

Universitäts-Rechenzentrum Trier



AWS.SPSS.8

Trier, den 5.6.1997

Zeitreihenanalyse mit SPSS-Trends

1	KURZE CHARAKTERISIERUNG DER ZEITREIHENANALYSE	5
1.1	Beispiele für empirische Zeitreihen und Auswertungsfragestellungen.....	5
1.1.1	Guthaben im Nullsummenroulett bei strenger Präferenz für Rot.....	5
1.1.2	Monatliche Verkaufszahlen für ein Produkt	5
1.1.3	Lichtstärke eines pulsierenden Sterns in 580 aufeinanderfolgenden Nächten	6
1.1.4	Anzahl defekter Produkte in 187 aufeinanderfolgenden Fertigungsintervallen.....	6
1.1.5	Zahnpasta-Marktanteil von "CREST" in 276 aufeinanderfolgenden Wochen mit zwischenzeitlicher Werbekampagne.....	7
1.2	Begriffsklärung: Empirische und theoretische Zeitreihe	7
1.3	Beispiele für theoretische Zeitreihen	8
1.4	Ziele und Phasen der Zeitreihenanalyse	9
1.5	Voraussetzung vieler zeitreihenanalytischer Methoden: Stationarität.....	10
1.6	Wichtige Hilfsmittel für die Analyse stationärer Zeitreihen.....	10
1.6.1	Die (partielle) Autokorrelationsfunktion (P)ACF	10
1.6.2	Der Box-Ljung-Test.....	11
1.6.3	Beispiel: ACF, PACF und Box-Ljung-Test bei der Modellprüfung	11
2	KURZE CHARAKTERISIERUNG VON SPSS-TRENDS.....	13
2.1	Verfügbarkeit an der Universität Trier	13
2.2	Trends-Optionen zur Modellidentifikation	13
2.2.1	Plotten der Zeitreihe in horizontaler bzw. vertikaler Richtung	13
2.2.2	Plotten der (partiellen) Autokorrelationsfunktion einer Zeitreihe.....	13
2.2.3	Plotten der Kreuzkorrelationsfunktion zweier Zeitreihen	14
2.3	Die von Trends unterstützten zeitreihenanalytischen Methoden	14
2.3.1	Exponentielles Glätten	14
2.3.2	ARIMA-Modellierung nach Box & Jenkins (1976)	14
2.3.3	Verfahren zur Saisonbereinigung.....	14
2.3.4	Spektralanalytische Verfahren	14
2.3.5	Regressionsanalytische Methoden	15
2.4	Trends-Optionen zur Modelldiagnose	16
2.5	Allgemeine Hinweise zur Analyse von Zeitreihendaten mit SPSS.....	16
2.5.1	Beispiel: Prognose von Verkaufszahlen per exponentieller Glättung.....	16
2.5.2	SPSS-Variablen bei der Zeitreihenanalyse.....	16
2.5.2.1	Datumsvariablen	16
2.5.2.2	Automatische Erzeugung neuer Variablen (Residuen, Prognosen etc.)	17
2.5.3	Schätz-, Validierungs- und Prognosebereich	18
2.5.3.1	Festlegung des Schätzbereichs	18
2.5.3.2	Festlegung des Validierungs- und/oder des Prognosebereichs.....	19
2.5.3.3	Aktuelle Einstellungen für Schätz-, Validierungs- und Prognosebereich überprüfen	19
2.5.4	Zeitreihenspezifische Transformationen	19
2.5.4.1	Erstellung neuer Zeitreihen.....	19
2.5.4.2	Ersetzung fehlender Werte	20

3	EXPONENTIELLES GLÄTTEN.....	22
3.1	Grundprinzipien des Verfahrens.....	22
3.2	Anwendung auf die SALES-Reihe.....	23
3.2.1	Modellspezifikation und Schätzergebnisse	23
3.2.2	Modellbeurteilung	26
4	ARIMA-MODELLE	29
4.1	Wiederholung: Grundprinzipien der ARIMA-Modellierung.....	29
4.1.1	Der autoregressive Prozeß.....	29
4.1.2	Der Moving Average - Prozeß	29
4.1.3	Der ARMA-Prozeß	29
4.1.4	Der ARIMA-Prozeß	29
4.1.5	Zur Modell-Identifikation	29
4.1.6	Zur Modell-Beurteilung	32
4.2	Beispiel: Der Einsatz von ARIMA-Modellen in der Qualitätskontrolle	32
4.2.1	Einlesen der Zeitreihe und andere Vorbereitungen.....	32
4.2.2	Plot der Zeitreihe zur Kontrolle auf Stationarität.....	33
4.2.3	Untersuchung der ACF und der PACF	34
4.2.4	Modellschätzung	34
4.2.5	Modellbeurteilung	36
4.2.6	Erprobung des ARIMA-Modells als Störungsdetektor.....	37
5	LITERATUR.....	41
6	ANHANG	42
6.1	Trends-Systemparameter.....	42
7	STICHWORTVERZEICHNIS	43

Herausgeber: Universitäts-Rechenzentrum Trier
 Universitätsring 15
 D-54286 Trier
 Tel.: (0651) 201-3417, Fax.: (0651) 3921

Leiter: Prof. Dr.-Ing. Manfred Paul
 Autor: Bernhard Baltes-Götz
 Mail: baltes@uni-trier.de

Copyright © 1997; URT

Vorwort

Dieser Kurs ist gedacht für Teilnehmer mit einer zeitreihenanalytischen Grundausbildung, die ihre Kenntnisse mit dem Zeitreihenmodul **Trends** aus dem SPSS-System anwenden wollen. Im Einzelnen werden folgende Themen behandelt:

- Kurze Wiederholung zeitreihenanalytischer Grundbegriffe
- Darstellung der Analysemöglichkeiten von Trends
- Beschreibung allgemeiner Bedienungsmerkmale von Trends
- Beispiele zur Anwendung wichtiger Standardmethoden mit Trends

Als Software kommt SPSS 6.1 für Windows zum Einsatz, jedoch können praktisch alle vorgestellten Verfahren auch mit jüngeren SPSS-Versionen unter Windows, MacOS oder Linux realisiert werden.

Das Manuskript ist als PDF-Dokument zusammen mit den im Kurs benutzen Dateien auf dem Webserver der Universität Trier von der Startseite (<http://www.uni-trier.de/>) ausgehend folgendermaßen zu finden:

[Rechenzentrum](#) > [Studierende](#) > [EDV-Dokumentationen](#) >
[Statistik](#) > [Zeitreihenanalyse mit SPSS-Trends](#)

Hinweise auf Unzulänglichkeiten im Manuskript werden mit Dank entgegen genommen

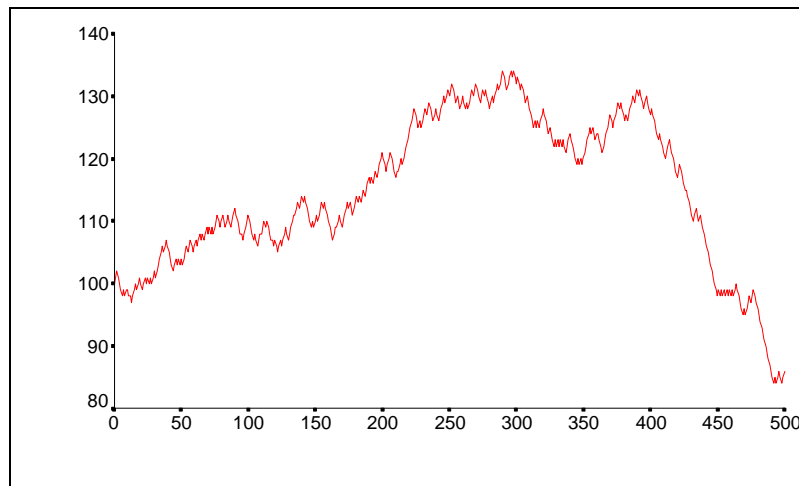
1 Kurze Charakterisierung der Zeitreihenanalyse

1.1 Beispiele für empirische Zeitreihen und Auswertungsfragestellungen

In den ersten drei Beispielen sehen Sie Zeitreihen in "eigendynamischer Fluktuation". Die Zeitreihenanalyse wird jeweils bestrebt sein, das Bildungsgesetz herauszufinden. Demgegenüber scheinen die beiden restlichen Zeitreihen Strukturbrüche aufzuweisen. Hier kann die Zeitreihenanalyse z.B. versuchen, das Bildungsgesetz aus einer Phase eigendynamischer Fluktuation zu gewinnen, um die exogenen Effekte von dem normalen stochastischen Geschehen separieren zu können.

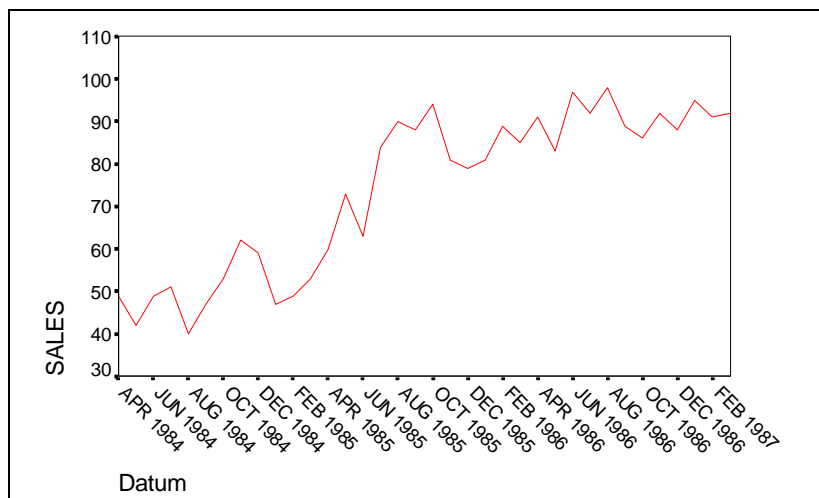
1.1.1 Guthaben im Nullsummenroulett bei strenger Präferenz für Rot

Im Vergleich zur rauhen Wirklichkeit verzichtet die Bank beim Nullsummenroulett auf das Einziehen des Kapitals bei Zero. Unser Spieler hat ein Startkapital von 100 Chips und setzt stets einen Chip auf Rot. Folglich kann sein Guthaben bei intaktem Spielgerät in jeder Runde entweder um einen Chip ansteigen (mit Wahrscheinlichkeit 18/37) oder um einen Chip abnehmen (ebenfalls mit Wahrscheinlichkeit 18/37) oder gleich bleiben (mit Wahrscheinlichkeit 1/37). Der Kontostand vollführt dabei eine sogenannte Irrfahrt (vgl. Abschnitt 1.3).



Fragestellung: Anhand einer empirischen Zeitreihe kann für ein konkretes Spielgerät geprüft werden, ob es tatsächlich Irrfahrten produziert.

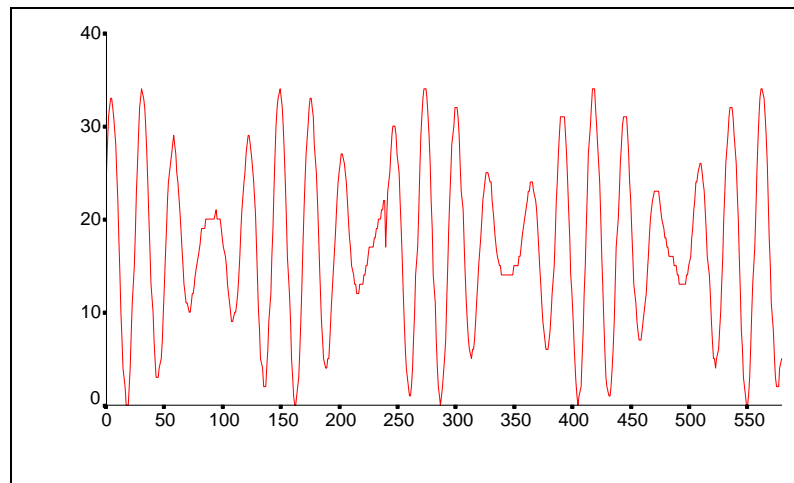
1.1.2 Monatliche Verkaufszahlen für ein Produkt¹



¹ Das Beispiel stammt aus dem SPSS-Trends-Handbuch (1988, A6ff)

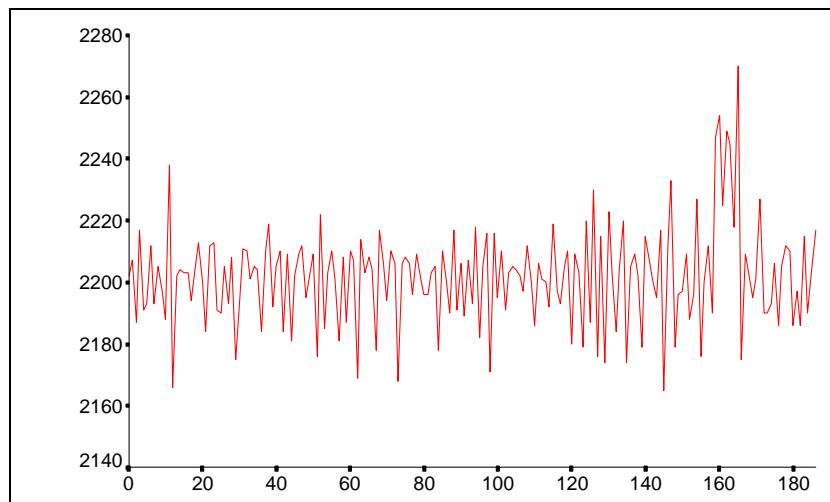
Fragestellung: Zur Planung der Produktion muß der weitere Verlauf der Verkaufszahlen möglichst präzise vorhergesagt werden.

