eine Abweichung der Größe 10 ebenso stark ein wie 100 Abweichungen der Größe 1. Gegebenenfalls sollte man die im folgenden Abschnitt angegebene Alternative einsetzen.

Der **Erwartungswert** aller U_t sollte gleich 0 sein – dies drückt inhaltlich aus, daß keine systematischen Effekte im Modell für $m(t)$ vernachlässigt wurden. Überprüft werden kann diese Annahme dadurch, daß die Residuen $\hat{u}_t$ gegen andere relevante Variablen, z. B. die Zeit t aufgetragen werden, um systematische Zusammenhänge zu entdecken. In Reihen mit Saisonschwankungen ist die Annahme $E[U_t] = 0$ bei einer Trendbestimmung im allgemeinen nicht erfüllt, da die Residuen $\hat{u}_t = x_t - \hat{m}_t$ systematisch mit der Saison variieren. Es werden sich daher im Regelfall andere Trendschätzungen ergeben, wenn man von vornherein ein gemeinsames Modell für Trend und Saison unterstellt. Man achte bei einer Zeichnung der $\hat{u}_t$ gegen t insbesondere auf sogenannte **Strukturbrüche**, d.h. tiefgreifende Veränderungen des Gesamtbildes. Methoden zur genaueren Überprüfung dieser Erscheinung geben z. B. Brown/Durbin/Evans [1975].

Die Annahme **konstanter Varianz** kann sehr gut in der Zeichnung der Residuen $\hat{u}_t$ gegen die Schätzungen $\hat{m}_t$ überprüft werden. Wenn die Varianz mit steigendem Trendwert zunimmt, so entsteht eine typische keilförmige Struktur:

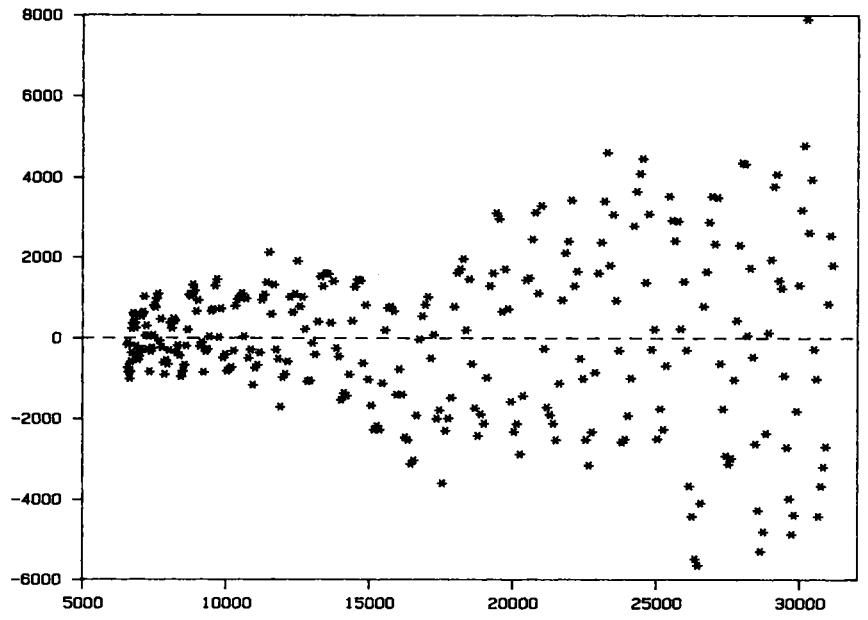


Abb. 1.4.3.1 *Residuen der Originalreihe STROM gegen die angepaßten Werte eines kubischen Polynoms*

Es sollte dann z. B. zu logarithmierten Werten übergegangen werden, wie es im Beispiel 1.4.2.1 geschehen ist. Weitere Transformationen behandeln wir in der Bemerkung 2.2.5. Auch das Auftreten nichtlinearer Beziehungen (etwa aufgrund der Vernachlässigung eines quadratischen Terms) kann in einer derartigen Zeichnung klar erkannt werden.

Die **Unabhängigkeitsannahme** impliziert insbesondere, daß die Störungen u_t nicht autokorreliert sind. Graphisch erkennt man **Autokorrelation** zum Lag 1 aus einer Zeichnung der $\hat{u}_t$ gegen die Zeit. Die Residuen sollten zufällig um 0 schwanken. Allzu langsame Schwankungen deuten auf positive, allzu schnelle und regelmäßige Schwankungen auf negative Autokorrelation hin. Im Beispiel der Jahreshöchstlasten ist etwa positive Autokorrelation zu beobachten. Eine andere Möglichkeit, Autokorrelationen zu erkennen, bieten Scatterdiagramme, in denen $\hat{u}_{t+\tau}$ gegen $\hat{u}_t$ eingezeichnet wird.

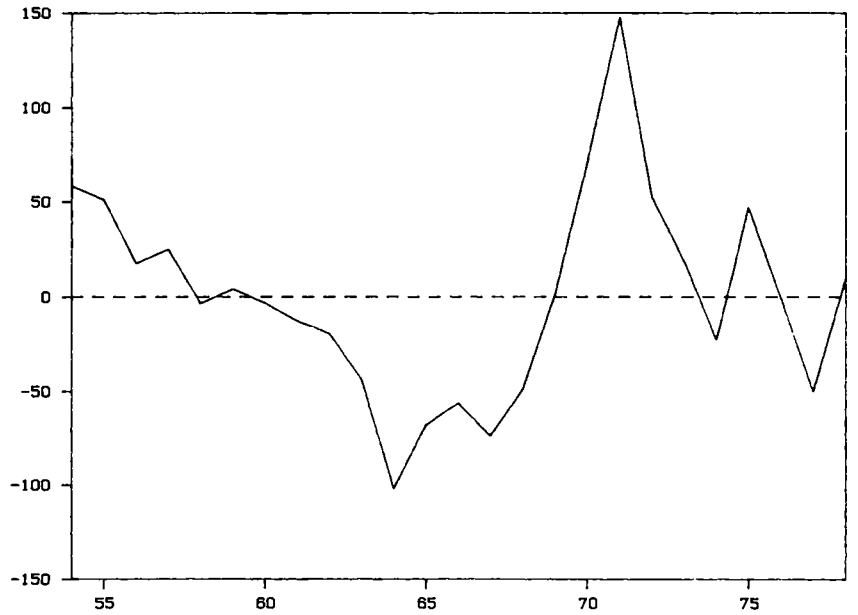


Abb. 1.4.3.2 *Residuen der Jahreshöchstlasten gegen die Zeit*

Ein Test auf Autokorrelation zum Lag 1 ist der **Durbin-Watson-Test**. Die Teststatistik ist

$$d = \sum_{t=2}^{N} (\hat{u}_t - \hat{u}_{t-1})^2 / \sum_{t=1}^{N} \hat{u}_t^2 .$$

Die asymptotische Verteilung von d unter der Nullhypothese unkorrelierter Störungen ist bei Kendall [1973] angegeben. Werte von d, die nahe bei 2 liegen, deuten auf Unkorreliertheit hin, solche nahe bei 0 auf positive und Werte nahe 4 auf negative Korrelation.

Im Kapitel über Spektralanalyse werden wir einen Test kennenlernen, der Autokorrelation beliebiger Ordnung überprüft.

Insbesondere die Annahme der Unabhängigkeit ist in der Zeitreihenanalyse praktisch niemals erfüllt. Dies hat zu einem gewissen Mißtrauen gegenüber Komponentenmodellen geführt und wesentlich die Hinwendung zur Theorie schwach stationärer Prozesse motiviert. Auf weitere Eigenschaften linearer Modelle sowie auf Tests der Parameter etc., gehen wir aus diesem Grunde nicht ein und verweisen auf die Ausführungen in Heinhold/Gaede [1979] und vor allem auf das Lehrbuch von Seber [1977].

1.4.4 Robuste Schätzung linearer Trendmodelle

Die Methode der kleinsten Quadrate ist bei normalverteilten Störungen im linearen Regressionsmodell in geeignetem Sinne optimal. Sie besitzt auch den Vorzug, daß eine Fülle von Methoden für weitergehende Analysen darauf aufgebaut sind. Jedoch sind wie erwähnt die Kleinste-Quadrate-Schätzungen sehr anfällig gegen Ausreißer, d. h. gegen einzelne, extreme Werte. Das häufige Auftreten von Ausreißern hat zu einer verstärkten Hinwendung zu robusten Verfahren geführt; dies sind Verfahren, welche „unempfindlich" gegen einzelne extreme Werte sind. Ein für die gesamte Zeitreihenanalyse relevanter Ansatz zur Gewinnung robuster Verfahren bildet die **M-Schätzung** von Huber [1964] mit ihren Verallgemeinerungen. Sie wird zunächst für das einfache Lageproblem eingeführt.

Bei einer univariaten Stichprobe $x_1, \ldots, x_N$ erfüllt das arithmetische Mittel $\bar{x}$ die Minimierungsaufgabe

$$\sum_{t=1}^{N} (x_t - a)^2 \stackrel{!}{=} \min_a,$$

und der Median $\tilde{x}$ ist Lösung von

$$\sum_{t=1}^{N} |x_t - a| \stackrel{!}{=} \min_a.$$

Die beiden Minimierungsprobleme lassen sich lösen, indem man die Ableitungen der linken Seiten Null setzt:

$$\sum_{t=1}^{N} \psi(x_t - a) = 0 \qquad [1.4.4.1]$$

mit $\psi(u) = u$ bzw. $\psi(u) = \operatorname{sign}(u)$.

Die zu $\varrho(u) = u^2$ gehörende Funktion $\psi(u) = u$ ist unbeschränkt. Die Konsequenz ist, daß weit von u weg liegende Beobachtungen x mit einem großen Gewicht in [1.4.4.1] eingehen. Dagegen haben bei der zu $\varrho(u) = |u|$ gehörenden Funktion $\psi(u) = \operatorname{sign}(u)$ alle Beobachtungen das gleiche Gewicht. Große Abstände haben keine größere Auswirkung als kleine. Der Median ist daher ungleich unempfindlicher gegen Ausreißer als das arithmetische Mittel.

Dementsprechend erhält man robuste Alternativen zu $\bar{x}$ über den Minimierungsansatz

$$\sum_{t=1}^{N} \varrho((x_t - a)/\sigma) \stackrel{!}{=} \min_a, \qquad [1.4.4.2]$$

wenn die Abstandsfunktion $\varrho(u)$ langsamer mit $|u|$ wächst als die quadratische Funktion u^2. Hier muß ein Skalenparameter σ mit aufgenommen werden, da die meisten der üblichen ϱ-Funktionen nicht adäquat auf Skalenänderungen reagieren. $\varrho(u) = u^2$ und $\varrho(u) = |u|$ sind Ausnahmen. Zur Bestimmung von σ kommen wir etwas weiter unten.

Über eine geeignete Wahl der Funktion $\varrho(u)$ versucht man nun, zwei Ziele möglichst gleichzeitig zu erreichen. Einmal wird eine große Unempfindlichkeit gegen Ausreißer angestrebt. Dazu müssen extreme Werte u ‚heruntergewichtet' werden. Zum anderen möchte man Schätzer, die im Fall der Normalverteilung keine wesentlich größere Varianz als $\bar{X}$ aufweisen, die also möglichst effizient sind. Das wird erreicht, indem in einem beschränkten Intervall um Null die Funktion $\psi(u)$ annähernd linear verläuft.

Auf der Basis allgemeiner Optimalitätsüberlegungen gelangte Huber zu folgender ψ-Funktion, welche die beiden oben genannten kombiniert:

$$\psi(u) = \begin{cases} u & |u| < 1.5 \\ 1.5 \cdot \operatorname{sign}(u) & |u| \geq 1.5 \end{cases}$$

Hier werden nur „zu große" Abstände heruntergewichtet.

Um noch sicherer gegen extreme Werte zu sein, haben andere Autoren sogenannte ‚redescending' ψ-Funktionen vorgeschlagen. Dazu gehört z. B. auch Tukey's biweight:

$$\psi(u) = \begin{cases} \frac{u}{6}\left(1 - \frac{u^2}{36}\right)^2 & |u| \leq 1 \\ 0 & |u| > 1 \end{cases}$$

sowie die ψ-Funktion von Hampel:

$$\psi(u) = \begin{cases} u & |u| < 1.5 \\ 1.5 \cdot \operatorname{sign}(u) & 1.5 < |u| \leq 3.6 \\ 0.341 \cdot (8 - |u|) \cdot \operatorname{sign}(u) & 3.6 < |u| \leq 8.0 \\ 0 & |u| > 8.0 \end{cases}$$

Bei allen Vorschlägen gibt es frei wählbare Parameter, für die hier schon die Werte geeigneter Empfehlungen eingesetzt wurden, vgl. Goodall [1983].

In vielen Fällen kann σ robust durch den MAD, den Median der absoluten Abweichungen vom Median,

$$\hat{\sigma} = 1.483 \cdot \operatorname{median}\{|x_t - \operatorname{median}\{x_s\}|\}$$

geschätzt werden. Der Faktor 1.483 macht den Schätzer im Fall der Normalverteilung erwartungstreu.

Anstatt der Streuung σ mit dem MAD vorab zu schätzen, kann man ihre Bestimmung auch in die M-Schätzung einbeziehen. Dazu ist [1.4.4.2] nicht nur bzgl. a sondern zugleich bzgl. σ zu minimieren. Dies führt zu [1.4.4.1] und zu

$$\sum_{t=1}^{N} \psi\left(\frac{x_t - a}{\hat{\sigma}}\right) \cdot \frac{x_t - a}{\hat{\sigma}} = 0.$$

Um dem Robustheitsgesichtspunkt Rechnung zu tragen, wird auch der zweite Faktor der einzelnen Summanden robustifiziert und als Argument der ψ Funktion geschrieben. Dann ist die Forderung, die Summe soll Null werden, nicht mehr stimmig. Stattdessen orientiert man sich an der Eigenschaft der KQ-Schätzer $\Sigma((x_t - \hat{\mu})/\hat{\sigma})^2 - 1 = 0$ und formuliert die simultanen M-Schätzer für die Lage und Streuung als Lösungen der Gleichungen

$$\sum_{t=1}^{N} \psi\left(\frac{x_t - a}{\hat{\sigma}}\right) = 0,$$

$$\sum_{t=1}^{N} \chi\left(\frac{x_t - a}{\hat{\sigma}}\right) = a,$$

wobei $\chi(u) = \psi(u)^2$ (oder ähnlich) gewählt wird und $a = (N-1) \cdot E[\chi(Z)]$ mit einer standardnormalverteilten Zufallsvariablen Z. Zu den statistischen Eigenschaften kommen wir im Kapitel 5.

Die Verallgemeinerung der M-Schätzung des Lageparameters auf die Schätzung des Parametervektors im Regressionsmodell ist leicht. Der M-Schätzer für den Parametervektor β im Regressionsmodell [1.4.2.1], der auf der Abstandsfunktion $\varrho(u)$ basiert, ist der Vektor $(\hat{\beta}_1, \ldots, \hat{\beta}_k)$ für den gilt:

$$\sum_{t=1}^{N} \varrho\left(\frac{x_t - \hat{\beta}_1 m_1(t) - \ldots - \hat{\beta}_k m_k(t)}{\hat{\sigma}}\right) =$$

$$\min_{\beta_1, \ldots, \beta_k} \sum_{t=1}^{N} \varrho\left(\frac{x_t - \beta_1 m_1(t) - \ldots - \beta_k m_k(t)}{\hat{\sigma}}\right) \qquad [1.4.4.3]$$

$\hat{\sigma}$ ist eine Schätzung des Parameters, der die Streuung der Störungen U_t charakterisiert. (Wie eben kann σ gemeinsam mit β geschätzt werden. Der Einfachheit halber betrachten wir nur den Fall einer Vorabschätzung von σ.) Bei einer differenzierbaren Abstandsfunktion $\varrho(u)$ mit der Ableitung $\varrho'(u) = \psi(u)$ führt die Aufgabe [1.4.4.3] zum Problem die simultane Lösung der folgenden Gleichungen zu bestimmen:

$$\sum_{t=1}^{N} \psi\left(\frac{x_t - \hat{\beta}_1 m_1(t) - \ldots - \hat{\beta}_k m_k(t)}{\hat{\sigma}}\right) \cdot m_i(t) = 0 \qquad i = 1, \ldots, k$$

BEISPIEL 1.4.4.1

Die Werte der Reihe CITI sind die wöchendlichen „Dutch auction rates" der Citibank Vorzugsaktien in Prozent der „commercial paper rate". Der Zeitraum reicht von Januar 1986 bis Oktober 1990. Der mittlere Peak von $t = 95$ bis $t = 102$ resultiert aus einer Unsicherheit bzgl. der Änderung der Gesetze. Die extremen Ausschläge am Ende der Reihe reflektiert die plötzliche Beunruhigung der Investoren bzgl. des finanziellen Wohlergehens der Unternehmung. Der Plot der Reihe zeigt eine leicht ansteigende Tendenz, so daß für weitere Analysen eine Trendbereinigung sinnvoll zu sein scheint.

Die Abbildung zeigt den robust mit Hubers M-Schätzer bzw. den nach der KQ-Methode ermittelten linearen Trend. Es ist deutlich zu sehen, wie der nicht

robust ermittelte Trend durch die großen Werte am Ende angehoben wird. Die geschätzten Koeffizienten sind dabei:

$$\hat{\beta}_{KQ} = \begin{pmatrix} 70.804 \\ 0.070 \end{pmatrix} \quad \text{und} \quad \hat{\beta}_{M\,Huber} = \begin{pmatrix} 71.54 \\ 0.053 \end{pmatrix}.$$

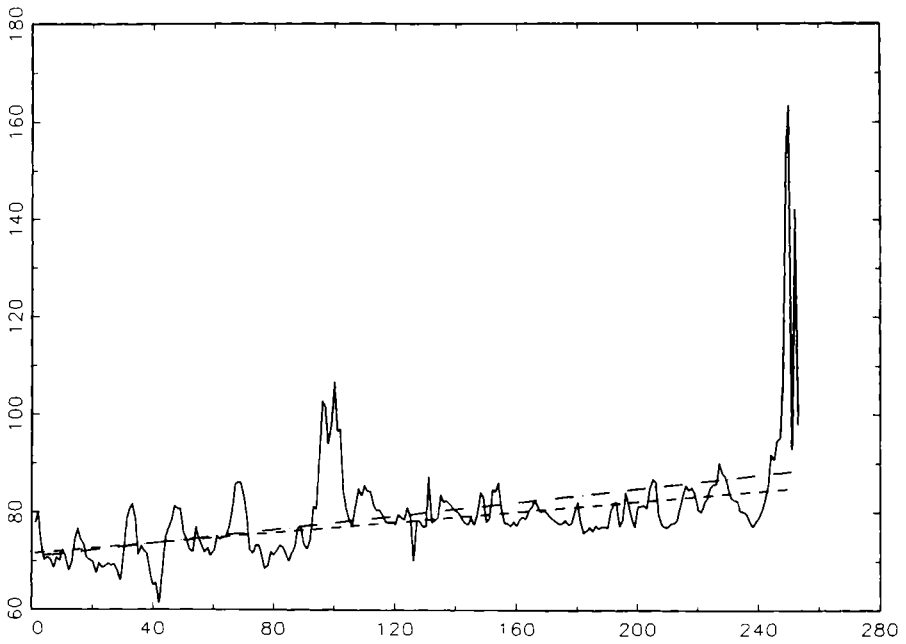


Abb. 1.4.4.1 *Reihe CITI mit linearem Trend (– · – · – = KQ-Methode, – – – = M-Schätzer)*

■

1.4.5 Nichtlineare Trendmodelle

Wir haben bisher ausschließlich **lineare Trendmodelle** betrachtet, d.h. solche Modelle, bei denen die Parameter β_j linear mit den bekannten Zeitfunktionen $m_j(t)$ verknüpft sind. Gelegentlich lassen sich auch **nichtlineare** Modelle durch Transformation der beobachteten Reihe auf ein lineares Regressionsmodell zurückführen. Typische Beispiele sind **Exponentialmodelle**

$$x_t = e^{\beta_1 m_1(t) + \beta_2 m_2(t) + \ldots + \beta_k m_k(t) + u_t}$$

oder **Potenzmodelle**

$$x_t = m_1(t)^{\alpha_1} \cdot m_2(t)^{\alpha_2} \cdot \ldots \cdot m_k(t)^{\alpha_k} \cdot u_t,$$

bei denen der Übergang zu den Logarithmen $\log x_t$ ein lineares Modell ergibt.

Aus inhaltlichen Erwägungen werden nun jedoch mitunter Modelle konstruiert, die sich nicht auf ein lineares Modell reduzieren lassen. Ein Beispiel ist bereits ein

Exponentialmodell, in dem der Störterm u_t nicht multiplikativ, sondern additiv auftritt, z. B.

$$x_t = e^{\beta_1 + \beta_2 t} + u_t .$$

Ein in der Praxis der Zeitreihenanalyse relevanteres Modell behandelt das folgende Beispiel.

BEISPIEL 1.4.5.1

Es ist oft unrealistisch, für Wachstumsprozesse eine unbeschränkte Expansionsfähigkeit zu unterstellen. Vielmehr ist häufig eine obere Grenze für das Wachstum einer Erscheinung anzunehmen. So werden z. B. biologische oder ökologische Restriktionen verhindern, daß ein Organismus oder eine Population über alle Grenzen wächst. Ein typisches Modell für derartige Wachstumsvorgänge ist die sogenannte **logistische Funktion**

$$m(t) = \frac{B}{C + e^{-At}}, \qquad [1.4.5.1]$$

welche für $t \to \infty$ einer Sättigungsgrenze $G = \frac{B}{C}$ zustrebt. Die von $m(t)$ beschriebene S-förmige Wachstumskurve ist in vielen empirischen Datensätzen zu beobachten (vgl. Abb. 1.4.5.1).

Das zugrundeliegende Modell wird deutlicher, wenn wir durch Differenzieren von $m(t)$ zu den Wachstumsraten $r(t) = m'(t)/m(t)$ übergehen:

$$r(t) = A \cdot \left(1 - \frac{m(t)}{G}\right). \qquad [1.4.5.2]$$

Diese sogenannte Verhulst'sche Differentialgleichung beschreibt die Geschwindigkeit des Wachstums in Abhängigkeit vom erreichten Wert $m(t)$. Wenn $m(t)$ weit unterhalb der Sättigungsgrenze G liegt, so ist die Wachstumsrate praktisch gleich A und das System wächst exponentiell. Je stärker sich $m(t)$ der Grenze G nähert, desto stärker wird die Wachstumsrate gesenkt und damit das Wachstum abgebremst.