1.1.3 Lichtstärke eines pulsierenden Sterns in 580 aufeinanderfolgenden Nächten



Fragestellung: Läßt sich die relativ "harmonische" Bewegung der Zeitreihe um eine Ruhelage herum als Kombination einiger weniger kosinusförmiger Schwingungskomponenten erklären? Dies ist aufgrund des physikalischen Hintergrunds zu erwarten. Welche Frequenzen haben die beteiligten Schwingungskomponenten?

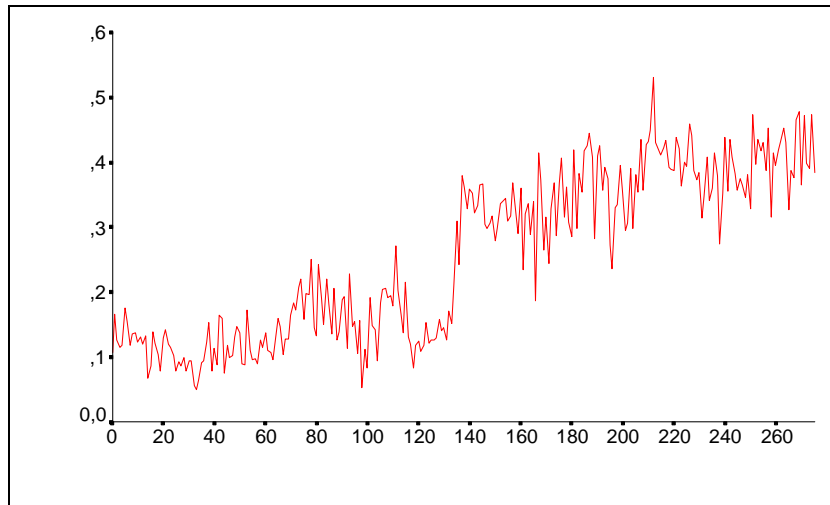
1.1.4 Anzahl defekter Produkte in 187 aufeinanderfolgenden Fertigungsintervallen¹



Fragestellung: Sind die auffälligen Spitzen in den Intervallen von 160 bis 170 als normales Fluktuationsverhalten zu erklären, oder sehen wir hier einen Defekt der Produktionsanlage (mit anschließender erfolgreicher Reparatur)? Ein Fertigungsingenieur muß schnell feststellen, ob die Anlage wegen eines Defekts angehalten werden muß. Er darf aber keinen Fehlalarm auslösen, weil ein Produktionsstopp ebenfalls Kosten verursacht.

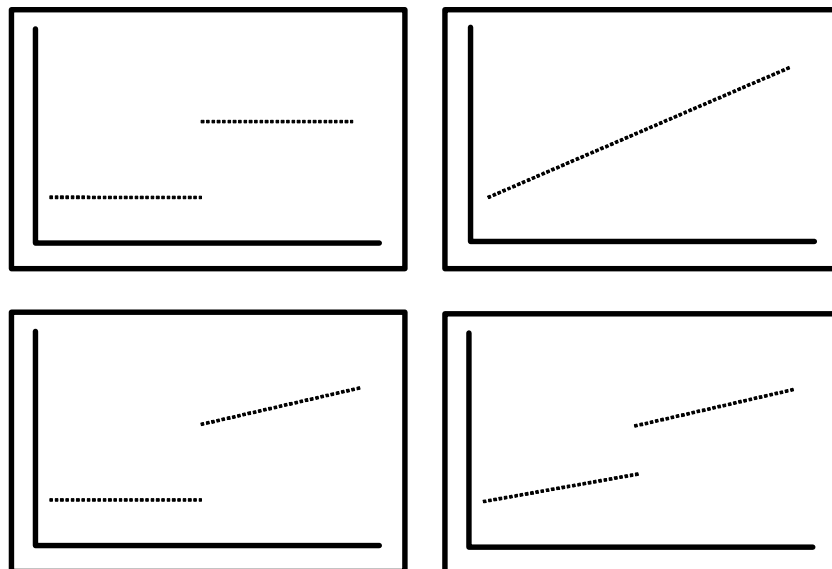
¹ Das Beispiel stammt aus dem SPSS-Trends-Handbuch (1994, S. 57ff)

1.1.5 Zahnpasta-Marktanteil von "CREST" in 276 aufeinanderfolgenden Wochen mit zwischenzeitlicher Werbekampagne¹



Fragestellung:

Wurde der Marktanteil durch die Werbekampagne (in der 135. und 136. Woche) signifikant im Sinne einer Treppenfunktion erhöht, liegt ein genereller linearer Trend vor oder ist ein kombiniertes Modell anzunehmen? Man wird also versuchen, zwischen den folgenden konkurrierenden Modellen für die Zeitreihe zu entscheiden:



1.2 Begriffsklärung: Empirische und theoretische Zeitreihe

Eine endliche Folge $x_1, x_2, x_3, \dots, x_N$ zeitlich äquidistanter Beobachtungen eines Merkmals bei einer Beobachtungseinheit wird als **empirische Zeitreihe** bezeichnet. Bei einer über reine Deskription hinausgehenden statistischen Analyse wird die empirische Zeitreihe als *Teil einer* zufälligen Realisation einer **theoretischen Zeitreihe**, bestehend aus einer Folge $(X_t)_{t \in T}$ von Zufallsvariablen, aufgefaßt. Dabei ist T eine Indexmenge mit den Beobachtungszeitpunkten. Verwendet man dazu z.B. die Menge der natürlichen Zahlen (eventuell um die Null erweitert), so hat die theoretische Zeitreihe folgende Gestalt:

$$X_1, X_2, X_3, \dots \quad \text{bzw.} \quad X_0, X_1, X_2, \dots$$

Nimmt man die Menge der ganzen Zahlen, dann sieht die theoretische Zeitreihe folgendermaßen aus:

$$\dots, X_{-2}, X_{-1}, X_0, X_1, X_2, \dots$$

¹ Das Beispiel stammt aus dem SPSS-Trends-Handbuch (1994, S. 137ff)

Wir gehen in diesem Manuskript stets davon aus, daß entweder die natürlichen oder die ganzen Zahlen als Indexmenge dienen.

Wesentlich an der statistischen Betrachtungsweise: Die theoretische Zeitreihe gehört zu einem **Zufallsexperiment**. Realisiert sich ein Ergebnis ω dieses Zufallsexperimentes, dann nimmt jede Zufallsvariable X_t den Wert $X_t(\omega)$ an, und wir erhalten eine Folge von Zahlen als Realisation der theoretischen Zeitreihe.

Im Roulett-Beispiel aus Abschnitt 1.1.1 haben wir ...

- als theoretische Zeitreihe: den Guthabensverlauf während eines prinzipiell unbegrenzt langen Roulettspiels unter den in Abschnitt 1.1.1 angegebenen Bedingungen.
Diese theoretische Zeitreihe ist zwar ein recht abstraktes Objekt, und doch lassen sich viele Aussagen darüber machen. Man kann z.B. leicht herleiten, daß die Varianz ihrer Zufallsvariablen mit dem Zeitindex ansteigt. Zudem ist das zugehörige Zufallsexperiment leicht durchzuführen. Mit Hilfe der Computersimulation ist es sogar möglich, zahlreiche unabhängige Realisationen der Zeitreihe zu erzeugen.
- als empirische Zeitreihe: den Guthabensverlauf aus den ersten 501 Spielen einer konkreten Realisation
Die Beschränkung auf ein *endliches Teilstück* der betrachteten Realisation ist aus praktischen Gründen unumgänglich. Unter bestimmten Voraussetzungen ist eine statistische Analyse trotzdem möglich (siehe Abschnitt 1.5 zum Begriff der Stationarität). Wir untersuchen nämlich in der Regel Bildungsgesetze einer Zeitreihe, die sich in allen geeignet langen Realisationsteilstücken artikulieren sollten.

Das Generieren von empirischen Zeitreihen ist nicht immer so leicht möglich. Im Beispiel aus Abschnitt 1.1.5 (Zahnpasta-Marktanteil von Crest) muß zur Durchführung des zugehörigen Zufallsexperiments eine neue Zahnpastasorte auf einem Markt eingeführt und außerdem eine Werbekampagne durchgeführt werden. Es wäre ein aufwendiges, wenn nicht gar unmögliches Unterfangen, zur Gewinnung einer zweiten Realisation der theoretischen Zeitreihe dieselbe Prozedur auf einem anderen, vergleichbaren und unabhängigen Markt zu wiederholen.

Bei einer über reine Deskription hinausgehenden statistischen Analyse benutzen wir eine empirische Zeitreihe dazu, ein mathematisches Modell für die zugehörige theoretische Zeitreihe zu ermitteln bzw. zu prüfen. In unserem Roulett-Beispiel ist die theoretische Zeitreihe aufgrund der in Abschnitt 1.1.1 gegebenen Beschreibung eine sogenannte **Irrfahrt**. (Eine formale Definition dieses Begriffs folgt im nächsten Abschnitt.) Anhand einer empirischen Zeitreihe kann für ein konkretes Spielgerät geprüft werden, ob es tatsächlich Irrfahrten produziert.

1.3 Beispiele für theoretische Zeitreihen

i) Weißes Rauschen

Eine theoretische Zeitreihe $(E_t)_{t \in T}$ wird als **weißes Rauschen** bezeichnet, wenn gilt:

- Alle Zufallsvariablen haben den Erwartungswert Null und dieselbe Varianz σ^2 :
$$\mathbf{E}(E_t) = 0, \text{Var}(E_t) =: \sigma^2 < \infty \quad \forall t \in T$$
- Die Kovarianz zwischen den Zufallsvariablen zu zwei verschiedenen Zeitpunkten ist Null:
$$\text{Kov}(E_t, E_s) = 0 \quad \forall t, s \in T, \text{ mit } t \neq s$$

Im Roulett-Beispiel in Abschnitt 1.1.1 wird ein weißes Rauschen zur Konstruktion der Kontostands-Irrfahrt verwendet. Dort realisiert sich nämlich in jedem Durchgang eine unabhängige Zufallsvariable E_t , $t = 1, 2, \dots$ mit der folgenden, von t unabhängigen, Verteilung:

$$P(-1) = 18/37, P(0) = 1/37, P(1) = 18/37$$

Damit haben wir für alle E_t den konstanten Mittelwert Null und dieselbe Varianz. Wegen der Unabhängigkeit verschwinden alle Kovarianzen.

ii) Irrfahrt (engl.: Random Walk)

Ist $(E_t)_{t \in \mathbb{N}_0}$ ein weißes Rauschen (mit " \mathbb{N}_0 " als Symbol für die Menge der natürlichen Zahlen inklusive Null), dann wird die im folgenden induktiv definierte theoretische Zeitreihe als **Irrfahrt** (engl.: **Random Walk**) bezeichnet:

$$X_t := \begin{cases} s + E_0 & \text{für } t = 0, s \in \mathbb{R} \\ X_{t-1} + E_t & \text{für } t = 1, 2, 3, \dots \end{cases}$$

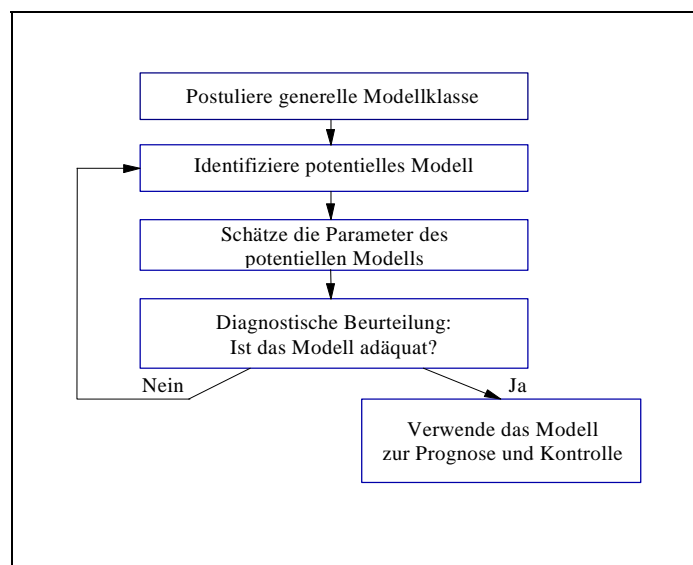
Ein Random Walk entscheidet sich also in jedem Zeittakt per Zufall für eine Weiterbewegung nach oben oder unten. Durch die Konstante s in der Definitionsvorschrift für X_0 können wir einen beliebigen Startwert festlegen (vgl. Beispiel iv in Abschnitt 1.1).

1.4 Ziele und Phasen der Zeitreihenanalyse

Ziele der Zeitreihenanalyse:

- Deskription
- Erklärung (Modellbildung)
Welcher Prozeß hat die empirische Zeitreihe mutmaßlich generiert?
- Prognose (des weiteren Zeitreihenverlaufs)
- Kontrolle (des weiteren Zeitreihenverlaufs)
- Evaluation (Interventionsanalyse)

Die zentrale Aufgabe der Zeitreihenanalyse ist die Modellbildung. Sie sollte nach Box & Jenkins (1976, S. 19) in folgenden Schritten ablaufen:



Wenn etwa bei einer Zeitreihenanalyse die generelle Klasse der $ARIMA(p,d,q)$ -Modelle zugrundegelegt wird (siehe Abschnitt 4.1), dann sind in den weiteren Modellierungsschritten folgende Fragen zu klären:

- Identifikation eines potentiellen Modells
Welcher Differenzierungsgrad d ist nötig? Welche Ordnungen p bzw. q sollen für das AR- bzw. für MA-Polynom angenommen werden?
- Parameterschätzung
In dieser Phase werden mit EDV-Hilfe die Parameter des vorläufig akzeptierten Modells anhand der empirischen Zeitreihe geschätzt.