■

Es ist nicht möglich, die Parameter A, B, C im obigen Beispiel durch ein exaktes lineares Modell zu schätzen. Das logistische Modell ist wesentlich nichtlinear. Man kann jedoch versuchen, das Modell approximativ zu linearisieren. Dies ist der Grundgedanke des Gauß-Newton-Verfahrens.

Betrachten wir wieder das Ausgangsproblem, einer Folge von Werten $x_1, x_2, \ldots, x_N$ eine nichtlineare Funktion $m(t; \theta)$ anzupassen, die von einem Parameter θ abhängt. Das Kleinstquadrate-Kriterium

$$\sum_{t=1}^{N} (x_t - m(t; \theta))^2 \stackrel{!}{=} \min$$

führt hier in der Regel zu Normalgleichungen, die nicht mehr explizit lösbar sind.

Ein einfacher Ansatz zur Lösung der Normalgleichungen ist der folgende, der auf dem Grundgedanken der approximativen Linearisierung beruht. Beginnend mit einer Startlösung θ_0 suchen wir eine lineare Approximation der Funktion $m(t; \theta)$

in der Nähe von θ_0. Dies führt auf

$$m(t;\theta) \approx m(t;\theta_0) + (\theta - \theta_0) \cdot m'(t;\theta_0),$$

wobei $m'(t;\theta_0)$ die Ableitung von $m(t;\theta)$ nach θ an der Stelle θ_0 bezeichnet. Wir gehen also von einem nichtlinearen Modell in θ

$$x_t = m(t;\theta) + u_t$$

zu einem linearen Modell in $\Delta = (\theta - \theta_0)$

$$x_t \approx m(t;\theta_0) + \Delta \cdot m'(t;\theta_0) + u_t$$

über, das jedoch nur approximativ gültig ist. Eine Korrektur Δ der Anfangslösung kann dann mit Hilfe des üblichen Kleinstquadrate-Ansatzes geschätzt werden; das Resultat $\hat{\Delta}$ liefert eine (wie man hofft: bessere) Lösung $\theta_1 = \hat{\Delta} + \theta_0$, die zu einem neuen Iterationsschritt benutzt werden kann. Diese Methode wird dann als **Gauß-Newton-Verfahren** bezeichnet. Wir illustrieren es zunächst an einem überschaubaren Zahlenbeispiel.

BEISPIEL 1.4.5.2

Den Werten

$$x_1 = 1.4 \quad x_2 = 1.9 \quad x_3 = 4.5 \quad x_4 = 7.8$$

liege das Modell

$$x_t = e^{\theta t} + u_t$$

zugrunde. Für θ soll eine Kleinstquadrateschätzung bestimmt werden.

Einen Startwert θ_0 erhalten wir z. B. aus der Forderung

$$x_1 = e^{\theta_0 1} \Rightarrow \theta_0 = \ln x_1 = 0.34.$$

Wir approximieren das Modell durch ein lineares Modell. Wegen

$$\frac{d}{d\theta} e^{\theta t} = t \cdot e^{\theta t}$$

ist dieses linearisierte Modell:

$$x_t \approx e^{\theta_0 t} + (\theta - \theta_0) \cdot t \cdot e^{\theta_0 t} + u_t.$$

Mit $\Delta = \theta - \theta_0$ lautet das Kleinstquadrate-Kriterium:

$$\sum_{t=1}^{4} (x_t - e^{0.34t} - \Delta t e^{0.34t})^2 \stackrel{!}{=} \min_{\Delta}$$

Differentiation nach Δ und Nullsetzen der Ableitung gibt die Normalgleichung in $\hat{\Delta}$:

$$\sum_{t=1}^{4} (t e^{0.34t})^2 \cdot \hat{\Delta} = \sum_{t=1}^{4} (x_t - e^{0.34t})(t e^{0.34t});$$

deren Lösung ist $\hat{\Delta} = 0.226$.

Als neue Näherung θ_1 für $\hat{\theta}$ haben wir somit

$$\theta_1 = \theta_0 + \hat{\Delta} = 0.566.$$

Die Verbesserung drückt sich in dem Vergleich der Summe der Abweichungsquadrate aus:

$$\sum_{t=1}^{4} (x_t - e^{0.34t})^2 = 18.28$$

$$\sum_{t=1}^{4} (x_t - e^{0.566t})^2 = 5.82.$$

■

Wenn **mehrere Parameter** in $m(t)$ zu schätzen sind, kann im Prinzip der gleiche Ansatz verwendet werden. Wir führen dies für den Fall von zwei Parametern aus; der allgemeine Fall wird analog behandelt.

Die Trendfunktion $m(t)$ hänge ab von α und β,

$$m(t) = m(t; \alpha, \beta).$$

Die Verallgemeinerung des Ansatzes für einen Parameter führt zu partiellen Ableitungen, die an der Stelle der ersten Näherung (α_0, β_0) zu bestimmen sind:

$$m_1(t) = \frac{\partial}{\partial \alpha} m(t; \alpha, \beta)\,|_{\alpha_0, \beta_0}$$

$$m_2(t) = \frac{\partial}{\partial \beta} m(t; \alpha, \beta)\,|_{\alpha_0, \beta_0}.$$

Das linearisierte Trendmodell ist dann mit $m_0(t) = m(t; \alpha_0, \beta_0)$, $\Delta\alpha = \alpha - \alpha_0$ und $\Delta\beta = \beta - \beta_0$:

$$(x_t - m_0(t)) \approx m_1(t) \cdot \Delta\alpha + m_2(t) \cdot \Delta\beta + u_t.$$

Daraus können Kleinstquadrateschätzer $\hat{\Delta}\alpha, \hat{\Delta}\beta$ der Differenzen bestimmt werden, die zu neuen Schätzungen $\alpha_1 = \alpha_0 + \hat{\Delta}\alpha$, $\beta_1 = \beta_0 + \hat{\Delta}\beta$ führen. Mit diesen kann das Verfahren dann iteriert werden.

BEISPIEL 1.4.5.3

Wir wollen das logistische Modell [1.4.5.1] an die Reihe USPOP (Bevölkerung der USA) von 1790 bis 1910 anpassen (Quelle: World-Almanac [1978]). Wir suchen zunächst eine Startlösung mit Hilfe der Methode von Hotelling/Davis (vgl. Davis [1941]). Dazu approximieren wir die Verhulst'sche Differentialgleichung 1.4.5.2 durch diskrete Wachstumsraten der beobachteten Werte:

$$r_t = \frac{x_{t+1} - x_t}{x_t} \approx A - \frac{A}{G} x_t.$$

Diese Approximation hat den Vorteil, daß zwischen x_t und r_t näherungsweise eine lineare Beziehung $r_t = \beta_1 - \beta_2 x_t$ besteht. Für β_1 und β_2 können somit leicht Schätzungen $\hat{\beta}_1, \hat{\beta}_2$ nach der Kleinstquadrate-Methode gewonnen werden, die wiederum zu Ausgangsschätzungen für A und G führen: $A_0 = \hat{\beta}_1$, $G_0 = \hat{\beta}_1/\hat{\beta}_2$. Zur Gewinnung eines Startwertes für C gehen wir aus von der Ausgangsgleichung [1.4.5.1] in der mit $G = B/C$ wegen $m(t) \approx x_t$ gilt:

$$x_t \approx \frac{G \cdot C}{C + e^{-At}} \quad t = 1, \ldots, N$$

Einsetzen von A_0 und G_0 sowie Auflösen nach C ergibt

$$C \approx \frac{x_t e^{-A_0 t}}{G_0 - x_t}.$$

Ein plausibler Startwert C_0 ist das arithmetische Mittel aller rechts stehenden Ausdrücke. Der schließlich noch notwendige Startwert für B ist dann $B_0 = G_0 \cdot C_0$. Für USPOP sind diese Anfangsschätzungen unter der Iteration 0 in der folgenden Tabelle aufgeführt, zusammen mit der entsprechenden Abstandsquadratsumme Q.

Verbesserte Lösungen bringt nun die lineare Approximation. Durch partielles Differenzieren von $m(t; A, B, C)$ finden wir:

$$m_1(t) = (tB_0 e^{-A_0 t})/(C_0 + e^{-A_0 t})^2$$
$$m_2(t) = 1/(C_0 + e^{-A_0 t})$$
$$m_3(t) = -B_0/(C_0 + e^{-A_0 t})^2.$$

Die Kleinstquadrate-Methode liefert dann Korrekturen, die zu neuen Schätzungen benutzt werden können. Der Verlauf der Iteration ist in folgender Tabelle dargestellt:

Iteration	A	B	C	Q
0	0.3641	2.1735	0.01328	198.616
1	0.3153	2.8156	0.01569	48.265
2	0.3135	2.9212	0.01485	2.590
3	0.3135	2.9216	0.01489	2.587
4	0.3135	2.9216	0.01489	2.587

Nach drei Iterationen ist der Prozeß praktisch konvergiert, die Abstandsquadratsumme Q nimmt nicht mehr ab. Die geschätzte Grenze für das Bevölkerungswachstum der USA ist

$$\hat{G} = \frac{B_4}{C_4} = 196.25.$$

Daß auch das logistische Modell bei Prognosen mit Vorsicht verwendet werden sollte, zeigt der Census von 1970 mit der Bevölkerung von 203.21 Mio.

Ein alternatives Schätzverfahren, das von $\ln x_t$ ausgeht, entnehme man Nelder [1961].

■

Der Leser lasse sich nicht davon täuschen, daß die Methode der approximativen Linearisierung in unseren Beispielen gut funktioniert hat. In vielen Fällen wird das Verfahren nicht zum Ziel führen, insbesondere, wenn die Startwerte schlecht gewählt wurden. Es ist dann nötig, den Ansatz in bestimmter Weise zu modifizieren, vgl. Bard [1974], Luenberger [1973].