- Modellbeurteilung

Bei der Beurteilung der Modellgültigkeit werden vor allem die Modellresiduen (geschätzte Werte der Störungen) untersucht. Die Residuen eines gut passenden Modells verhalten sich wie ein weißes Rauschen. Die Beurteilung der einzelnen Parameterschätzungen kann z.B. anhand ihrer Standardfehler erfolgen.

1.5 Voraussetzung vieler zeitreihenanalytischer Methoden: Stationarität

Eine theoretische Zeitreihe $(X_t)_{t \in T}$ heißt (**schwach**) **stationär**, wenn gilt:

- Alle Zufallsvariablen haben denselben Erwartungswert:

$$E(X_t) = \mu \quad \forall t \in T$$

- Die Kovarianz zwischen zwei Zufallsvariablen hängt ausschließlich von der zeitlichen Distanz ab:

$$\text{Kov}(X_t, X_u) = \text{Kov}(X_{t,r}, X_{u,r}) \quad \forall t, u, r \in T$$

Weil die Varianz einer Variablen gleich ihrer Kovarianz mit sich selbst ist (Distanz Null), sind bei einer stationären Zeitreihe also insbesondere auch alle Varianzen gleich.

Bemerkungen:

- Die Restriktion zeitlicher Invarianz wesentlicher stochastischer Merkmale der Zeitreihe ist für eine statistische Analyse aufgrund einer einzigen (!) Realisation des Prozesses unabdingbar, zumal in der Praxis nicht die gesamte Realisation vorliegt sondern nur ein Ausschnitt der Länge N (siehe Abschnitt 1.2).
- Zahlreiche nicht-stationäre Zeitreihen lassen sich in stationäre Zeitreihen transformieren (z.B. durch Trend-Elimination, Differenzenbildung).

Übungen:

- Untersuchen Sie die in Abschnitt 1.3 definierten theoretischen Zeitreihen (weißes Rauschen und Irrfahrt (Random Walk)) auf Stationarität?
- Zu einer Zeitreihe $(X_t)_{t \in T}$ wird die Zeitreihe $(D_t)_{t \in T}$ mit den ersten Differenzen folgendermaßen definiert:

$$D_t := X_t - X_{t-1}$$

Zeigen Sie, daß die Zeitreihe mit den ersten Differenzen zu einer Irrfahrt stationär ist.

1.6 Wichtige Hilfsmittel für die Analyse stationärer Zeitreihen

1.6.1 Die (partielle) Autokorrelationsfunktion (P)ACF

Nur für stationäre Zeitreihen ist die **Autokorrelationsfunktion** (engl. Abk.: ACF) definiert, die für jede natürliche Zahl τ die Korrelation zweier um τ Takte voneinander entfernter Variablen X_t und $X_{t+\tau}$ angibt. Analoges gilt für die **partielle Autokorrelationsfunktion** (engl. Abk.: PACF), die für jede natürliche Zahl τ die partielle Korrelation zweier um τ Takte voneinander entfernter Variablen X_t und $X_{t+\tau}$ bei Kontrolle der dazwischenliegenden Zufallsvariablen angibt. ACF und PACF spielen im Rahmen der ARIMA-Modellierung (vgl. Abschnitt 4) eine wichtige Rolle bei der Modelidentifikation. Außerdem liefern sie Informationen zur Modellbeurteilung (siehe nächsten Abschnitt).

1.6.2 Der Box-Ljung-Test

Mit dem **Box-Ljung-Test** kann die Nullhypothese geprüft werden, daß die ersten m Autokorrelationen einer Zeitreihe gleich Null sind. Seine Prüfgröße Q_{BL} ist folgendermaßen definiert ($N =$ Stichprobengröße, $\hat{\rho}_\tau =$ geschätzte Autokorrelation zur Distanz τ):

$$Q_{BL} := N(N + 2) \sum_{\tau=1}^m \frac{\hat{\rho}_\tau^2}{N - \tau}$$

Sie ist bei Gültigkeit der Nullhypothese asymptotisch (d.h. für $N \longrightarrow \infty$) χ_m^2 -verteilt, so daß die Nullhypothese zumindest für hinreichend große empirische Zeitreihen ($N \geq 100$) getestet werden kann (vgl. Schlittgen & Streitberg 1989, S. 243f; Hartung 1989, S. 690f).

Die Residuen zu den häufig eingesetzten ARIMA-Modellen (vgl. Abschnitt 4) müssen bei Modellgültigkeit ein weißes Rauschen sein, und bei dieser Zeitreihe sind alle Autokorrelationen gleich Null. Daher wird der Box-Ljung-Test in der Zeitreihenanalyse häufig zur Prüfung der Gültigkeit von ARIMA-Modellen eingesetzt. Bei Berechnung aus der Residualreihe zu einem gültigen Modell mit $(p+q)$ Parametern folgt die Prüfgröße Q_{BL} allerdings einer asymptotisch einer χ_{m-p-q}^2 -Verteilung, falls $m > (p+q)$.

Ein Modell zu einer empirischen Zeitreihe sollte nur dann akzeptiert werden, wenn der Box-Ljung-Test zur Residualzeitreihe für jedes m die zugehörige Nullhypothese beibehält.

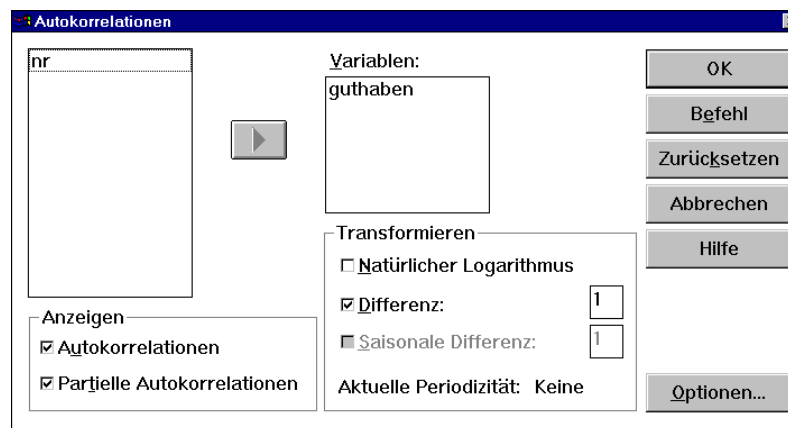
1.6.3 Beispiel: ACF, PACF und Box-Ljung-Test bei der Modellprüfung

Am Beispiel der Nullsummenroulett-Zeitreihe (vgl. Abschnitt 1.1.1) soll der Einsatz der oben beschriebenen Hilfsmittel zur Modellprüfung mit SPSS demonstriert werden. Eine empirische Realisation dieser Zeitreihe finden Sie in der SPSS-Datendatei **ROULETT.SAV** an der im Vorwort vereinbarten Stelle. Wenn es sich wirklich um eine Irrfahrt handelt, ist die Zeitreihe der ersten Differenzen ein weißes Rauschen. Bei dieser theoretischen Zeitreihe verschwinden alle (partiellen) Autokorrelationen. Somit kann die Irrfahrts - Hypothese durch Untersuchung der (P)ACF zu der empirischen Differenzenzeitreihe überprüft werden.

Um die beiden Funktionen anzufordern, rufen wir mit dem Menübefehl

Grafik, Zeitreihen, Autokorrelationen...

die folgende Dialogbox auf:



Hier wählen wir die Variable GUTHABEN und markieren im Bereich **Transformieren** das Kontrollkästchen **Differenz**, ohne den voreingestellten Differenzierungsgrad Eins zu verändern. Nach einem **OK** erhalten wir im Ausgabefenster die folgenden Ergebnisse:

Autocorrelations: GUTHABEN

Transformations: difference (1)

Lag	Auto-Corr.	Stand. Err.	-1	-.75	-.5	-.25	0	.25	.5	.75	1	Box-Ljung	Prob.
1	,046	,045					.		*			1,067	,302
2	,046	,045					.		*			2,131	,345
3	,013	,044					.		*			2,221	,528
4	,081	,044					.		**			5,533	,237
5	-,027	,044					.		*			5,912	,315
6	,003	,044					.		*			5,918	,432
7	,034	,044					.		*			6,503	,482
8	,013	,044					.		*			6,596	,581
9	,038	,044					.		*			7,334	,602
10	-,056	,044					.		*			8,935	,538
11	,024	,044					.		*			9,227	,601
12	,012	,044					.		*			9,297	,677
13	,006	,044					.		*			9,313	,749
14	,057	,044					.		*			10,974	,688
15	,102	,044					.		**			16,321	,361
16	,044	,044					.		*			17,345	,364

Plot Symbols: Autocorrelations * Two Standard Error Limits .

Total cases: 501 Computable first lags after differencing: 499

Partial Autocorrelations: GUTHABEN

Transformations: difference (1)

Lag	Pr-Aut-Corr.	Stand. Err.	-1	-.75	-.5	-.25	0	.25	.5	.75	1
1	,046	,045					.		*		
2	,044	,045					.		*		
3	,009	,045					.		*		
4	,078	,045					.		**		
5	-,036	,045					.		*		
6	-,001	,045					.		*		
7	,035	,045					.		*		
8	,005	,045					.		*		
9	,040	,045					.		*		
10	-,063	,045					.		*		
11	,021	,045					.		*		
12	,015	,045					.		*		
13	-,002	,045					.		*		
14	,068	,045					.		*		
15	,089	,045					.		**		
16	,029	,045					.		*		

Plot Symbols: Autocorrelations * Two Standard Error Limits .

Total cases: 501 Computable first lags after differencing: 499

Erwartungsgemäß liegen alle (partiellen) Autokorrelationen nahe bei Null. Im Box-Ljung-Test wird die Nullhypothese für alle Lags (Differenzen) beibehalten. Wir gehen also davon aus, daß die Zeitreihe der ersten Differenzen ein weißes Rauschen und infolgedessen die undifferenzierte Zeitreihe eine Irrfahrt.

Im weiteren Verlauf des Kurses werden wir noch mehrfach (P)ACF-Funktionen betrachten.

2 Kurze Charakterisierung von SPSS-Trends

Trends ist ein SPSS-Zusatzmodul, das zahlreiche Methoden der modernen Zeitreihenanalyse bereitstellt. Die zugehörigen Prozeduren können alternativ per Dialogbox oder über Kommandos aufgerufen werden.

2.1 Verfügbarkeit an der Universität Trier

Trends steht an der Universität Trier im Rahmen folgender SPSS-Versionen zur Verfügung:

- SPSS für Windows
- SPSS für Macintosh
- SPSS für UNIX (Rechner: SUN-SPARC Station 1000 unter SOLARIS)

In diesem Manuskript wird zwar in erster Linie die Windows-Version behandelt, doch sind die meisten Angaben relativ leicht übertragbar

2.2 Trends-Optionen zur Modellidentifikation

Trends kann mehrere Plots erzeugen, die in Textdrucker- und in hochauflöser Qualität verfügbar sind:

2.2.1 Plotten der Zeitreihe in horizontaler bzw. vertikaler Richtung

Menü: **Grafik, Sequenz..., Format: Zeit auf horizontaler Achse = Ja bzw. Nein**

Syntax: **TSPLLOT** bzw. **CASEPLOT**

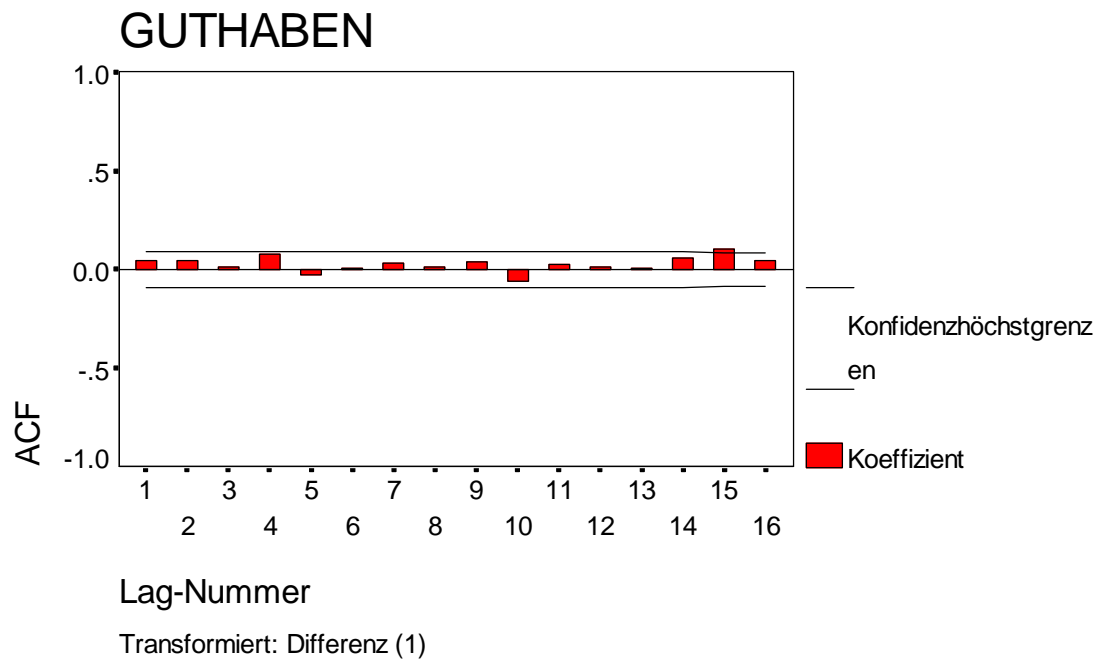
Die hochauflösten Zeitreihenplots in Abschnitt 1.1 wurden alle mit SPSS für Windows erstellt.