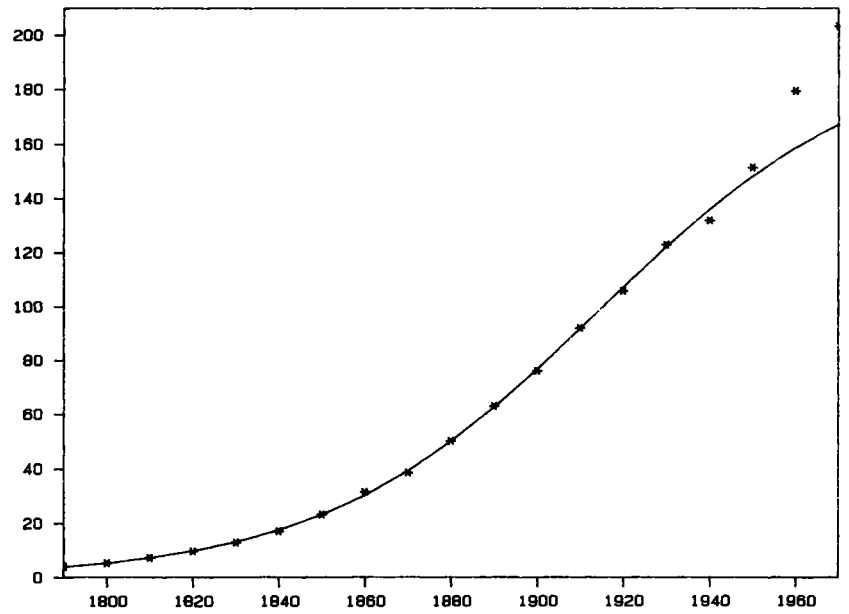


Abb. 1.4.5.1 *Logistisches Wachstum der US Population USPOP 1790–1970 (die Kurve wurde auf Grund der Werte von 1790–1910 geschätzt).*

1.4.6 Splines

Hat man für den Verlauf des Trends bzw. der glatten Komponente keine funktionale Vorstellung, bieten sich glättende Splines an, um die wesentliche Struktur der zeitlichen Entwicklung herauszuarbeiten. Dabei wird lediglich davon ausgegangen, daß die langfristige Entwicklung relativ glatt verläuft. Die übliche mathematische Operationalisierung der Glattheit basiert auf dem Konzept der Ableitung einer Funktion $g(t)$. Je häufiger $g(t)$ differenzierbar sie ist und je kleiner die Ableitungsordnung k ist, für die $g^{(k)}(t)$ in etwa Null ist, desto glatter erscheint $g(t)$ auch visuell. Beim Splines-Ansatz ist es üblich, $k = 2$ zu wählen. Es gilt dann, einen Trend zu bestimmen, der das Glattheitsmaß

$$\int \left[\frac{\partial^2}{\partial t^2} g(t) \right]^2 \mathrm{d}t$$

minimiert. Natürlich soll $g(t)$ auch den Verlauf der Zeitreihe wiedergeben, also gleichzeitig

$$\sum_{t=1}^{N} [x_t - g(t)]^2$$

möglichst klein machen. Diese beiden Forderungen widersprechen sich jedoch. Die Glattheitsbedingung wird genau dann Null, wenn $g(t)$ eine lineare Trendfunktion ist. Die Übereinstimmung von $g(t)$ und x_t ist andererseits am besten, wenn

$g(t)$ die Reihe interpoliert. Zur Lösung des Widerspruches wird die Übereinstimmung formal als Nebenbedingung

$$\sum_{t=1}^{N} [x_t - g(t)]^2 \leq S$$

formuliert. Dabei ist $S \geq 0$ eine vorgegebene Konstante.

Da $g(t)$ nur an diskreten, gleichabständigen Zeitpunkten interessiert, ist es sinnvoll, die Forderung bzgl. der zweiten Ableitung zu modifizieren. Für die erste Ableitung erhalten wir

$$\frac{\partial}{\partial t} g(t) = \lim_{h \to 0} \frac{g(t+h) - g(t)}{h} \to \frac{g(t+1) - g(t)}{1} = g(t+1) - g(t)$$

Die wiederholte Anwendung dieser Ersetzung führt auf das modifizierte Glattheitsmaß für den Trend:

$$\sum_{t=3}^{N} [g_t - 2g_{t-1} + g_{t-2}]^2.$$

Mit einem Lagrange-Parameter β^2 kann die Aufgabe der **Trendbestimmung mittels glättender Splines** auch geschlossen formuliert werden:

$$\beta^2 \sum_{t=3}^{N} [g_t - 2g_{t-1} + g_{t-2}]^2 + \sum_{t=1}^{N} [x_t - g_t]^2 \stackrel{!}{=} \min_{(g_t)}. \qquad [1.4.6.1]$$

β^2 besitzt eine Interpretation als Glattheitsparameter. Je größer β^2 ist, desto mehr Gewicht erhält die Summe der quadrierten zweiten Differenzen von g_t. Entsprechend wird sie gegenüber der Anpassung an die beobachteten Werte x_t ‚bevorzugt'.

Die Minimierungsaufgabe [1.4.6.1] läßt sich leicht lösen, wenn man beachtet, daß sie sich formal auch über einen Regressionsansatz ergibt. Mit

$$z_t = \begin{cases} x_t & \text{für } t = 1, \ldots, N \\ 0 & \text{für } t = N+1, \ldots, 2N-2 \end{cases}$$

lautet das formale Regressionsmodell:

$$z_t = \begin{cases} g_t + U_t & \text{für } t = 1, \ldots, N \\ \beta(g_t - 2g_{t-1} - g_{t-2}) + U_t & \text{für } t = N+1, \ldots, 2N-2 \end{cases}$$

Dieses enthält in Matrizenform die Gestalt:

$$\begin{pmatrix} x_1 \\ \vdots \\ x_N \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & & & & & & & \\ 0 & 1 & & & & & & & \\ & & \ddots & & & & & & \\ & & & & & & 1 & 0 & \\ 0 & & & & & & 0 & 1 & \\ 1\beta & -2\beta & 1\beta & 0 & \ldots & & & 0 & \\ 0 & 1\beta & -2\beta & 1\beta & & & & & \\ 0 & & 0 & & \ldots & 1\beta & -2\beta & 1\beta & 0 \\ 0 & & 0 & & & 0 & 1\beta & -2\beta & 1\beta \end{pmatrix} \begin{pmatrix} g_1 \\ \vdots \\ g_N \end{pmatrix} + \begin{pmatrix} U_1 \\ \vdots \\ U_N \\ U_{N+1} \\ \vdots \\ U_{2N-2} \end{pmatrix}$$

Die Methode der kleinsten Quadrate minimiert nun gerade die Zielfunktion [1.4.6.1]. Sie führt auf die Normalgleichungen (für $N \geq 5$, $3 \leq t \leq N-2$):

$$
\begin{array}{llllll}
\beta^2\hat{g}_1 & -2\beta^2\hat{g}_2 & +\beta^2\hat{g}_3 & & +\hat{g}_1 & = x_t \\
-2\beta^2\hat{g}_1 & +5\beta^2\hat{g}_2 & -4\beta^2\hat{g}_3 & +\beta^2\hat{g}_4 & +\hat{g}_2 & = x_2 \\
\beta^2\hat{g}_{t-2} & -2\beta^2\hat{g}_{t-1} & +5\beta^2\hat{g}_t & -4\beta^2\hat{g}_{t+1}+\beta^2\hat{g}_{t+2} & +\hat{g}_t & = x_t \quad \text{für} \quad 3 \leq t \leq N-2 \\
\beta^2\hat{g}_{N-3} & -4\beta^2\hat{g}_{N-2} & +5\beta^2\hat{g}_{N-1} & -2\beta^2\hat{g}_N & +\hat{g}_{N-1} & = x_{N-1} \\
\beta^2\hat{g}_{N-2} & -2\beta^2\hat{g}_{N-1} & +\beta^2\hat{g}_N & & +\hat{g}_N & = x_N
\end{array}
$$

Da das Gleichungssystem Band-liniiert ist, d.h. mit g_t kommen nur die vier benachbarten Koeffizienten in der gleichen Zeile vor, läßt es sich numerisch recht einfach berechnen. Sowohl der Speicheraufwand als auch die benötigte Anzahl von Rechenoperationen sind sehr gering, vgl. dazu Golub/vanLoan [1983].

BEISPIEL 1.4.6.1

Das östliche Afrika sieht sich vielen Problemen gegenüber, die mit der Klimaentwicklung zusammenhängen. Herausragend ist dabei das Wasserproblem. Um das Auftreten und die Dauer von Trockenheitsperioden besser verstehen zu lernen, betrachtet man heutzutage intensiver historische Zeitreihen über die Wasserstände des Nil. Für die Jahre 715 bis 1284 liegen die jährlichen Minima (d.h. die jeweils am 3. Juli am Nilometer in der Nähe von Kairo abgelesenen Wasserstände) lückenlos vor. Siehe dazu Popper [1951] und Reusing/Skala [1990].

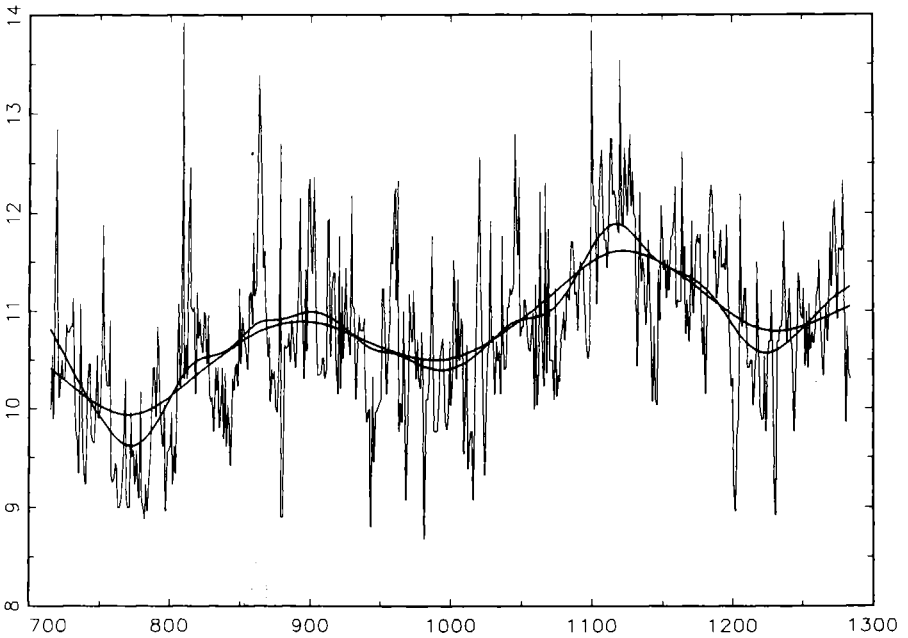


Abb. 1.4.6.1 *Jährliche NIL-Minima [m] mit Splines-Trendfunktionen ($\beta = 200$ bzw. $\beta = 750$)* ■

Hier ist nun ein interessanter Aspekt die längerfristige Entwicklung des mittleren Verlaufes der Reihe NIL. Dieser wird durch die mit dem Spline-Ansatz ermittelte Trendfunktion deutlich. ■

Der Spline-Ansatz basiert auf der Unabhängigkeit der Reste U_t. Schimek [1992] untersucht die Auswirkungen von Abhängigkeiten der Störungen und kommt zu dem Schluß, daß sie bei der Trendbestimmung nicht schwerwiegend sind. Sollen aber zusätzlich Fehlerintervalle bestimmt werden, so ändert sich das Bild.

Über eine Vergrößerung des Glattheitsparameters kann die Auswirkung von Ausreißern nur partiell ausgeglichen werden. Mit einem größeren Wert von β^2 geht eine insgesamt abgeflachtere glatte Komponente einher. Die geringere Ausreißerempfindlichkeit wird durch eine schlechter mit dem Reihenverlauf harmonierende glatte Komponente erkauft, vgl. Abb. 1.4.6.1. Dies zeigt den Bedarf an einer robusten Variante der Splines-Glättung. Eine Möglichkeit, die Splines-Glättung zu robustifizieren, ist offensichtlich: In dem obigen Zielkriterium werden wie bei der M-Schätzung die Abstandsquadrate $(x_t - g_t)^2$ durch andere Abstände $\varrho(x_t - g_t)$ ersetzt. Siehe dazu Schlittgen [1990].

1.4.7 Trendzerlegung von Zeitreihen

Insbesondere bei Kursreihen tritt das Problem auf, **lokale Trends** in einer gegebenen Reihe zu bestimmen, d. h. die Reihe in steigende und fallende Segmente zu zerlegen. Ein solches Segment kann bei einer Feinanalyse u. U. wieder in Trendsegmente kleinerer Ordnung zerfallen. Plastisch ist diese Sichtweise in folgendem Zitat von Edwards/Maggee [1966, S. 15] beschrieben:

"The major trends in stock prices are like the tides ..., the waves are the intermediate trends ..., the surface of the water is constantly agitated by wavelets [and] ripples, ... these are analogous to the market's minor trends."