2.2.2 Plotten der (partiellen) Autokorrelationsfunktion einer Zeitreihe

Menü: **Grafik, Zeitreihen, Autokorrelationen...**

Syntax: **ACF** bzw. **PACF**

In Abschnitt 1.6 haben Sie bereits (P)ACF-Plots in Textdrucker-Qualität gesehen. Als hochauflöste Grafik sieht die ACF der einmal differenzierten ROULETT-Reihe so aus:



2.2.3 Plotten der Kreuzkorrelationsfunktion zweier Zeitreihen

Menü: **Grafik, Zeitreihen, Kreuzkorrelationen...**

Syntax: **CCF**

Wenn zeitversetzte Effekte einer unabhängigen Variable auf eine abhängige Variable vermutet werden, kann die Inspektion der Kreuzkorrelationsfunktion nützlich sein. Ihr Wert für die natürliche Zahl τ ist die Korrelationen der ersten Zeitreihe mit der um τ Zeiteinheiten verschobenen zweiten Zeitreihe.

2.3 Die von Trends unterstützten zeitreihenanalytischen Methoden

Trends kann Modelle aus verschiedenen Familien zeitreihenanalytischer Methoden schätzen.

2.3.1 Exponentielles Glätten

Menü: **Statistik, Zeitreihen, Exponentielles Glätten...**

Syntax: **EXSMOOTH**

Die Methode des exponentiellen Glättens ist ein relativ einfaches Verfahren, das die Erstellung von Einschnitt-Prognosen (von X_{t+1} aufgrund von X_1, \dots, X_t) erlaubt, wobei kein wesentliches theoretisches Verständnis der Zeitreihe nötig ist. Eine kurze Beschreibung der Methode und ein Anwendungsbeispiel folgen in Abschnitt 0.

2.3.2 ARIMA-Modellierung nach Box & Jenkins (1976)

Menü: **Statistik, Zeitreihen, ARIMA...**

Syntax: **ARIMA**

Das ARIMA-Prozedur von Trends ermöglicht univariate saisonale und nicht-saisonale ARIMA-Modelle nach Box & Jenkins (1976). Der Schätzalgorithmus arbeitet nach dem Maximum-Likelihood - Prinzip und toleriert fehlende Daten. Eine Beschreibung des sehr häufig angewendeten ARIMA-Ansatzes folgt in Abschnitt 4 im Zusammenhang mit einem Anwendungsbeispiel.

Weil die ARIMA-Prozedur über den klassischen ARIMA-Ansatz hinausgehend auch Regressoren einbeziehen kann, erlaubt sie auch eine multiple Regressionsanalyse mit ARIMA-Modell für die Fehler (vgl. Abschnitt 2.3.5).

2.3.3 Verfahren zur Saisonbereinigung

Menü: **Statistik, Zeitreihen, X11 ARIMA...** bzw.

Statistik, Zeitreihen, Saisonale Zerlegung...

Syntax: **X11ARIMA** bzw. **SEASON**

Bei diesem traditionellen Ansatz der Zeitreihenanalyse werden saisonale Muster aus der Zeitreihe eliminiert, um die übrigen Komponenten besser analysieren zu können. Trends bietet die ältere Methode Census I (Kommando SEASON) und die aktuellere Methode Census II X11 (Kommando X11ARIMA) an.

Die Verfahren zur Saisonbereinigung werden in diesem Kurs nicht behandelt.

2.3.4 Spektralanalytische Verfahren

Menü: **Grafik, Zeitreihen, Spektralanalyse...**

Syntax: **SPECTRA**

Spektralanalytische Methoden eignen sich vor allem für Zeitreihen, die mutmaßlich durch Überlagerung kosinusförmiger Schwingungskomponenten mit verschiedenen Frequenzen entstanden sind.

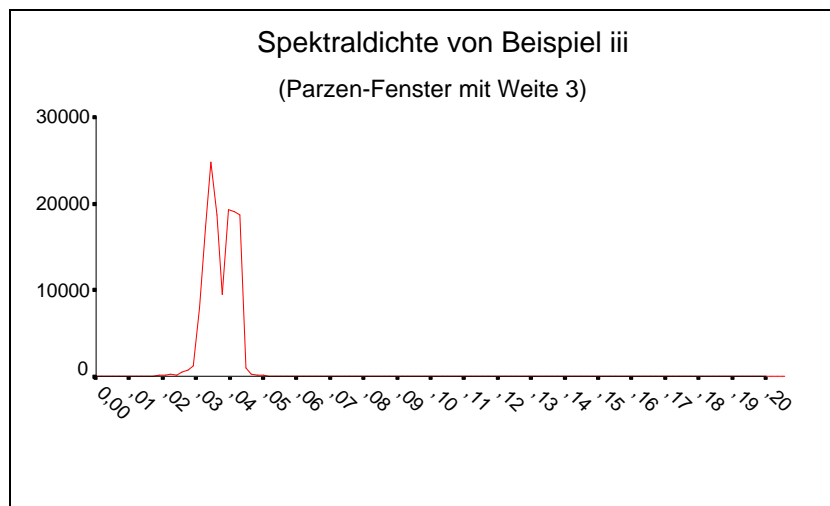
Die in Abschnitt 1.1.3 vorgestellten Lichtstärke-Daten eines pulsierenden Sterns lassen sich z.B. fast perfekt als Überlagerung zweier harmonischer Schwingungen mit einer 24- bzw. 29-tägigen Periode darstellen:

$$\hat{x}_t = 17,10 + 7,65 \cos\left(2\pi \frac{1}{29} t\right) - 6,50 \sin\left(2\pi \frac{1}{29} t\right) - 7,08 \sin\left(2\pi \frac{1}{24} t\right), t = 1, 2, \dots$$

In einem verallgemeinerten Sinn läßt sich *jede* stationäre Zeitreihe als "Summe" von kosinusförmigen Funktionen darstellen, wobei aber i.a. *alle* Frequenzen $0 \leq \lambda < 0,5$ einbezogen werden müssen. Die größte "entdeckbare" Frequenz 0,5 entspricht dabei einer Schwingung mit einer Periodenlänge von zwei Zeittakten. Bei täglich erhobenen Daten können also z.B. nur Schwingungskomponenten mit einer Periodenlänge von mindestens zwei Tagen identifiziert werden.

Das Periodogramm bzw. die durch Glättung daraus gewonnene Schätzung der Spektraldichte vermitteln einen Eindruck davon, in welcher Stärke einzelne Schwingungskomponenten (Frequenzen) in einer stationären Zeitreihe enthalten sind. Trends erstellt Plots für uni- und bivariate Spektralanalysen.

Da die durchaus interessante spektralanalytische Methodologie im weiteren Kursverlauf nicht mehr aufgegriffen wird, soll an dieser Stelle die Spektraldichte zu unserem Beispiel iii angegeben werden:



Es finden sich erwartungskonform (vgl. obige Formel) zwei ausgeprägte Spitzen bei den Frequenzen $1/29$ ($\approx 0,0345$) und $1/24$ ($\approx 0,0417$).

2.3.5 Regressionsanalytische Methoden

Trends bietet einige regressionsanalytische Verfahren, welche die Besonderheiten von autokorrelierten Zeitreihendaten berücksichtigen:

- Multiple Regression mit AR(1)- bzw. ARIMA(p,d,q)-Modell für die Residuen

Menü: **Statistik, Zeitreihen, Autoregression...** bzw.
Statistik, Zeitreihen, ARIMA...

Syntax: **AREG, ARIMA**

Häufig soll eine Kriteriums-Zeitreihe durch Regressor-Zeitreihen erklärt werden. Die übliche Regressionsanalyse ist i.a. nicht anwendbar, weil die Modellresiduen eine serielle Abhängigkeit aufweisen. Oft kann diese serielle Abhängigkeit durch einen autoregressiven Prozeß erster Ordnung (AR(1)-Prozeß, siehe Abschnitt 4.1) modelliert werden. In solchen Fällen ist mit AREG durch explizite Berücksichtigung der AR(1)-Komponente eine unverfälschte inferenzstatistische Analyse möglich. Wird für die Residuen ein allgemeiner ARIMA-Prozeß unterstellt, dann ist die gleichnamige Prozedur zu verwenden.

- Kurvenanpassung

Menü: **Statistik, Regression, Kurvenanpassung...**

Syntax: **CURVEFIT**

Mit dieser Prozedur aus dem SPSS-Basissystem kann man einer Zeitreihe verschiedene Kurven als Funktionen der Zeit anpassen. Einziger Regressor ist hier also die Zeit.

2.4 Trends-Optionen zur Modelldiagnose

Zur Prüfung der Modellgültigkeit stehen u.a. folgende Methoden für die Untersuchung der Residuen zur Verfügung:

- Plots der Residual-(P)ACF und Test der Nullhypothese, daß die ersten m Autokorrelationen einer Zeitreihe gleich Null sind. Hierzu kann mit dem Menübefehl **Grafik, Zeitreihen, Autokorrelationen...** oder mit dem Kommando **ACF** die empirische Autokorrelationsfunktion berechnet, geplottet und mit dem Box-Ljung-Test beurteilt werden (vgl. Abschnitt 1.6).
- Grafische Normalverteilungsanalyse mit dem Menübefehl **Grafik, Normalverteilung....** oder mit dem Kommando **NPLOT**.

Die Analyse der Residuen mit dem **FIT**-Kommando aus dem SPSS-Basissystem gibt Aufschluß über die Erklärungsleistung eines Modells und erlaubt den Vergleich verschiedener Modelle.

Einzelne Parameterschätzungen können z.B. anhand ihrer Standardfehler beurteilt werden.

2.5 Allgemeine Hinweise zur Analyse von Zeitreihendaten mit SPSS

2.5.1 Beispiel: Prognose von Verkaufszahlen per exponentieller Glättung

Im Verlauf von Abschnitt 2.5 wird zur Demonstration elementarer Trends-Eigenschaften die Zeitreihe SALES mit Verkaufszahlen für ein Produkt analysiert. Dabei wird mit der Methode der exponentiellen Glättung (siehe Abschnitt 3) ein Prognose-Modell entwickelt, das in der ökonomischen Praxis etwa dazu benutzt werden kann, eine Produktion an den zu erwartenden Absatz anzupassen. Die Bereitstellung der Daten kann durch Öffnen der SPSS – Datendatei **SALES.SAV** (zu finden an der im Vorwort vereinbarten Stelle) oder durch Ausführung der folgenden SPSS-Kommandos geschehen:

```
data list free /sales.
begin data.
49 42 49 51 40 47 53 62 59 47 49 53
60 73 63 84 90 88 94 81 79 81 89 85
91 83 97 92 98 89 86 92 88 95 91 92
end data.
execute.
```

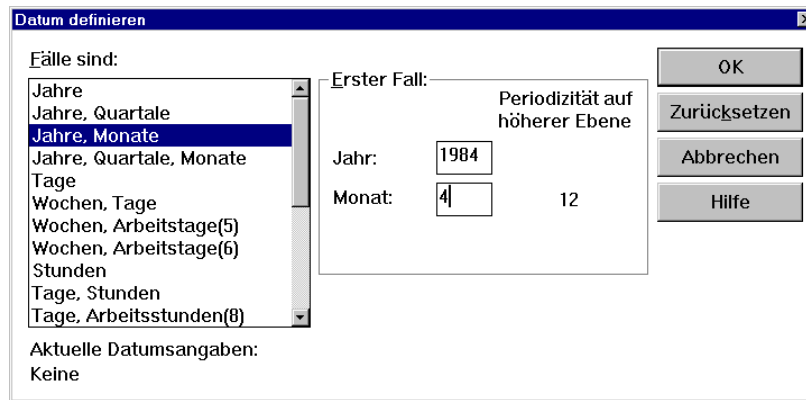
2.5.2 SPSS-Variablen bei der Zeitreihenanalyse

Eine empirische Zeitreihe wird als normale SPSS-Variable in der aktiven Arbeitsdatei gespeichert, d.h. jeder *Zeitpunkt* wird als *Fall* aufgefaßt. Folglich können auf Zeitreihen auch alle SPSS-Transformationsbefehle angewandt werden.

2.5.2.1 Datumsvariablen

Bei der Zeitreihenanalyse mit SPSS kann man optional mit **Datumsvariablen** arbeiten, die für jede Beobachtung den Beobachtungszeitpunkt angeben (z.B. als Datums- oder Zeitangabe). Sie verbessern die Lesbarkeit der Ausgabe und erleichtern die in Abschnitt 2.5.2.2 beschriebene Auswahl eines bestimmten Teils der Zeitreihe zur Schätzung eines Modells (historische Periode) bzw. zur Validierung des Modells (Validierungsperiode). Wegen der Gleichabständigkeit aller Beobachtungen brauchen die Datumsvariablen nicht eingelesen zu werden, sondern können per Menü mit **Daten, Datum definieren...** oder mit dem Kommando **DATE** generiert werden.