Das Wort „Trend" wird in der sogenannten technischen Analyse von Kursreihen häufig nicht nur im Sinne der reinen Deskription einer Zeitreihe verstanden, sondern mit der Vorstellung verbunden, daß die Zeitreihe bestimmten Einflüssen von außen unterliegt. Die Hoffnung von Spekulanten ist dann, solche Trends und vor allem die Trendwechselpunkte rechtzeitig zu erkennen und gewinnbringend auszunutzen. Die bisher betrachteten, in der Mehrzahl graphisch ausgerichteten Verfahren (Charts) sind selbst für die Beschreibung der ‚Trendwechsel' nur bedingt geeignet. Ein ganz neuartiges und vielversprechendes Verfahren zur Beschreibung der mit lokalen Trends zusammenhängenden Charakteristiken einer – nicht als äquidistant vorausgesetzten – Zeitreihe wurde von Wegscheider [1994] vorgeschlagen. Wie weiter unten deutlich werden wird, sind diese Charakteristiken keineswegs nur für Kursreihen von Interesse. Wichtige, inhaltlich interpretierbare Aspekte lassen sich damit z. B. auch in EEG-Reihen und musikalischen Anwendungen ermitteln. Bei Wegscheider's Ansatz wird jedem Zeitpunkt t der Reihe eine Maßzahl $p(t)$, das sogenannte **Trend-Umkehr-Potential** zugeordnet, wobei absolut großen Potentialen Umkehrpunkte von großen Trends entsprechen. **Trendumkehrpunkte** sind dabei solche Zeitpunkte, die am Ende der einen Trendrichtung und gleichzeitig am Beginn der dieser entgegengesetzten Trendrichtung stehen.

BEISPIEL 1.4.7.1

Die in Abb. 2.1.3 auf S. 94 dargestellte Reihe IBM der Aktienkurse an 369 aufeinanderfolgenden Börsentagen zeigt sehr deutlich die beschriebene Struktur der Trends unterschiedlicher Größenordnung. Neben einigen großen Trends sind auch die kleineren offensichtlich. Die Zeitpunkte sind nicht äquidistant, was bei der Analyse von Kursreihen i. d. R. vernachlässigt wird. ■

Die Berechnung der Wegscheider-Potentiale geschieht iterativ, wobei schrittweise die „uninteressantesten" Punkte der Zeitreihe gelöscht werden. Diesen Punkten wird ihr (vertikaler) Abstand als Potential zugeordnet. Die Iteration endet mit den beiden globalen Extrema (Maximum und Minimum) der Zeitreihe. Um diese Iteration exakt beschreiben zu können, ist etwas Notation erforderlich.

DEFINITION 1.4.7.2

Sei $T \subseteq \{1, 2, \ldots, N\}$ eine nichtleere Teilmenge von Zeitpunkten. Der **linke Randpunkt** $t_{\min}$ von T sei der kleinste, der **rechte Randpunkt** $t_{\max}$ der größte Wert von T. Alle anderen Punkte heißen **innere Punkte** von T. Wir schreiben $T^<$ für $T \setminus \{t_{\max}\}$. Für $t \in T$ mit $t > t_{\min}$ bezeichne t_L den **linken Nachbarn** von t in T, formal

$$t_L = \max\{t' : t' \in T, t' < t\},$$

und entsprechend t_R den rechten Nachbarn von t in T für $t < t_{\max}$. ■

Da wir Trendumkehrpunkte bestimmen wollen, sind Punkte uninteressant, an denen die Reihe konstant ist bzw. monoton steigt oder fällt. Es interessieren also nur lokale Maxima und Minima. Die entsprechende Teil-Reihe wird in zwei Schritten erzeugt. Im ersten Schritt wird von einem Bindungsstrang, d. h. einer Sequenz gleicher Werte, nur der jeweils letzte ausgewählt. Im zweiten Schritt werden die Zeitpunkte t, für die x_t im Trend liegt, gelöscht. Die Potentiale der in diesen beiden Schritten zu löschenden Zeitpunkte werden Null gesetzt. Daran schließt sich ein Reduktionsschritt an, der wiederholt wird, bis nur noch $t_{\min}$ und $t_{\max}$ übrigbleiben.

DEFINITION 1.4.7.3 Trend-Umkehr-Potentiale

Sei $(x_t)_{t \in T}$ mit $T_0 \subseteq \{1, 2, \ldots, N\}$ eine Zeitreihe mit mindestens zwei Werten. Die durch den folgenden Algorithmus eindeutig bestimmten Werte $p(t)$ heißen **Trend-Umkehr-Potentiale**.

Schritt 1: Für alle $t < t_{\max}$ mit $x_{t_R} - x_t = 0$ wird $p(t) = 0$ gesetzt und t als gelöscht betrachtet.
Sei T_1 die Menge der nicht gelöschten Zeitpunkte.

Schritt 2: Falls T_1 einelementig ist, wird $p(t) = 0$ gesetzt und die Iteration abgebrochen.
Für alle inneren Punkte t von T_1 mit $x_{t_L} < x_t < x_{t_R}$ bzw. $x_{t_L} > x_t > x_{t_R}$ setzt man $p(t) = 0$ und betrachtet t als gelöscht.
Sei T_2 die Menge der nicht gelöschten Zeitpunkte.

Schritt 3: Sei t' der kleinste Zeitpunkt, an dem der minimale Abstand von je zwei aufeinanderfolgenden Beobachtungen in T_2 beginnt:

$$t' = \min\{t : t \in T_2^<, \quad |x_{t_R} - x_t| = \min\{|x_{s_R} - x_s| : s \in T_2^<\}\}.$$

i) Falls t' und t'_R beides innere Punkte oder beides Randpunkte von T_2 sind, wird ihr Potential definiert durch $p(t') = |x_{t'_R} - x_{t'}|$.
t' und t'_R werden als gelöscht betrachtet.

ii) Für $t' = t_{\min}$ und $t'_R < t_{\max}$ wird $p(t') = |x_{t'_R} - x_{t'}|$ gesetzt und t' als gelöscht betrachtet.
Für $t'_R = t_{\max}$ und $t' > t_{\min}$ wird $p(t_{\max}) = |x_{t'_R} - x_{t'}|$ gesetzt und $t_{\max}$ als gelöscht betrachtet.

Sei T_3 die Menge der nicht gelöschten Zeitpunkte.

Schritt 4: Setze $T_2 = T_3$ und gehe zurück zu Schritt 3, falls T_2 nicht leer ist. ■

BEMERKUNG 1.4.7.4

Nach Schritt 1 gilt für $(x_t)_{t \in T_1}$ offenbar, daß alle Differenzen $x_{t'_R} - x_{t'} \neq 0$ sind, d.h. $(x_t)_{t \in T_1}$ ist bindungsfrei. Die nach Schritt 2 verbleibende Zeitreihe ist in etwas lockerer Sprechweise eine Zig-Zag-Reihe. Genauer gilt für drei aufeinanderfolgende Zeitpunkte $t_L, t, t_R \in T_2$ entweder $x_{t_L} < x_t > x_{t_R}$ oder $x_{t_L} > x_t < x_{t_R}$. Die Vorzeichen der ersten Differenzen sind also alternierend. Nach dem vorletzten Durchgang bleiben in der Tat zwei Punkte übrig. Waren es vorher vier, so sind zwei oder einer gelöscht worden. Waren es drei, so waren beide Reihensegmente Randstücke, von denen nur eines mittels Löschung eines Randpunktes herausgeschnitten wurde. Es verbleiben genau die globalen Extrema der Ausgangsreihe vor dem letzten Durchgang des Rekursionsschrittes. Somit bekommen die Punkte, an denen sich das Maximum und das Minimum der Reihe befinden, die Spannweite der Zeitreihenwerte als Potential zugeordnet.

BEISPIEL 1.4.7.5

Um den Algorithmus zu illustrieren, ist in der folgenden Tabelle eine Zeitreihe angegeben sowie die Teilreihen, welche nach den verschiedenen Schritten verblieben sind. In der untersten Zeile sind die Potentiale zu den einzelnen Zeitpunkten angegeben.

	Zeitpunkte											
	1	2	3	4	5	6	7	8	9	10	11	12
nach	Zeitreihenwerte											
Schritt	2	1	1	6	5	9	8	4	5	5	5	2
1	2		1	6	5	9	8	4			5	2
2	2		1	6	5	9		4			5	2
3			1	6	5	9		4			5	2
4			1			9		4			5	2
5			1			9						2
6			1			9						
7												
	Potential-Werte											
	−1	0	8	−1	1	−8	0	1	0	0	−1	−7

■

Wegscheider hat den Begriff des Potentials nicht auf Grund der Verbindungen zur Potentialtheorie, vgl. Doob [1984] gewählt. Er wollte damit vielmehr auf die Möglichkeiten „potentieller" Gewinne anspielen. Ein „perfekter Prophet", d.h. jemand,

der die gesamte Kursreihe kennt, könnte ja vor jedem Anstieg kaufen und vor jedem Abstieg verkaufen. Sein Gesamtgewinn wäre dann $\Sigma_{t\in T^<}|x_{t_R}-x_t|$.

DEFINITION 1.4.7.6

Das **Bewegungspotential** einer Zeitreihe $(x_t)_{t\in T}$ ist definiert als

$$P(T)=\sum_{t\in T^<}|x_{t_R}-x_t|. \qquad [1.4.7.1]$$

Es gibt nun einen verblüffend einfachen Zusammenhang zwischen Bewegungspotential und Trend-Umkehr-Potentialen.

SATZ 1.4.7.7 Zerlegungssatz für Bewegungspotentiale

Das Bewegungspotential $P(T)$ einer Zeitreihe $(x_t)_{t\in T}$ ergibt sich aus den Punktpotentialen $p(t)$ vermöge

$$P(T)=\sum_{t\in T}|p(t)|-(x_{\max}-x_{\min}), \qquad [1.4.7.2]$$

wobei $x_{\max}$ den größten und $x_{\min}$ den kleinsten Wert der Reihe bezeichnen.

Beweis: Siehe Wegscheider [1994].

Da der Algorithmus sukzessive die jeweils kleinsten Trends entfernt, erlaubt die schrittweise Potentialreduktion nach 1.4.7.3 eine Potentialzerlegung in dem Sinne, daß man das Potential einer Reihe aufspalten kann in Anteile, die auf große Trends zurückgeführt werden können und Anteile, die zu mittleren und kleinen Trends gehören. Diese Zerlegung gewährt über die Darstellung der Trends hinaus einen Einblick in die Trendstruktur einer Zeitreihe. Die Möglichkeit der Potentialreduktion zeigt auch, daß die hier im Ansatz vorgestellte Methode auf der Stufe der Fourierzerlegung angesiedelt ist.