Durch Abschicken der folgenden Dialogbox:



bzw. des äquivalenten Kommandos:

```
DATE YEAR 1984 MONTH 4 12.
```

werden drei neue Variablen erstellt:

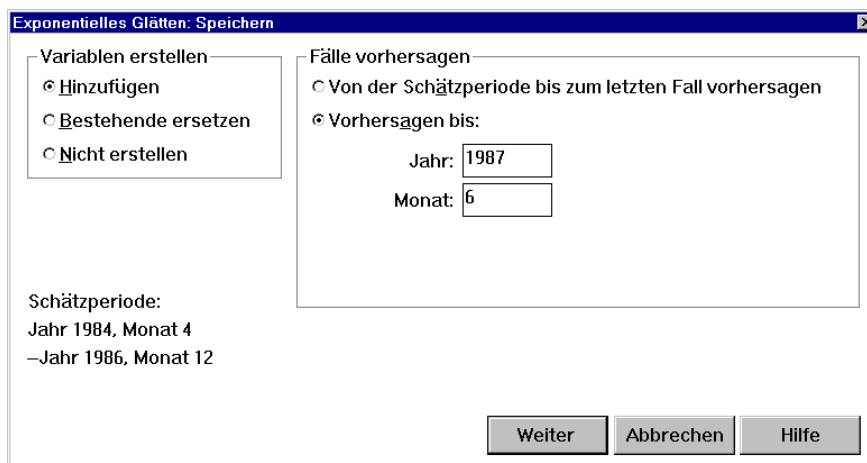
- DATE_ ist eine Datumsvariable im Format "MMM YYYY". Sie enthält in unserem Beispiel für die erste Beobachtung den Wert "APR 1984", für die zweite Beobachtung den Werte "MAY 1984" usw.
- YEAR_ und MONTH_ sind numerische Variablen, welche die Jahres- bzw. Monatsangabe zu den einzelnen Beobachtungen enthalten.

Wir werden gleich sehen, wie diese Variablen zur bequemen Auswahl von Zeitbereichen verwendet werden können.

Außerdem wird die Periodizität 12 vereinbart, was aber nur bei Verfahren zur saisonalen Zerlegung relevant ist, die in diesem Kurs nicht behandelt werden.

2.5.2.2 Automatische Erzeugung neuer Variablen (Residuen, Prognosen etc.)

Die Trends-Prozeduren AREG, ARIMA, EXSMOOTH, SEASON und X11ARIMA erzeugen neue Variablen, die Residuen, prognostizierte Werte, Konfidenzintervalle etc. enthalten. Diese Variablen werden in der aktiven SPSS-Arbeitsdatei gespeichert. Sie können über ihre im Ausgabefenster vermerkten Namen angesprochen und beliebig weiterverarbeitet, z.B. geplottet, werden. Sie können die automatische Erzeugung neuer Variablen prozedur-spezifisch in der **Speichern**-Subdialogbox beeinflussen, z.B. beim exponentiellen Glätten:



Die drei Optionen im Bereich **Variable erstellen** bedeuten:

- Hinzufügen

Die Prozedur erzeugt neue Variablen und speichert diese permanent in der Arbeitsdatei unter eindeutigen Namen ab, ohne bereits vorhandene Variablen zu überschreiben. Die automatisch gebildeten Variablennamen werden im Ausgabefenster protokolliert, z.B.:

The following new variables are being created:

NAME	LABEL
FIT_2	Fit for SALES from EXSMOOTH, MOD_2 NN A ,10
ERR_2	Error for SALES from EXSMOOTH, MOD_2 NN A ,10

Unter dieser Einstellung kann es also im Verlauf einer Sitzung zu einer Variableninflation kommen.

- Bestehende ersetzen

Die Prozedur erzeugt neue Variablen und speichert diese in der Arbeitsdatei ab, wobei bereits vorhandene Variablen vom selben Typ überschrieben werden, z.B.:

The following new variables are being created:

NAME	LABEL
FIT#1	Fit for SALES from EXSMOOTH, MOD_2 NN A ,10
ERR#1	Error for SALES from EXSMOOTH, MOD_2 NN A ,10

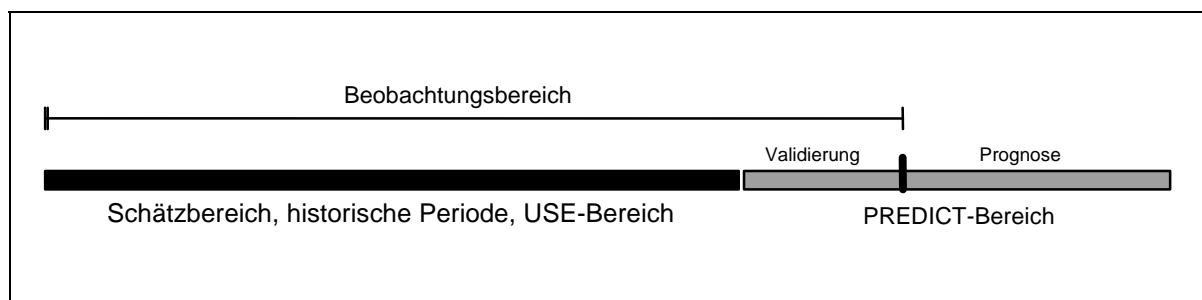
Die Numerierung bezieht sich hier auf die in der aktuellen Prozedur zu analysierenden Zeitreihen. Wird während der Sitzung nur eine Zeitreihe untersucht, bleibt es z.B. bei einer einzigen Residualvariablen, die den Namen ERR#1 erhält.

- Nicht erstellen

Die Prozedur erzeugt keine neue Variablen.

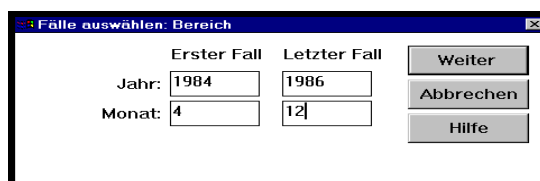
2.5.3 Schätz-, Validierungs- und Prognosebereich

Häufig ist es sinnvoll, nur einen Teil der verfügbaren Zeitreihe zur Modellentwicklung zu verwenden (Schätzbereich, historische Periode) und dann mit diesem Modell zu Validierungszwecken Werte für den restlichen Teil der Zeitreihe (Validierungsbereich) zu "prognostizieren". Natürlich ist man oft auch an einer echten Prognose über das beobachtete Zeitintervall hinaus interessiert:



2.5.3.1 Festlegung des Schätzbereichs

Zur Festlegung des *Schätzbereichs* wird mit **Daten, Fälle auswählen ...**, **Nach Fall- oder Zeitbereich** die folgende Dialogbox angefordert:



In diesem Beispiel soll die historische Periode im Dezember 1986 enden. Denselben Effekt erreicht man auch mit dem folgenden **USE**-Kommando:

```
use year 1984 month 4 thru year 1986 month 12.
```

Im **USE**-Kommando werden die Datumsvariablen YEAR_ und MONTH_ (vgl. Abschnitt 2.5.2) verwendet, wobei die abschließenden Unterstriche weggelassen werden müssen.

2.5.3.2 Festlegung des Validierungs- und/oder des Prognosebereichs

Die Festlegung eines *Validierungs-* und/oder eines *Prognosebereichs* kann für eine betroffene Prozedur in der **Speichern**-Subdialogbox erfolgen. Mit der Dialogbox in Abschnitt 2.5.2.2 zum Verkaufszahlenbeispiel werden die letzten drei Beobachtungszeitpunkte (1/87, 2/87 und 3/87) als Validierungsperiode festgelegt und für die Monate 4/87, 5/87 und 6/87, zu denen keine Beobachtungen vorliegen, Prognosen angefordert.

Denselben Effekt erreicht man auch mit dem folgenden **PREDICT**-Kommando:

```
predict thru year 1987 month 6.
```

Im **PREDICT**-Kommando werden die Datumsvariablen YEAR_ und MONTH_ (vgl. Abschnitt 2.5.2) verwendet, wobei die abschließenden Unterstriche weggelassen werden müssen.

Die Festlegung eines *Validierungs-* und/oder eines *Prognosebereichs* wirkt sich auf die Trends-Prozeduren AREG, ARIMA, CURVEFIT und EXSMOOTH aus (bis zu einer neuen Festlegung).

Endet der Schätzbereich vor der letzten Beobachtung, dann behandelt SPSS-Trends automatisch die Zeitpunkte nach der Schätzbereich als Validierungsbereich.

2.5.3.3 Aktuelle Einstellungen für Schätz-, Validierungs- und Prognosebereich überprüfen

Nach Ausführung der Zeitreihen-Prozedur **Experimentelles Glätten** repräsentieren in der SPSS-Arbeitsdatei drei neue Fälle den oben definierten Prognosebereich. Außerdem sind die Nummern aller Fälle im Validierungs- und im Prognosebereich durchgestrichen, was den Ausschluß von den Schätzprozeduren zur Folge hat. Die fällige Abbildung wird in Abschnitt 3.2.1 nachgeliefert.

Mit dem Kommando

```
tshow.
```

können Sie SPSS auffordern, die aktuellen Einstellungen für Schätz-, Validierungs- und Prognosebereich im Ausgabefenster zu protokollieren, z.B.:

```
Date Related Settings -

Use:      YEAR 1984,  MONTH APRIL  TO   YEAR 1986,  MONTH DECEMBER

Predict:  First observation following use period  to   YEAR 1987,  MONTH JUNE

Time Series date variables defined:  YEAR_ MONTH_ DATE_
                                      Period: 12, imputed by SPSS
```

2.5.4 Zeitreihenspezifische Transformationen

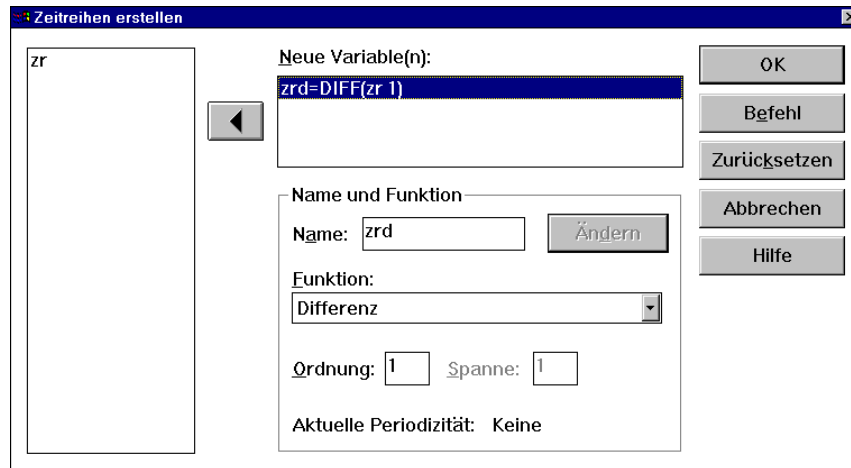
Bereits das SPSS-Basispaket bietet einige nützliche Transformationsmöglichkeiten für Zeitreihen, die allerdings in unserem Beispiel nicht benötigt werden:

2.5.4.1 Erstellung neuer Zeitreihen

Mit dem Menübefehl **Transformieren, Zeitreihen erstellen...** bzw. mit dem Kommando **CREATE** können neue Zeitreihen aus vorhandenen erzeugt werden, wobei z.B. folgende Funktionen zur Verfügung stehen:

- Gleitmittel
- Differenzenbildung
- kumulierte Summe
- schnelle Fourier-Transformation

Mit der folgenden Dialogbox wird die neue Variable (bzw. Zeitreihe) ZRD durch Bildung der ersten Differenz aus der Zeitreihe ZR erstellt:



Wird diese Dialogbox abgeschickt, dann führt SPSS im Hintergrund den folgenden Befehl aus:

```
CREATE /zrd=DIFF(zr 1).
```

2.5.4.2 Ersetzung fehlender Werte

Viele Prozeduren können nicht auf Zeitreihen angewendet werden, die *im Mittelteil* fehlende Werte aufweisen (z.B. Exponentielle Glättung, Spektralanalyse). Mit dem Menübefehl **Transformieren, Missing-Werte ersetzen...** bzw. mit dem Kommando **RMV** können fehlende Werte durch Anwendung verschiedener Interpolationsmethoden ersetzt werden.

Aufgrund der folgenden Dialogbox entsteht die neue Variable (bzw. Zeitreihe) ZRE aus der alten Variablen ZR durch lineare Interpolation fehlender Werte:



Wird die Dialogbox abgeschickt, dann führt SPSS im Hintergrund den folgenden Befehl aus:

```
RMV /zre=LINT(zr).
```

Wenn z.B. der X-Wert zur achten Beobachtung fehlt, und die Werte der Nachbarn vorhanden sind, dann wird der fehlende Wert ersetzt durch :

$$\frac{X(7) + X(9)}{2}$$

Mit dieser linearen Interpolation wird leicht die Option **Linearer Trend an dem Punkt** verwechselt. Beim letztgenannten Verfahren wird eine lineare Regression:

$$\hat{X} = \alpha + \beta I$$

der vorhandenen Werte auf die automatische Fallnummernvariable (mit Werten $I = 1, 2, 3, \dots$) gerechnet. Der fehlende Wert zum achten Fall wird dann durch die Regressionsprognose $\alpha + \beta \cdot 8$ ersetzt.

3 Exponentielles Glätten

3.1 Grundprinzipien des Verfahrens

Bei den Verfahren aus dieser Familie steht die Anwendung zur Prognose im Vordergrund (siehe z.B. Schlittgen & Streitberg, 1989, S. 380ff). Aus der gesamten Vergangenheit einer Zeitreihe bis zum Zeitpunkt t können Prognosen für ihren weiteren Verlauf erstellt werden. Wir wollen uns auf die **Einschritt-Prognose** \hat{X}_t von X_{t+1} aufgrund der Informationen bis zum Zeitpunkt t beschränken.