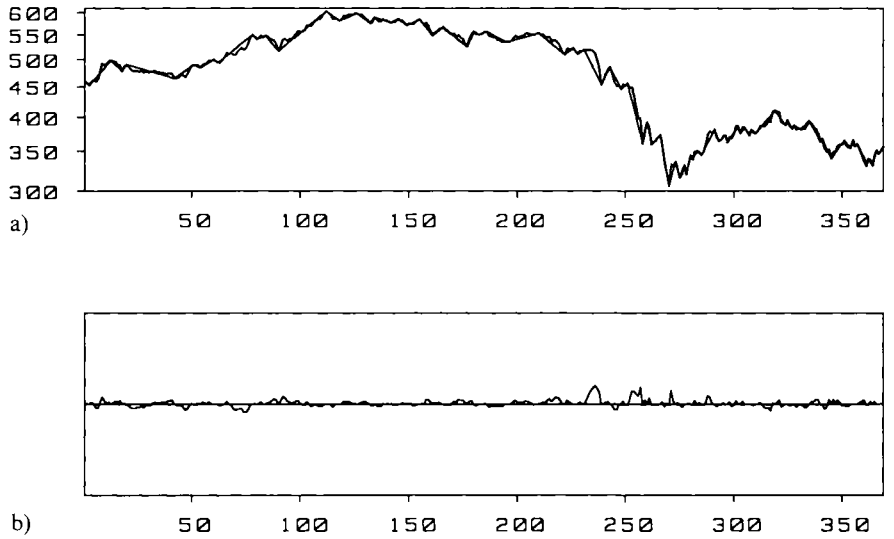


Abb. 1.4.7.1 *Zerlegung der Reihe IBM in zwei Filterwertbänder: a) Reihe und Bewegungen von mindestens 1%, b) Bewegungen unter 1%* ■

BEISPIEL 1.4.7.8

Für die bereits oben betrachtete Reihe IBM erhalten wir die angegebene Aufsplittung in zwei Trendbänder. Eine weitere Aufspaltung ist in diesem Fall nicht sinnvoll. – Sie besteht im wesentlichen aus großen und kleinen Bewegungen; Bewegungen mittlerer Größenordnung fehlen. ■

Mit Hilfe der Trend-Umkehr-Potentiale kann zu einem vorgegebenem Trendniveau eine idealisierte Version der Ausgangszeitreihe angegeben werden, die sämtliche Trends der Zeitreihe enthält, deren ‚Höhenunterschied' das vorgegebene Niveau überschreitet oder einhält.

DEFINITION 1.4.7.9

$(x_t)_{t\in T}$ sei eine Zeitreihe mit mindestens zwei Elementen und sei $f > 0$. Mit

$$T^f = \{t: t \in T, p(t) > f\}$$

heißt dann $(x_t)_{t\in T^f}$ die f-Filterreihe von $(x_t)_{t\in T}$.

Das Bewegungspotential von $(x_t)_{t\in T^f}$, $P(T^f)$, wird als ***f*-Potential** von $(x_t)_{t\in T}$ bezeichnet. ■

Man beachte, daß f normalerweise die gleiche Einheit wie die Reihe hat. Die f-Potentiale messen die Dispersion der Zeitreihe, die auf Trends einer bestimmten Mindestlänge (genauer: einem bestimmten Niveauunterschied) zurückgeführt werden kann. Aus den Potentialreihen ergeben sich auf natürliche Weise eine Vielzahl weitere Dispersionsmaße für die Ausgangsreihe. Dazu sei auf die Originalarbeit von Wegscheider verwiesen.

1.5 Transformation von Zeitreihen durch Filter

1.5.1 Vorbemerkung

Neben der Bestimmung des globalen Trends ist oft eine Glättung der Zeitreihe von Interesse. Glättung bedeutet Ausschaltung von irregulären Schwankungen durch lokale Approximationen. Der Vorteil liegt u.a. darin, daß man mit Polynomen niedrigen Grades auskommt.

Im klassischen Komponentenmodell entspricht dies der näherungsweisen Bestimmung der glatten Komponente.

Es ist naheliegend, im ersten Ansatz zum Glätten einer Zeitreihe die Beobachtung x_t durch ein lokales arithmetisches Mittel y_t zu ersetzen:

$$y_t = \frac{1}{2q+1} \sum_{u=-q}^{+q} x_{t-u} \quad t = q+1, \ldots, N-q\,.$$

Dies ist ein Beispiel für eine lineare Transformation einer Zeitreihe (x_t) in eine andere (y_t). Da lineare Transformationen von großer Bedeutung sind, sollen die einzelnen Transformationen unter dem Blickwinkel eines allgemeinen Konzepts besprochen werden.

DEFINITION 1.5.1.1
Eine lineare Transformation L einer Zeitreihe (x_t) in eine andere (y_t) gemäß

$$y_t = Lx_t = \sum_{u=-q}^{s} a_u x_{t-u} \quad t = s+1, \ldots, N-q$$

wird als **linearer Filter** L bezeichnet. Anstatt durch das formale Symbol L wird der Filter auch durch seine **Gewichte** a_u in der Form $(a_{-q}, \ldots, a_s)$ bzw. (a_u) angegeben. Die Anwendung eines Filters auf eine Zeitreihe (x_t) wird als **Filtration** der Reihe bezeichnet. (x_t) heißt auch der **Input** und die gefilterte Reihe (y_t) der **Output** des Filters.

■

Bei der Filtration einer Zeitreihe (x_t) ist zu beachten, daß die gefilterte Reihe (y_t) i.a. kürzer ist als die Input-Reihe. Im Fall $s > 0$ wird der Anfang, im Fall $q > 0$ das Ende gekappt.

Relevante Beispiele für lineare Filter werden in den folgenden Unterabschnitten besprochen. Auf die Angabe der Summationsgrenzen soll im folgenden zur Vereinfachung der Schreibweise verzichtet werden.

1.5.2 Gleitende Durchschnitte

Das am Beginn des Abschnittes erwähnte Verfahren zur Glättung von Zeitreihen wird als einfacher gleitender Durchschnitt bezeichnet. Der entsprechende Filter hat die Gewichte $a_u = 1/(2q+1)$ mit $\Sigma a_u = 1$. Wir lassen nun beliebige gewichtete Durchschnitte zu, verlangen jedoch $\Sigma a_u = 1$.

DEFINITION 1.5.2.1
Ein linearer Filter (a_u) mit $\Sigma a_u = 1$ heißt ein **gleitender Durchschnitt**. Im Fall $a_u = 1/(2q+1)$, $u = -q, \ldots, q$, spricht man von einem **einfachen gleitenden Durchschnitt**.

■

Wir betrachten die Glättungseigenschaft von einfachen gleitenden Durchschnitten anhand eines Beispiels. Es ist von der Konstruktion dieser Filter her klar, daß die Glättung um so stärker ist, je mehr Werte jeweils einbezogen werden. Bei der Frage, wie groß q zu wählen ist, ist zu beachten, daß an den Rändern der Zeitreihe jeweils q Werte gekappt werden.

BEISPIEL 1.5.2.2
Die folgende Abbildung zeigt die Reihe KONKURSE sowie 2 einfache gleitende Durchschnitte, bei denen je 3 bzw. je 5 Werte einbezogen wurden. Formelmäßig ergeben sich also

$$y_t = \tfrac{1}{3}(x_{t-1} + x_t + x_{t+1})$$
$$z_t = \tfrac{1}{5}(x_{t-2} + x_{t-1} + x_t + x_{t+1} + x_{t+2})$$

Einfache gleitende Durchschnitt können auch für gerade Anzahlen von Werten bestimmt werden. Der Output ist dann aber jeweils der Mitte zwischen 2 Zeitpunk-

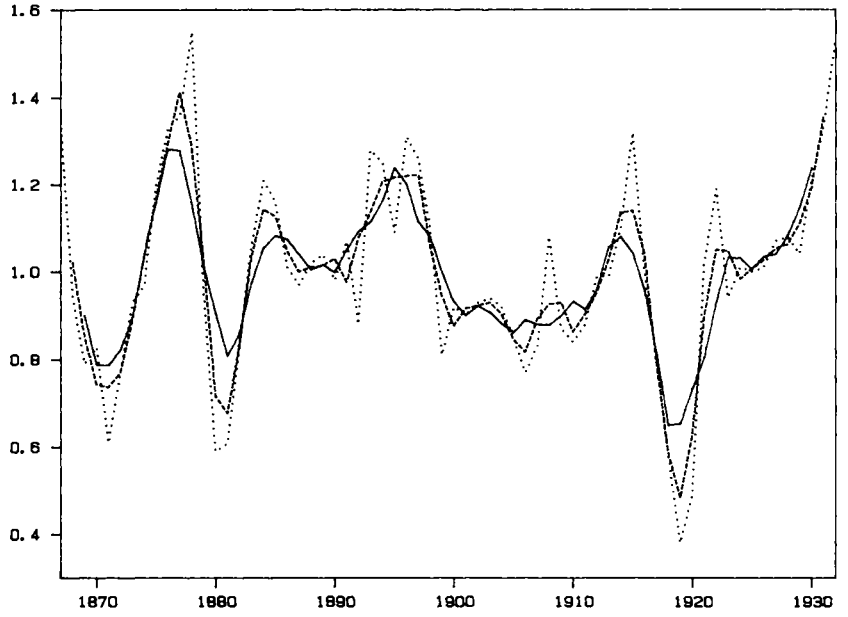


Abb. 1.5.2.1 *KONKURSE* (· · · · ·) *mit gleitenden Durchschnitten der Länge 3* (- - - -) *sowie der Länge 5* (—)

■

ten zugeordnet, so etwa $\frac{1}{4}(x_1 + x_2 + x_3 + x_4)$ dem Zeitpunkt 2.5. Man kann diesen ungewünschten Effekt dadurch vermeiden, daß man jeweils über zwei aufeinanderfolgende Werte des Outputs mittelt. Im Beispiel berechnet man also

$$y_{2.5} \triangleq \tfrac{1}{4}(x_1 + x_2 + x_3 + x_4)$$

und

$$y_{3.5} \triangleq \tfrac{1}{4}(x_2 + x_3 + x_4 + x_5),$$

als Mittelwert entsteht

$$\begin{aligned} y_3 &\triangleq \tfrac{1}{2} y_{2.5} + \tfrac{1}{2} y_{3.5} \\ &= \tfrac{1}{8} x_1 + \tfrac{1}{4} x_2 + \tfrac{1}{4} x_3 + \tfrac{1}{4} x_4 + \tfrac{1}{8} x_5 . \end{aligned}$$

Die **lokale Approximation einer Zeitreihe durch Polynome** führt ebenfalls zu gleitenden Durchschnitten, wenn jeweils nur der für den mittleren Zeitpunkt erhaltene Wert y_t des Polynoms als geglätteter Wert anstelle von x_t genommen wird. Wir führen dies nur für ein formales Beispiel aus. Eine ausführliche Diskussion befindet sich bei Kendall [1973].