Leitidee beim Entwurf des Verfahrens war, die Prognose \hat{X}_t möglichst einfach aus der letzten Prognose \hat{X}_{t-1} und der neu hinzu gekommenen Beobachtung X_t berechnen zu können. Man entschied sich für die gewichtete Summe:

$$\hat{X}_t := \alpha X_t + (1 - \alpha) \hat{X}_{t-1}$$

Der Name des Verfahrens rührt daher, daß diese Prognose auch als exponentiell gewichtete Summe aller bisherigen Zufallsvariablen dargestellt werden kann:

$$\hat{X}_t = \sum_{u=0}^{\infty} \alpha(1 - \alpha)^u X_{t-u}, \alpha \in [0,1]$$

Bei α gleich Eins beruht die neue Schätzung ausschließlich auf der aktuellen Beobachtung, der Vergangenheit wird also keinerlei Bedeutung mehr beigemessen. Je mehr sich α dem Wert Null nähert, desto geringer ist der Einfluß der aktuellen Beobachtung im Vergleich zur Vergangenheit, desto stärker ist dementsprechend die Glättungswirkung.

Da bei einer empirischen Zeitreihe $(x_t)_{t=1, \dots, N}$ keine unendliche Vergangenheit vorliegt, benötigt das Prognoseverfahren einen Startwert. SPSS-Trends verwendet per Voreinstellung das arithmetische Mittel der empirischen Zeitreihe:

$$\hat{x}_0 := \bar{x}$$

Allerdings können Sie auch einen alternativen Wert festlegen. Davon ausgehend werden die Einschritt-Prognosen zu einer empirischen Zeitreihe dann wie in dem oben angegebenen induktiven Schema ermittelt:

$$\hat{x}_t := \alpha x_t + (1 - \alpha) \hat{x}_{t-1}$$

Sie können Trends auffordern, durch Ausprobieren einen α -Wert zu bestimmen, der zu einer möglichst kleine Summe quadrierter Abweichungen zwischen den Prognosen $\hat{x}_t(\alpha)$ und den Beobachtungen x_{t+1} führt:

$$\sum_{t=0}^{N-1} (\hat{x}_t(\alpha) - x_{t+1})^2 = \min_{0 \leq \alpha \leq 1}$$

Das bislang beschriebene Prognoseverfahren ist nur für Zeitreihen *ohne* Trend und *ohne* Saisonkomponente geeignet. Trends unterstützt ein verallgemeinertes Verfahren, in dem über die Parameter γ (Gamma), δ (Delta) und ϕ (Phi) auch Trend-, Saison- und *ausklingende* Trendeffekte berücksichtigt werden können.

Wir wollen z.B. annehmen, daß die untersuchte Zeitreihe $(X_t)_{t \in N}$ aus einer trend- und saisonfreien Zeitreihe $(Y_t)_{t \in N}$ durch Addieren eines linearen Trends entstanden sei:

$$X_t = b_0 + b_1 t + Y_t, t = 1, 2, 3, \dots, b_0, b_1 \in \mathbb{R}$$

SPSS-Trends erstellt seine Prognose \hat{X}_t für den nächsten Zeitpunkt $t+1$ (mit Abstand Eins zum aktuellen Zeitpunkt) dann mit Hilfe eines Schätzers \hat{H}_t für die Höhe zum Zeitpunkt t und eines Trendschätzers \hat{S}_t für die Steigung zum Zeitpunkt t :

$$\hat{X}_t := \hat{H}_t + \hat{S}_t$$

Die bei der Prognose \hat{X}_t (der Zeitreihe zum Zeitpunkt $t+1$) verwendeten Schätzer \hat{H}_t und \hat{S}_t werden folgendermaßen definiert:

$$\hat{H}_t := \alpha X_t + (1 - \alpha) \hat{X}_{t-1} = \hat{X}_{t-1} + \alpha(X_t - \hat{X}_{t-1}) = \hat{H}_{t-1} + \hat{S}_{t-1} + \alpha(X_t - \hat{X}_{t-1})$$

$$\hat{S}_t := \hat{S}_{t-1} + \alpha\gamma(X_t - \hat{X}_{t-1})$$

Der aktuelle Prognosefehler $(X_t - \hat{X}_{t-1})$ führt also mit der generellen Anpassungsgeschwindigkeit α zu Korrekturen der Niveau- und Trendschätzer. Über den zusätzlichen Parameter γ kann die spezielle Anpassungsgeschwindigkeit für die Trendschätzung feingesteuert werden. Bei $(\gamma = 0)$ bleibt z.B. der Steigungsschätzer für alle Zeitpunkte gleich dem Initialwert \hat{s}_0 , den SPSS für eine empirische Zeitreihe folgendermaßen wählt:

$$\hat{s}_0 := \frac{x_N - x_1}{N - 1}$$

Der Vollständigkeit halber soll noch der bei SPSS voreingestellte Startwert für den Lageschätzer mitgeteilt werden:

$$\hat{h}_0 := x_1 - \frac{1}{2} \hat{s}_0$$

Offenbar wird mit Hilfe des schon bekannten initialen Trendschätzers \hat{s}_0 "vorsichtig" (Faktor $\frac{1}{2}$) nach links extrapoliert.

Als Vorteile der exponentiellen Glättung sind neben der formalen Einfachheit vor allem der geringe Rechenaufwand und der niedrige Speicherplatzbedarf zu nennen. Bei Hinzutreten einer neuen Beobachtung muß zum Aktualisieren der Prognose nicht mehr die gesamte bisherige Zeitreihe durchlaufen werden, sondern es muß nur auf die bisherige Schätzung zurückgegriffen werden. Daher wird das Verfahren vor allem dort gewählt, wo sehr viele Zeitreihen gleichzeitig prognostiziert werden müssen.

Im Vergleich zu den später ausführlicher darzustellenden ARIMA-Modellen bietet die exponentielle Glättung im allgemeinen eine schlechtere Prognosequalität. Dies erklärt sich daraus, daß ein exponentielles Glättungsverfahren gerade die bestmöglichen Prognosen im Rahmen eines speziellen ARIMA-Modells ergibt. Die einfache exponentielle Prognoseformel für Modelle ohne Trend- und Saisonkomponente ist z.B. optimal in einem ARIMA(0,1,1)-Modell (Schlittgen & Streitberg 1989, S. 386). Bei der ARIMA-Modellierung kann jedoch für eine konkret vorliegende Zeitreihe das günstigste Modell frei gewählt und die Prognose darauf begründet werden.

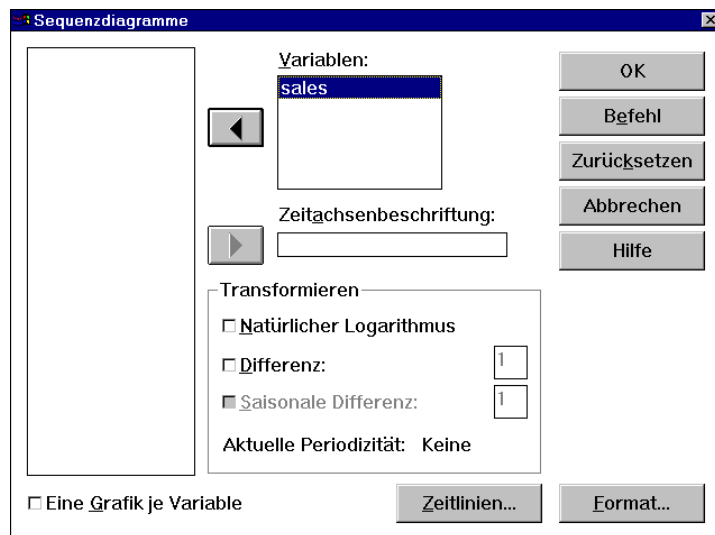
3.2 Anwendung auf die SALES-Reihe

3.2.1 Modellspezifikation und Schätzergebnisse

Wir wenden nun das Verfahren der exponentiellen Glättung auf die SALES-Zeitreihe an, die wir schon aus den Abschnitten 1.1.2 und 2.5 kennen. Die in Abschnitt 2.5 vorgenommenen Vereinbarungen und Einstellungen bleiben aktiv (z.B. bzgl. der Schätzperiode).

Man sollte eine Zeitreihe routinemäßig zu Beginn der Analyse grafisch darstellen, um einen ersten Eindruck von ihrem Verhalten zu gewinnen. Trends bietet dazu die beiden Kommandos **TS**PLOT und **CASE**PLOT für eine Darstellung in horizontaler bzw. vertikaler Richtung. Sie sind in der Windows-Benutzeroberfläche über den Menübefehl **Grafik, Sequenz...** verfügbar.

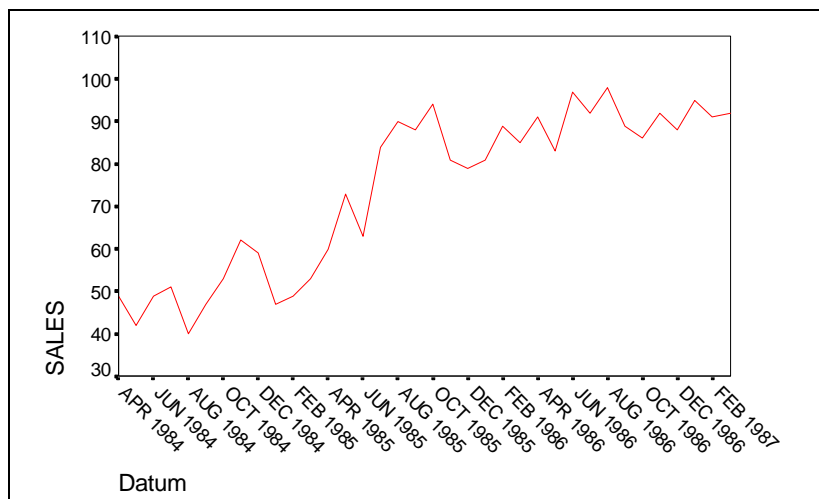
Vor der Definition der Schätzperiode erhalten wir für unsere Beispiel mit der Dialogbox:



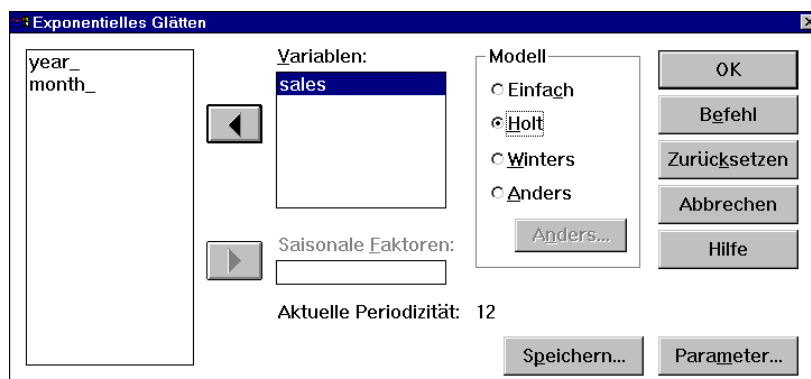
bzw. mit dem äquivalenten Kommando:

```
TSPLLOT VARIABLES= sales /NOLOG /FORMAT NOFILL NOREFERENCE.
```

die schon aus Abschnitt 1.1.2 bekannte Abbildung:



Die Zeitreihe zeigt einen deutlichen linearen Trend sowie keine nennenswerte Saisonalität. In der mit **Statistik, Zeitreihen, Exponentielles Glätten...** aufgerufenen Dialogbox wählen wir daher das Modell LN (linear trend, no seasonal component), das in SPSS nach seinem Urheber auch als "Holt"-Modell bezeichnet wird:



Zunächst sollen die günstigsten Werte für die Parameter α und γ über eine sogenannte Gittersuche bestimmt werden: SPSS soll z.B. für α alle Werte zwischen 0 und 1 im Abstand 0,1 ausprobieren. Daher aktivieren wir die **Parameter**-Subdialogbox und fordern dort für α und γ eine Gittersuche an:

Außerdem verlangen wir mit der in Abschnitt 2.5.2.2 abgebildeten **Speichern**-Subdialogbox Prognosen bis zum Juni 1987.

Die schließlich per **OK** in der Hauptdialogbox ausgelösten Berechnungen können wir auch mit folgenden Kommandos anfordern:

```
TSET PRINT=DEFAULT NEWVAR=ALL .
PREDICT THRU YEAR 1987 MONTH 6 .
EXSMOOTH /VARIABLES=sales
/MODEL=HOLT
/ALPHA=GRID(0 1 .1)
/GAMMA=GRID(0 1 .2)
/INITIAL=CALCULATE.
```

SPSS-Trends berechnet die Summe der quadrierten Abweichungen:

$$\sum_{t=0}^{N-1} (\hat{x}_t(\alpha, \gamma) - x_{t+1})^2$$

für jede Parameter-Kombination auf dem Gitter und informiert über die 10 Spitzenreiter:

The 10 smallest SSE's are:	Alpha	Gamma	SSE
	,6000000	,0000000	1894,10722
	,7000000	,0000000	1895,61539
	,5000000	,0000000	1927,22676
	,8000000	,0000000	1929,89344
	,9000000	,0000000	1998,41128
	,4000000	,0000000	2001,56915
	1,0000000	,0000000	2105,84009
	,3000000	,0000000	2133,54584
	,6000000	,2000000	2169,76937
	,7000000	,2000000	2175,52716

Wir wählen also für die Anpassungsgeschwindigkeit α den Wert 0,6 und für γ den Wert 0,0, d.h. die Wachstumsrate wird für alle Zeitpunkte konstant geschätzt (vgl. Abschnitt 3.1). Zu diesem günstigsten Modell erzeugt SPSS in der Arbeitsdatei die Variable FIT_1 mit den Prognosen \hat{x}_t und ERR_1 mit den Fehlern $(x_t - \hat{x}_t)$ (vgl. Abschnitt 2.5.2.2):

The following new variables are being created:

NAME	LABEL
FIT_1	Fit for SALES from EXSMOOTH, MOD_1 HO A ,60 G ,00
ERR_1	Error for SALES from EXSMOOTH, MOD_1 HO A ,60 G ,00

3 new cases have been added.

Im letzten Satz dieses Ergebnisausschnitts protokolliert SPSS, daß aufgrund unserer Vereinbarung einer Prognoseperiode (vgl. Abschnitt 2.5.3.2) drei neue Fälle (Zeitpunkte) in die Arbeitsdatei aufgenommen worden sind. Für die von uns vereinbarte Validierungsperiode (Januar bis März 1987) wurden die Prognosen natürlich *ohne* Verwendung der Beobachtungen aus diesem Zeitraum erstellt. Anstatt der (vorhandenen) Beobachtungen wurden die vom Modell für diese Zeitpunkte geschätzten Werte herangezogen.

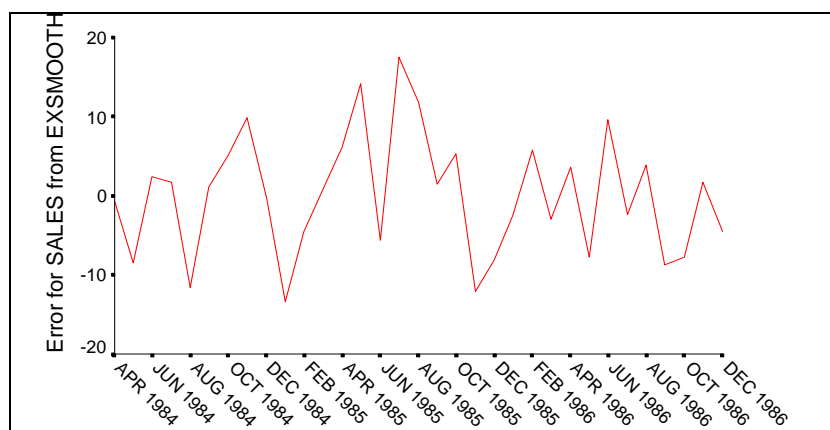
In unserem kleinen Beispiel sind auch die Fehler in den Zwei- und Dreischritt-Prognosen akzeptabel:

36.err_1	sales	year_	month_	date_	fit_1	err_1
31	86,00	1986	10	OCT 1986	93,69161	-7,69161
32	92,00	1986	11	NOV 1986	90,29539	1,70461
33	88,00	1986	12	DEC 1986	92,53691	-4,53691
34	95,00	1987	1	JAN 1987	91,03351	3,96649
35	91,00	1987	2	FEB 1987	92,25226	-1,25226
36	92,00	1987	3	MAR 1987	93,47101	-1,47101
37	,	1987	4	APR 1987	94,68976	,
38	,	1987	5	MAY 1987	95,90851	,
39	,	1987	6	JUN 1987	97,12726	,
40						

Zur Beurteilung der Modellgültigkeit und Modelltauglichkeit wollen wir nun die Residualzeitreihe ERR_1 genau untersuchen.

3.2.2 Modellbeurteilung

Da wir bei der exponentiellen Glättung eigentlich gar kein Modell postuliert, identifiziert und geschätzt haben (vgl. Abschnitt 1.4), kann (und muß) auch keine Prüfung der Modellgültigkeit erfolgen. Man wird das gewonnene Prognoseverfahren eher nach *Nützlichkeit* beurteilen. Trotzdem sollte i.a. auch anhand der Residuen grob geprüft werden, ob der Ansatz der exponentiellen Glättung problemadäquat ist. Der Residuenplot sollte ein "relativ zufälliges "Zucken" um die Nulllage zeigen, ohne prägnantes Muster. In unserem Beispiel scheint dies der Fall zu sein:



Eine Überprüfung der ACF zu den Residuen aus einer exponentiellen Glättung mit dem Box-Ljung-Test erscheint unangemessen, denn wir haben schon in Abschnitt 3.1 festgehalten, daß ein exponentielles Glättungsverfahren gerade die bestmöglichen Prognosen im Rahmen eines speziellen ARIMA-Modells ergibt. Nun ist kaum damit zu rechnen, daß dieses implizit gewählte Modell perfekt zu den Daten paßt. Folglich werden die Residuen der experimentellen Glättung i.a. *kein* weißes Rauschen sein, sondern noch serielle Abhängigkeit enthalten.

Wir wollen trotzdem für unser Beispiel die ACF zu der Residualreihe betrachten. Dazu verwenden wir den Menübefehl **Grafik, Zeitreihen, Autokorrelationen...** oder das folgende Kommando:

```
acf err_1.
```

Wir finden relativ kleine Autokorrelationen, wollen jedoch aus den eben erwähnten Gründen die Box-Ljung-Tests nicht beachten:

```
Autocorrelations:  ERR_1  Error for SALES from EXSMOOTH, MOD_16 HO
```

Lag	Auto-Corr.	Stand. Err.	-1	-.75	-.5	-.25	0	.25	.5	.75	1	Box-Ljung	Prob.
			+-----+-----+-----+-----+-----+										
1	,058	,166					*					,121	,728
2	,034	,164					*					,163	,922
3	-,141	,161				***						,924	,820
4	-,155	,158				***						1,876	,759
5	,102	,156					**					2,304	,806
6	-,127	,153				***						2,993	,810
7	-,227	,150				*****						5,284	,625
8	,001	,147					*					5,284	,727
9	,103	,144					**					5,794	,760
10	,260	,141					*****					9,189	,514
11	,125	,138					***					10,015	,529
12	-,267	,135				*****						13,925	,306
13	,003	,132					*					13,925	,379
14	-,267	,128				*****						18,265	,195
15	,020	,125					*					18,291	,248
16	,020	,121					*					18,319	,306

Plot Symbols: Autocorrelations * Two Standard Error Limits .

Mit dem Kommando **FIT**, zu dem es *kein* Dialogbox-Äquivalent gibt, können wir nun anhand verschiedener Fehlermaße zur Residualzeitreihe ERR_1 analysieren, ob das akzeptierte Modell praktisch verwertbare Prognoseleistungen erbringt. Die im Kommando angegebene Anzahl der Freiheitsgrade stammt aus der Ausgabe zur exponentiellen Glättung:

```
fit errors=err_1 /obs=sales /dfe=31.
```

FIT Error Statistics

Error Variable		ERR_1
Observed Variable		SALES
N of Cases	Use	33
	Predict	3
Deg Freedom	Use	31
	Predict	3
Mean Error	Use	,0609
	Predict	,4144
Mean Abs Error	Use	6,1656
	Predict	2,2299
Mean Pct Error	Use	-,8585
	Predict	,4001
Mean Abs Pct Err	Use	9,0997
	Predict	2,3834
SSE	Use	1894,1072
	Predict	19,4651

MSE	Use	61,1002
	Predict	6,4884
RMS	Use	7,8167
	Predict	2,5472
Durbin-Watson	Use	1,8727
	Predict	1,4017

Unser Modell produziert offenbar brauchbare Prognosen: z.B. beträgt der mittlere absolute prozentuale Fehler in der Validierungsperiode nur 2,38%.

Am Ende der FIT-Ausgabe steht die Durbin-Watson - Statistik, die häufig zur Beurteilung der Autokorrelation erster Ordnung eingesetzt wird. Diese Statistik ist jedoch nur interpretierbar für Residuen aus Regressionsmodellen, wenn unter den erklärenden Variablen keine zeitverzögerte Variante der abhängigen Variablen auftaucht. In unserem Fall ist der Durbin-Watson-Test also nicht anwendbar.

Einen guten Eindruck von der Prognosequalität erhält man auch durch den gemeinsamen Plot der beiden Variablen SALES und FIT_1, wobei letztere die Modellvorhersage enthält. Damit alle Zeitpunkte einbezogen werden, muß die Restriktion der Schätzperiode aufgehoben werden. Dazu kann man z.B. nach dem Menübefehl **Daten, Fälle auswählen ...** in der Dialogbox **Fälle auswählen** die Option **Alle Fälle** reaktivieren.

Danach werden nicht nur alle eingelesenen Zeitpunkte berücksichtigt, sondern auch die drei von der exponentiellen Glättung erzeugten Zeitpunkte im Prognosebereich.

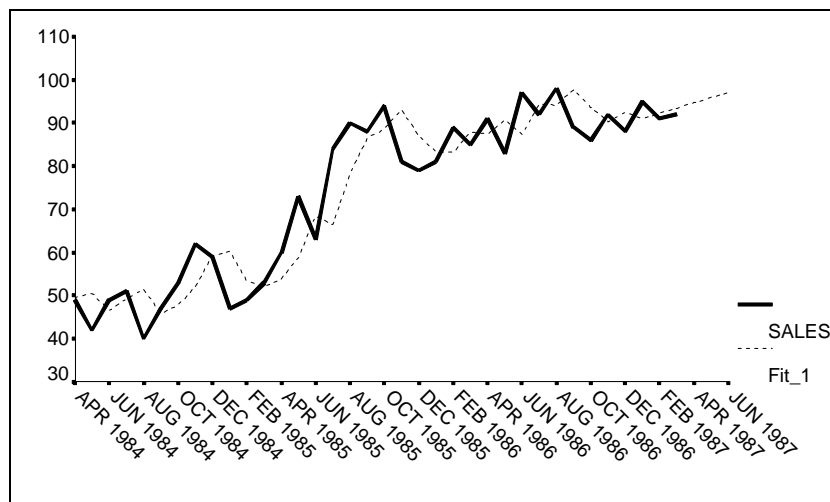
Das Syntaxäquivalent lautet:

```
use all.
```

Nun kann die gewünschte Abbildung mit der Dialogbox **Sequenzdiagramme** oder mit dem Kommando:

```
tsplot sales fit_1.
```

angefordert werden. Wir stellen fest, daß die FIT_1-Kurve eine leicht verzögerte und deutlich geglättete Variante der SALES-Kurve ist:



4 ARIMA-Modelle

4.1 Wiederholung: Grundprinzipien der ARIMA-Modellierung

Grundlegend für die Entwicklung der modernen Zeitreihenanalyse war die Entdeckung, daß periodische Bewegungen durch autoregressive Prozesse (siehe Abschnitt 4.1.14.1.2) oder durch gewichtete Summation von unabhängigen Zufallsvariablen (siehe Abschnitt 4.1.2) entstehen können. Die so produzierten Schwingungen zeichnen sich durch zufällig variierende Amplituden und Periodenlängen aus, so daß die zugrundeliegenden Modelle aussichtsreiche Kandidaten für die Erklärung nicht-physikalischer (z.B. ökonomischer) Zeitreihen sind.

In den folgenden Definitionen wird der Einfachheit halber angenommen, daß der voraussetzungsgemäß für alle Zufallsvariablen X_t identische Erwartungswert μ (Stationarität!) gerade Null ist. In der Praxis wird μ aus den Daten geschätzt und dann von allen Beobachtungen subtrahiert.

4.1.1 Der autoregressive Prozeß

Eine Zeitreihe $\dots, X_{-2}, X_{-1}, X_0, X_1, X_2, \dots$ heißt **autoregressiver Prozeß von der Ordnung p** (kurz: **AR(p)-Prozeß**), wenn für ein weißes Rauschen $\dots, E_{-2}, E_{-1}, E_0, E_1, E_2, \dots$ gilt:

$$X_t = \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + E_t.$$

4.1.2 Der Moving Average - Prozeß

Eine Zeitreihe $\dots, X_{-2}, X_{-1}, X_0, X_1, X_2, \dots$ heißt **moving average - Prozeß von der Ordnung q** (kurz: **MA(q)-Prozeß**), wenn für ein weißes Rauschen $\dots, E_{-2}, E_{-1}, E_0, E_1, E_2, \dots$ gilt:

$$X_t = E_t - \vartheta_1 E_{t-1} - \dots - \vartheta_q E_{t-q}.$$

Bemerkung: Die Verwendung der negativen Vorzeichen hat ausschließlich historische bzw. darstellungstechnische Gründe.

4.1.3 Der ARMA-Prozeß

Eine Zeitreihe $\dots, X_{-2}, X_{-1}, X_0, X_1, X_2, \dots$ heißt **ARMA(p,q) - Prozeß**, wenn für ein weißes Rauschen $\dots, E_{-2}, E_{-1}, E_0, E_1, E_2, \dots$ gilt:

$$X_t = \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + E_t - \vartheta_1 E_{t-1} - \dots - \vartheta_q E_{t-q}.$$

Der ARMA(p,q)-Prozeß entsteht also durch Kombination eines AR(p)-Prozesses mit einem MA(q)-Prozeß.