BEISPIEL 1.5.2.3

Gegeben sei eine Zeitreihe $x_1, x_2, \ldots, x_N$. Den ersten fünf Werten soll ein Polynom $m(t)$ zweiten Grades angepaßt werden. Die Anpassung gemäß Abschnitt 1.4. führt zu

$$\sum_{t=1}^{5} (x_t - \beta_1 - \beta_2 t - \beta_3 t^2)^2 \stackrel{!}{=} \min.$$

Zur Vereinfachung werden bei diesem Problem o.B.d.A. die Zeitpunkte neu durchnumeriert von -2 bis $+2$. Dann ist

$$\sum_{t=-2}^{2} (x_t - \beta_1 - \beta_2 t - \beta_3 t^2)^2 \stackrel{!}{=} \min$$

zu lösen. Wegen $\sum_{t=-2}^{2} t = \sum_{t=-2}^{2} t^3 = 0$ reduziert sich das System der Normalgleichungen hier zu

$$\begin{aligned} 5\hat{\beta}_1 \quad + 10\hat{\beta}_3 &= \Sigma x_t \\ 10\hat{\beta}_2 \qquad\quad &= \Sigma t x_t \\ 10\hat{\beta}_1 \quad + 34\hat{\beta}_3 &= \Sigma t^2 x_t . \end{aligned}$$

Der Wert des Polynoms wird nur für den mittleren Zeitpunkt ($t = 0$ in der neuen Numerierung) benötigt. Daher genügt es, $\hat{\beta}_1$ zu bestimmen. Die erste und die dritte Gleichung liefern:

$$35\hat{\beta}_1 = 17\Sigma x_t - 5\Sigma t^2 x_t ,$$

bzw. in den einzelnen x_t ausgedrückt:

$$\hat{\beta}_1 = \tfrac{1}{35}(-3x_{-2} + 12x_{-1} + 17x_0 + 12x_{+1} - 3x_{+2}) .$$

Da die Berechnung des Polynomwertes offensichtlich unabhängig von der Numerierung der Zeitpunkte ist, erhalten wir als geglätteten Wert aus den ersten 5 Beobachtungen

$$y_3 = \tfrac{1}{35}(-3x_1 + 12x_2 + 17x_3 + 12x_4 - 3x_5) .$$

Entsprechend ergeben sich die geglätteten Werte von je 5 aufeinanderfolgenden Beobachtungen zu

$$y_t = \tfrac{1}{35}(-3x_{t-2} + 12x_{t-1} + 17x_t + 12x_{t+1} - 3x_{t+2}) .$$

Die Glättung durch ein Polynom 2. Grades ist also eine Filtration der Zeitreihe durch den linearen Filter $(-\frac{3}{35}, \frac{12}{35}, \frac{17}{35}, \frac{12}{35}, -\frac{3}{35})$.

Die Anpassung eines Polynoms 1. Grades an jeweils $2q + 1$ Werte entspricht gerade einem einfachen gleitenden Durchschnitt.

■

Bei einer Filtration durch einen Filter der Länge $2q + 1$ gehen an beiden Rändern der Zeitreihe jeweils q Werte verloren. Dies ist insbesondere für den aktuellen Rand der Zeitreihe oft nicht akzeptabel. Es existieren daher eine Reihe von Verfahren der **Randergänzung**.

Eine generelle Technik ist die Anwendung eines Prognoseverfahrens, um die Reihe q Werte in die Zukunft hinein zu verlängern. Auf diese erweiterte Reihe kann dann ein Filter der Länge $2q + 1$ angesetzt werden.

Im Fall der lokalen Approximation durch Polynome liegt es andererseits nahe, bei der Polynomanpassung an die letzten Reihenwerte die Koeffizienten des Polynoms zu schätzen und daraus angepaßte Werte für die Daten des aktuellen Rands zu berechnen.

BEISPIEL 1.5.2.4

Im Fall der Anpassung eines Polynoms 1. Grades an jeweils $2q+1$ Werte ergibt der letzte Ansatz die angepaßten Werte

$$y_{N-q+s} = \frac{1}{2q+1} \sum_{t=-q}^{q} \left(1 + \frac{3s \cdot t}{q(q+1)}\right) x_{N-q+t}$$

für $s = q, q-1, \ldots, 1$. Der Output läßt sich wieder als Anwendung eines Filters denken, dessen Gewichte allerdings nicht zeitinvariant sind. Für $q = 1$ z. B. findet man den letzten Outputwert

$$y_N = -\tfrac{1}{6} x_{N-2} + \tfrac{2}{6} x_{N-1} + \tfrac{5}{6} x_N .$$

■

Die Filter der hier betrachteten gleitenden Durchschnitte haben die Besonderheit $a_{-u} = a_u$ gemeinsam. Diese Eigenschaft wird durch die folgende Definition herausgehoben.

DEFINITION 1.5.2.5

Ein linearer Filter $(a_{-q}, \ldots, a_s)$ heißt **symmetrisch**, wenn $q = s$ und $a_{-u} = a_u$, für $u = 1, \ldots, q$. Sonst heißt er asymmetrisch.

■

Ein Beispiel für einen asymmetrischen Filter werden wir im folgenden Abschnitt kennenlernen.

1.5.3 Differenzenfilter

Bei der Anpassung eines Polynoms an eine Zeitreihe und bei der lokalen Glättung durch Polynome stellte sich die Frage, welcher Grad für das Polynom gewählt werden soll. Wir skizzieren hier den Lösungsansatz der **„Variaten Differenzen"**. Die Grundlage für den Ansatz liefert das folgende Lemma.

LEMMA 1.5.3.1

$f(t)$ sei ein Polynom vom Grade $p > 0 : f(t) = a_0 + a_1 t + \ldots + a_p t^p$.
Dann ist

$$g(t) = f(t) - f(t-1)$$

ein Polynom vom Grade höchstens $p-1$.

Beweis: Es ist

$$\begin{aligned} g(t) &= f(t) - f(t-1) \\ &= a_0 + a_1 t + \ldots + a_p t^p - (a_0 + a_1 (t-1) + \ldots + a_p (t-1)^p) \\ &= a_0 - a_0 \\ &\quad + a_1 t - a_1 t + a_1 \\ &\quad + \\ &\quad \vdots \\ &\quad + a_p t^p - a_p t^p + a_p \binom{p}{1} t^{p-1} - a_p \binom{p}{2} t^{p-2} + \ldots + a_p (-1)^p t^0 . \end{aligned}$$

Da sich die Glieder mit dem Exponenten p aufheben, folgt die Behauptung. ■

Die Differenzenbildung reduziert den Polynomgrad um 1. Wird das Verfahren $p-1$-mal angewandt, so erhält man bei einem Polynom p-ten Grades einen konstanten Wert.

Die Anwendung der Differenzenbildung auf Zeitreihen führt wieder auf eine lineare Transformation oder Filtration.

DEFINITION 1.5.3.2

Der lineare Filter Δ, der definiert ist durch

$$\Delta x_t = x_t - x_{t-1}, \quad t = 2, 3, \ldots, N$$

heißt **Differenzenfilter** 1. Ordnung.

Differenzenfilter p-ter Ordnung, $p > 1$, sind rekursiv definiert durch

$$\Delta^p x_t = \Delta^{p-1} x_t - \Delta^{p-1} x_{t-1}, \quad t = p+1, \ldots, N.$$

■

BEISPIEL 1.5.3.3

Der Differenzenfilter 2. Ordnung ist gegeben durch

$$\begin{aligned}\Delta^2 x_t &= \Delta x_t - \Delta x_{t-1} \\ &= x_t - x_{t-1} - (x_{t-1} - x_{t-2}) \\ &= x_t - 2x_{t-1} + x_{t-2}.\end{aligned}$$

Er kann also auch durch seine Gewichte angegeben werden:

$$a_0 = 1,\ a_1 = -2,\ a_2 = 1.$$

■

Ist bei dem der Trendbestimmung zugrundegelegten Ansatz [1.4.2.1] $m(t)$ also ein Polynom, so läßt sich durch Differenzenbildung der Trend eliminieren bzw. der Grad des Polynoms bestimmen.

Für ökonomische Zeitreihen ist es oft ausreichend, die Differenzen 1. Ordnung zu bilden, um Veränderungen im Niveau zu beseitigen. Da die ersten Differenzen als Zuwächse auch inhaltlich bedeutsam sind, wird dieses Vorgehen häufig praktiziert. Falls nötig, werden noch die zweiten Differenzen gebildet, um zusätzliche Nichtstationarität der Steigungen zu beseitigen. Speziell der Box-Jenkins-Ansatz verwendet die Differenzenbildung extensiv.

BEISPIEL 1.5.3.4

Aus der Reihe STROM ermitteln wir die Reihe (x_t) der Jahresproduktionen für die Jahre 1955–1974. Um dieser Reihe ein Trendpolynom anzupassen, werden die ersten und zweiten Differenzen berechnet und gezeichnet.

Da die ersten Differenzen noch einen Trend aufweisen, die zweiten jedoch um Null schwanken, ist der Reihe (x_t) ein Trendpolynom mit $p = 2$ anzupassen.
Bei saisonalen Daten benutzt man gerne sogenannte **saisonale Differenzen** (etwa $\Delta_{12} x_t = x_t - x_{t-12}$ für Monatsdaten) um Saisonschwankungen zu eliminieren, vgl. Abschnitt 1.7.

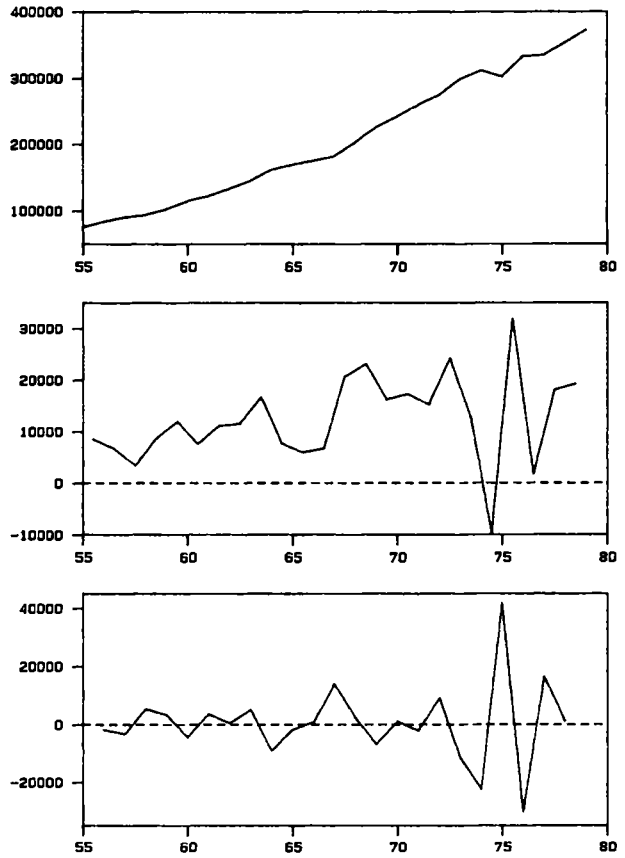
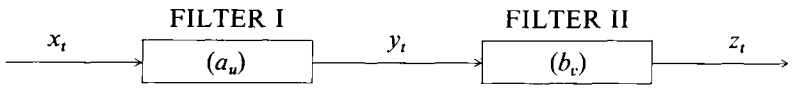


Abb. 1.5.3.1 *Jährliche Stromproduktion, erste und zweite Differenzen* ■

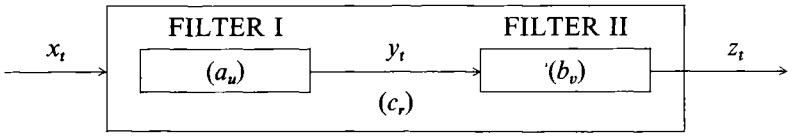
1.5.4 Faltungen

Höhere Differenzen sind ein Beispiel für die Hintereinanderausführung von Filtern. Schematisch läßt sich dies darstellen durch



Der Input ist hier die Reihe (x_t). Darauf wird der Filter I angewandt. Er hat als Output das Zwischenergebnis (y_t), das durch den Filter II schließlich in den Output (z_t) transformiert wird.

Wir wollen untersuchen, wie sich die beiden Filter I und II zusammensetzen lassen, damit der Output (z_t) durch die Anwendung eines einzigen Filters (c_t) aus dem Input (x_t) hervorgeht. Dies läßt sich mittels folgender Skizze veranschaulichen:



Mit den in der Skizze angegebenen Bezeichnungen erhalten wir

$$
\begin{aligned}
z_t &= \sum_v b_v y_{t-v} \\
&= \sum_v b_v \Big(\sum_u a_u x_{t-v-u}\Big) && \text{Umordnen der Summation} \\
&= \sum_u \sum_v b_v a_u x_{t-v-u} && \text{Substitution } r = u + v \\
&= \sum_r \Big(\sum_v b_v a_{r-v}\Big) x_{t-r} \\
&= \sum_r c_r x_{t-r}
\end{aligned}
$$

Die Gewichte des kombinierten Filters sind also

$$c_r = \sum_v b_v a_{r-v}$$

Dabei erstreckt sich die Summation über alle jeweils möglichen Indexkombinationen.

DEFINITION 1.5.4.1

Die **Faltung** zweier Filter mit den Gewichten (a_u) bzw. (b_v) ist der durch die Gewichte (c_r) gegebene Filter. Dabei ist

$$c_r = \sum_v b_v a_{r-v}.$$

Für die Faltung wird auch $(c_r) = (a_u) * (b_v)$ geschrieben.

■

BEISPIEL 1.5.4.2

Anschaulich läßt sich die Faltung zweier Filter (a_u) und (b_v) nach folgendem Schema illustrieren, das der Einfachheit halber für $(a_{-2}, a_{-1}, a_0, a_1) = (1, 0, 1, 2)$, $(b_{-1}, b_0, b_1) = (1, 1, 3)$ angegeben ist.

	$a_{-2} = 1$	$a_{-1} = 0$	$a_0 = 1$	$a_1 = 2$
$b_{-1} = 1$	1	0	1	2
$b_0 = 1$	1	0	1	2
$b_1 = 3$	3	0	3	6

Die Elemente der Matrix sind die Produkte $b_i a_j$ und man erhält für die Faltung

$$c_{-3} = 1,\ c_{-2} = 1 + 0 = 1,\ c_{-1} = 3 + 0 + 1 = 4,$$
$$c_0 = 0 + 1 + 2 = 3,\ c_1 = 3 + 2 = 5,\ c_2 = 6.$$

■

Die Faltungsoperation ist kommutativ und assoziativ:

$$(a_u) * (b_v) = (b_v) * (a_u);$$
$$(a_u) * ((b_v) * (c_w)) = ((a_u) * (b_v)) * (c_w).$$

Die Faltung ist formal eine Verknüpfung zweier (endlicher) Zahlenfolgen. Genau die gleiche Verknüpfung findet bei der Filtration einer Zeitreihe durch einen linearen Filter statt. Die Filtration einer Zeitreihe kann also auch als Faltung einer Zeitreihe mit einem linearen Filter aufgefaßt werden.

Faltungen können auf formales Ausmultiplizieren von Polynomen zurückgeführt werden, wie die folgenden Überlegungen zeigen. Grundlage dafür ist ein sehr einfacher Filter.

DEFINITION 1.5.4.3

Als **Shift** (oder Backshift) **-Operator** bezeichnen wir den Filter B mit

$$Bx_t = x_{t-1}.$$

■

Der Filter B ordnet also dem Input

$$\ldots x_1\, x_2\, x_3\, x_4 \ldots$$

den Output

$$\ldots x_0\, x_1\, x_2\, x_3 \ldots$$

zu; er verschiebt die Zeitreihe um eine Zeiteinheit. Hintereinanderausführungen von B schreiben wir als Potenzen; allgemein ist

$$B^p x_t = x_{t-p}.$$

Für $p > 0$ ergeben sich Rückwärtsverschiebungen um p Zeiteinheiten, für $p = 0$ die Identität $B^0 = 1$, für $p < 0$ Vorwärtsverschiebungen. Mit den Exponenten kann wie üblich gerechnet werden, es ist z. B. $B^p B^q = B^{p+q}$.

Jeder Filter kann nun als **gewichtete Summe der Shiftoperatoren** ausgedrückt werden. Ein Filter mit den Gewichten a_u ist in der Form

$$A(B) = \Sigma a_u B^u$$

darstellbar. Zum Beispiel gilt für den Differenzenfilter

$$\Delta = 1 - B,$$

denn es ist

$$\Delta x_t = x_t - x_{t-1} = 1x_t - Bx_t = (1 - B)x_t.$$

Faltungen können dann einfach durch Ausmultiplizieren und Zusammenfassen von Gliedern mit gleichen Exponenten berechnet werden. Für den Differenzenfilter 2. Ordnung ergibt sich

$$\Delta^2 = (1 - B)(1 - B) = 1 - 2B + B^2,$$

woraus die Gewichte $a_0 = 1, a_1 = -2, a_2 = 1$ hervorgehen.

Diese Darstellung erlaubt die Anwendung von Regeln der üblichen Algebra auf

Faltungsoperationen. Als Beispiel wollen wir die explizite Berechnung des Differenzenfilters p'ter Ordnung betrachten.

BEISPIEL 1.5.4.4

Der Satz über die Entwicklung eines Binoms

$$(a+b)^p = \sum_{s=0}^{p} \binom{p}{s} a^{p-s} b^s = a^p + pa^{p-1}b + \frac{p(p-1)}{2} a^{p-2}b^2 + \ldots + b^p$$

bringt für den Differenzenfilter p'ter Ordnung

$$\Delta^p = (1-B)^p = \sum_{s=0}^{p} \binom{p}{s} (-1)^s B^s,$$

woraus man die Gewichte von Δ^p direkt ablesen kann.

■

1.5.5 Exponentielle Glättung

Die bisher betrachteten Filter haben alle eine feste Anzahl von Gewichten. Oft ist es aber auch von Interesse, neu hinzukommende Werte zusätzlich einzubeziehen, so daß im Prinzip immer mehr Beobachtungen für die Ermittlung des Output berücksichtigt werden. Am geschicktesten geschieht dies, indem der Output y_N der ersten N Werte $x_1, \ldots, x_N$ mit der neuen Beobachtung x_{N+1} kombiniert wird. Auf diese Weise erhalten wir **rekursive Filter**.

BEISPIEL 1.5.5.1

Bei einer Zeitreihe $x_1, \ldots, x_N$ kann die Bestimmung des arithmetischen Mittels $\bar{x}$ rekursiv vorgenommen werden. Liegt $\bar{x}_N = \Sigma_{t=1}^{N} x_t / N$ vor und wird x_{N+1} beobachtet, so ergibt sich:

$$\bar{x}_{N+1} = \frac{1}{N+1} \sum_{t=1}^{N+1} x_t = \frac{N}{N+1} \bar{x}_N + \left(1 - \frac{N}{N+1}\right) x_{N+1} \qquad [1.5.5.1]$$

■

Lassen wir in [1.5.5.1] allgemeinere Gewichte β, $0 < \beta < 1$, zu, so erhalten wir den **Filter des einfachen exponentiellen Glättens**. Er wurde von Brown [1962] als eigenständiger Prognosefilter vorgeschlagen. Er ging bei der Ableitung von dem einfachen Modell

$$X_t = \mu + \varepsilon_t$$

mit unabhängigen, identisch verteilten Störungen aus. Das ist gerade das Modell, das im Prinzip auch dem Beispiel 1.5.5.1 zugrundegeliegt. Bezeichnen wir die Prognose des Wertes x_{t+h}, $h > 0$, unter Verwendung der Werte $x_1, \ldots, x_t$ mit $\hat{x}_{t,h}$, so ergibt sich die folgende Aufdatierungsformel für die Ein-Schritt-Prognosen, d.h. für $h = 1$:

$$\hat{x}_{t,1} = \beta \hat{x}_{t-1,1} + (1-\beta) x_t. \qquad [1.5.5.2]$$

Der geringe Speicherplatzbedarf ist neben der formalen Einfachheit einer der Gründe für die weite Verbreitung dieses Verfahrens. Er kommt insbesondere im ökonomischen Bereich zum Tragen, wenn mehrere hundert Zeitreihen zugleich prognostiziert werden müssen.

Die Prognose mit der Formel [1.5.5.2] wird als **einfaches exponentielles Glätten** bezeichnet. Der Name rührt daher, daß $\hat{x}_{t,1}$ auch darstellbar ist als exponentiell gewichtete Summe der vergangenen Werte. Bei unendlicher Vergangenheit gilt:

$$\hat{x}_{t,1} = (1-\beta) \sum_{u=0}^{\infty} \beta^u x_{t-u}. \qquad [1.5.5.3]$$

Für die praktische Anwendung ist eine Initialisierung sowohl in [1.5.5.2] als auch in [1.5.5.3] nötig. In der Regel wird $\hat{x}_{1,1} = x_1$ gewählt. Dies entspricht der Wahl $x_0 = x_{-1} = x_{-2} = \ldots = x_1$.

Der **Glättungsparameter** β steuert das Verhalten der 1-Schritt-Prognosen. Bei kleinem $\beta (\beta \approx 0)$ bekommt die letzte Beobachtung ein großes Gewicht. Die Prognosen können von einem zum nächsten Mal stark schwanken. Große Werte von β $(\beta \approx 1)$ bewirken eine größere Glättung, die Anpassung an eine neue Entwicklung, z. B. bei einem Strukturbruch, wird aber langsamer vollzogen.

BEISPIEL 1.5.5.2

Das **Verhalten der 1-Schritt-Prognosen** des exponentiellen Glättens wird anhand folgender vereinfachter Reihen in Abb. 1.5.5.1 charakterisiert. Diese spiegeln eine periodische Schwingung, einen Ausreißer oder Impuls, eine Niveauänderung und einen einsetzenden Trend wider. ■

Bei der Wahl des Glättungsparameters gibt es verschiedene Ansätze. Eine Möglichkeit, das für eine Reihe $(x_t)_{t=1,\ldots,N}$ geeignete β zu bestimmen, besteht darin, für verschiedene β die 1-Schritt-Prognosen mit den Reihenwerten zu vergleichen und den Wert zu wählen, für den gilt:

$$\sum_{t=m}^{N-1} [x_{t+1} - \hat{x}_{t,1}(\beta)]^2 \stackrel{!}{=} \min_{0<\beta<1}. \qquad [1.5.5.4]$$

Die untere Summationsgrenze m wird dabei so groß gewählt, daß der Effekt des Startwertes vernachlässigbar ist.

Für die meisten praktischen Belange reicht es, das Minimum über der diskreten Menge $= 0.1, 0.2, \ldots, 0.9$ zu bestimmen. Das so erhaltene β kann dann verwendet werden, solange sich keine signifikante Änderung im Verhalten der Zeitreihe ergibt. Um derartige Änderungen automatisch zu erfassen, wurden Varianten des exponentiellen Glättens vorgeschlagen, bei der β adaptiv angepaßt wird. Siehe dazu Günther [1980], Mertens [1981] und Trigg/Leach [1967]. Bei der Analyse zweier Studien stellte Ekern [1981] jedoch fest, daß die adaptive exponentielle Glättung keine besseren Prognoseergebnisse lieferte als das einfache Verfahren mit geeignet gewähltem Glättungsparameter.

BEISPIEL 1.5.5.3

Das einfache Modell, von dem Brown bei seiner Ableitung des exponentiellen Glättens ausgegangen ist, weist schon auf einen wesentlichen Grundzug des Verfahrens hin: Bei Reihen ohne systematischer Änderung wird vor allem das Mittel geschätzt.

Die Konsequenz dieser Eigenschaft kann bei empirischen Reihen zur Unbrauchbarkeit der Prognose führen. Dies wird etwa bei der Reihe SCHMIERSTOFF, der