4.1.4 Der ARIMA-Prozeß

Es gibt viele nicht-stationäre Zeitreihen, die durch ein- oder mehrmalige Differenzenbildung in eine stationäre Zeitreihe transformiert werden können. Genügt eine Zeitreihe nach d -maliger Differenzenbildung einem ARMA(p,q)-Modell, so wird sie als **ARIMA(p,d,q)-Prozeß** bezeichnet. Die ursprüngliche Zeitreihe kann dann aus der ARMA-Reihe durch Summation gewonnen werden, z.B. für $d = 1$:

$$Y_t = X_t - X_{t-1} \Leftrightarrow X_t = \sum_{j=0}^{\infty} Y_{t-j}$$

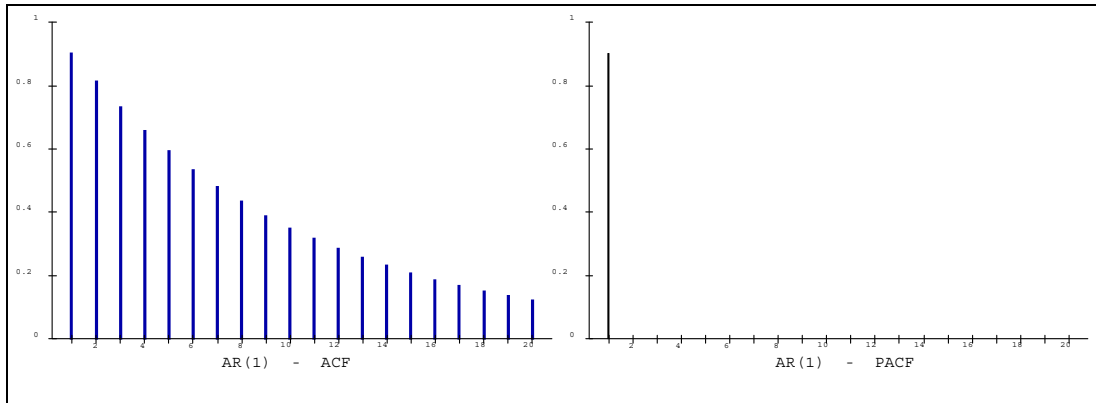
Den ARIMA($p,1,q$)-Prozeß erhält man also durch einmaliges Integrieren einer ARMA(p,q)-Zeitreihe.

4.1.5 Zur Modell-Identifikation

Wer sich zur ARIMA-Modellierung einer empirischen Zeitreihe entschließt, hat damit die Anwendbarkeit der generellen Klasse der ARIMA-Modelle postuliert. Die nächste Aufgabe besteht dann darin, durch Festlegung der Parameter d , p und q ein potentielles Modell zu bestimmen (vgl. das Ablaufschema von Box & Jenkins in Abschnitt 1.4).

Ist eine Differenzenbildung erforderlich und fehlen entsprechende Vorkenntnisse über die Zeitreihe, dann wählt man den Differenzierungsgrad d gerade so, daß die empirische Varianz der beobachteten Reihe möglichst klein wird.

Die Bestimmung von p und q ist problematisch. Man betrachtet dazu u.a. die Autokorrelationsfunktion (ACF) und die partielle Autokorrelationsfunktion (PACF), wobei die partielle Autokorrelation zum lag $t-s$ ($t > s$) die partielle Korrelation von X_t und X_s bei Kontrolle der dazwischen liegenden Zufallsvariablen X_{s+1}, \dots, X_{t-1} ist. Zumindest reine AR- bzw. MA-Prozesse lassen sich mit Hilfe ihrer ACF und ihrer PACF relativ gut identifizieren. Die folgende Abbildung zeigt für den AR(1)-Prozeß mit $\phi_1 = 0,9$ und $E_t \sim N(0,1)$ (d.h. E_t ist normalverteilt mit Erwartungswert Null und Varianz Eins) die theoretische ACF (links) und die theoretische PACF (rechts). Während die ACF exponentiell abfällt, bricht die PACF nach dem ersten lag ab.



Dieses Verhalten ist leicht nachzuvollziehen:

Zur ACF der stationären AR(1)-Zeitreihe:

Für den AR(1)-Prozeß gilt:

$$X_t = \phi_1 X_{t-1} + E_t$$

Wegen $\text{Kov}(E_t, X_t) = 0$ ist ϕ_1 gerade der Regressionskoeffizient für die einfache Regression von X_t auf X_{t-1} . Da die beiden Variablen identische Varianzen haben (Stationarität!) ist der Regressionskoeffizient identisch mit der Korrelation, die wir auch als die erste Autokorrelation der Zeitreihe bezeichnen. Da man für das weiße Rauschen eine positive Varianz annehmen wird, ist die Korrelation (also ϕ_1) betragsmäßig echt kleiner als Eins.

Durch zweimalige Anwendung des Zeitreihen-Bildungsgesetzes erhalten wir:

$$\begin{aligned} X_t &= \phi_1 X_{t-1} + E_t \\ &= \phi_1 (\phi_1 X_{t-2} + E_{t-1}) + E_t \\ &= \phi_1^2 X_{t-2} + \phi_1 E_{t-1} + E_t \end{aligned}$$

Folglich ist ϕ_1^2 gerade der Regressionskoeffizient für die einfache Regression von X_t auf X_{t-2} ist. Mit obigen Argumenten gibt daher ϕ_1^2 gerade die Korrelation zwischen X_t und X_{t-2} an, also die zweite Autokorrelation. Offensichtlich ist die k -te Autokorrelation im AR(1)-Prozeß gerade gleich ϕ_1^k , so daß (bei $\phi_1 < 1$) der in obiger Abbildung ersichtliche exponentielle Verlauf resultiert.

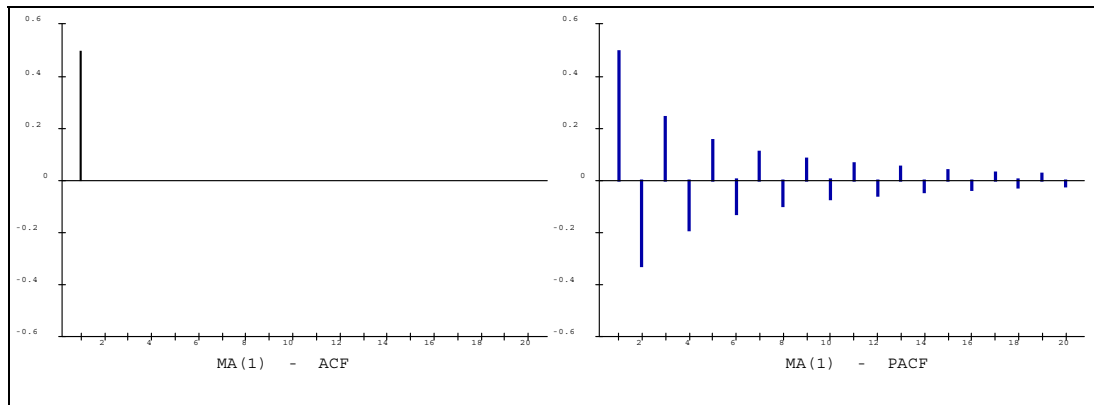
Zur PACF der stationären AR(1)-Zeitreihe:

Wird aus $X_t = \phi_1 X_{t-1} + E_t$ die Variable X_{t-1} auspartialisiert, so erhalten wir als X_{t-1} -bedingten Rest schlicht die Fehlervariable E_t , die definitionsgemäß für alle $u > 1$ mit X_{t-u} und E_{t-u} unkorreliert ist. Damit sind aber alle partiellen Korrelationen zwischen X_t und früheren Variablen X_{t-u} ($u > 1$) gleich Null.

Für AR-Prozesse gilt generell:

- Sie sind nur unter bestimmten Bedingungen stationär (vgl. Schlittgen & Streitberg 1989, S. 100ff)
- Die ACF geht exponentiell gegen Null, in Abhängigkeit von den Vorzeichen der Parameter eventuell oszillierend.
- Die PACF wird Null für lags, die größer sind als der Grad p , so daß der letzte lag-Wert mit einer von Null verschiedenen partiellen Autokorrelation den Grad angibt.

Wie die nächste Abbildung zeigt, ergibt sich beim MA(1)-Prozeß mit $\vartheta_1 = -0,9$ und $E_t \sim N(0,1)$ ein deutlich anderes Bild. Die ACF bricht nach dem ersten lag ab, während die PACF oszillierend exponentiell gegen Null konvergiert.



Auch diese Funktionsverläufe sind leicht zu verstehen:

Zur ACF der MA(1)-Zeitreihe:

Nach dem Zeitreihen-Bildungsgesetzes erhalten wir:

$$\begin{aligned} X_t &= E_t - \vartheta_1 E_{t-1} \\ X_{t-1} &= E_{t-1} - \vartheta_1 E_{t-2} \\ X_{t-2} &= E_{t-2} - \vartheta_1 E_{t-3} \end{aligned}$$

Damit haben X_t und X_{t-u} für $(u > 1)$ keine Störungsterme mehr gemeinsam, so daß sie unkorreliert sind.

Zur PACF der MA(1)-Zeitreihe:

Für das X_{t-1} -bedingte Residuum von X_t gilt:

$$\begin{aligned} \tilde{X}_{t|t-1} &= X_t - \beta_{t|t-1} X_{t-1} \\ &= X_t - \beta_{t|t-1} (E_{t-1} + \vartheta_1 E_{t-2}) \\ &= X_t - \beta_{t|t-1} E_{t-1} - \beta_{t|t-1} \vartheta_1 E_{t-2} \end{aligned}$$

Damit ist über das Auspartialisieren von X_{t-1} mit $\beta_{t|t-1} \vartheta_1 E_{t-2}$ wieder ein Störungsterm "eingeschleust" worden, der auch in X_{t-2} auftritt, so daß die *partielle* Autokorrelation zweiter Ordnung zwischen X_t und X_{t-2} i.a. von Null verschieden ist.

Für MA-Prozesse gilt generell:

- Sie sind stationär.
- Die PACF geht exponentiell gegen Null, in Abhängigkeit von den Vorzeichen der Parameter eventuell oszillierend.
- Die ACF wird Null für lags, die größer sind als der Grad q , so daß der letzte lag-Wert mit einer von Null verschiedenen Autokorrelation den Grad angibt.

Nach Bestimmung von d , p und q werden die Modellparameter mit der Maximum Likelihood - Methode geschätzt. Außerdem kann man Standardfehler zu den Parameterschätzungen bestimmen.

4.1.6 Zur Modell-Beurteilung

Zur Beurteilung von ARIMA-Modellen bietet Trends folgende Möglichkeiten:

- Mit dem Menübefehl **Grafik, Zeitreihen, Autokorrelationen...** oder mit dem Kommando **ACF** kann die empirische Autokorrelationsfunktion der Residuen berechnet, geplottet und mit dem Box-Ljung-Test beurteilt werden. Dieses Verfahren prüft für $m = 1, \dots, k (\leq N-2)$ die Hypothese, daß die ersten m Autokorrelationen der Modellresiduen gleich Null sind. Wenn für ein m das ausgegebene p -level den kritischen Wert (üblicherweise 0.05) unterschreitet, kann das Modell nicht akzeptiert werden. Leider muß die in Abschnitt 1.6.2 angegebene Freiheitsgrad-Korrektur bei einer ACF-Untersuchung einer Residualzeitreihe manuell vorgenommen werden (siehe Beispiel unten).
- Die ARIMA-Prozedur liefert mit dem Akaike Information Criterion (AIC, besonders geeignet für AR-Prozesse) und dem Schwartz Bayesian Criterion (SBC, allgemein verwendbar) zwei Gütekriterien, die vor allem für den Vergleich verschiedener ARIMA-Modelle zur selben Zeitreihe sehr nützlich sind. Sie beurteilen ein Modell nach seiner geschätzten Residualvarianz **und** der Anzahl verwendeter Parameter (Prinzip der Sparsamkeit!). Den SPSS-Handbüchern sind leider keine exakten Formeln für die Kriterien zu entnehmen, doch dürfte das SBC recht analog zu dem folgenden, unter dem Namen "Bayesian Information Criterion" (BIC) bekannten, Kriterium berechnet werden (vgl. Schlittgen & Streitberg 1989, S. 245ff):

$$BIC(p, q) := \ln(\hat{\sigma}_{p,q}^2) + \frac{(p + q)\ln(N)}{N}$$

4.2 Beispiel: Der Einsatz von ARIMA-Modellen in der Qualitätskontrolle

Wir wollen uns nun mit der in Abschnitt 1.1.4 vorgestellten Zeitreihe zum Qualitätsverhalten einer Produktionsanlage beschäftigen, wobei alle Phasen der Zeitreihenanalyse mit SPSS am Beispiel vorgeführt werden.¹ Das SPSS-Programm mit allen während der interaktiven Arbeit (explizit per Syntax-Fenster oder implizit per Dialogbox) verwendeten Trends-Kommandos findet sich in der Datei **QCTR.SPS**, die zusammen mit dem Manuskript verfügbar ist (siehe Vorwort). In einer Phase ungestörter Produktion soll für eine Fertigungsanlage ein ARIMA-Modell ihres Qualitätsverhaltens (Anzahl defekter Produkte) ermittelt werden. Mit Hilfe dieses Modells soll später zuverlässig entschieden werden können, ob die Zeitreihe der Produktionsfehler aus ihrem normalen Verlaufskorridor ausgebrochen ist, und daher die Fertigungsanlage angehalten werden muß.

Im Folgenden werden die wesentlichen Arbeitsschritte erläutert.

4.2.1 Einlesen der Zeitreihe und andere Vorbereitungen

Wir öffnen die SPSS-Datendatei **CHAP6.SAV** (zu finden an der im Vorwort vereinbarten Stelle) mit der Variablen bzw. Zeitreihe **SCORE** und wählen als Schätzbereich bzw. historische Periode (vgl. Abschnitt 2.5.3.1) eine Phase ungestörter Produktion (Fälle 1 bis 100). Dazu rufen wir mit **Daten, Fälle auswählen...** die zugehörige Dialogbox auf, steigen dann durch Markierung bei **Nach Zeit- oder Fallbereich** und Klick auf den Schalter **Bereich...** in die Subdialogbox zur Fallspezifikation ein, und legen dort fest:

¹ Das Beispiel wird im Kapitel 6 des Trends-Handbuchs (SPSS 1994) ausführlich diskutiert.

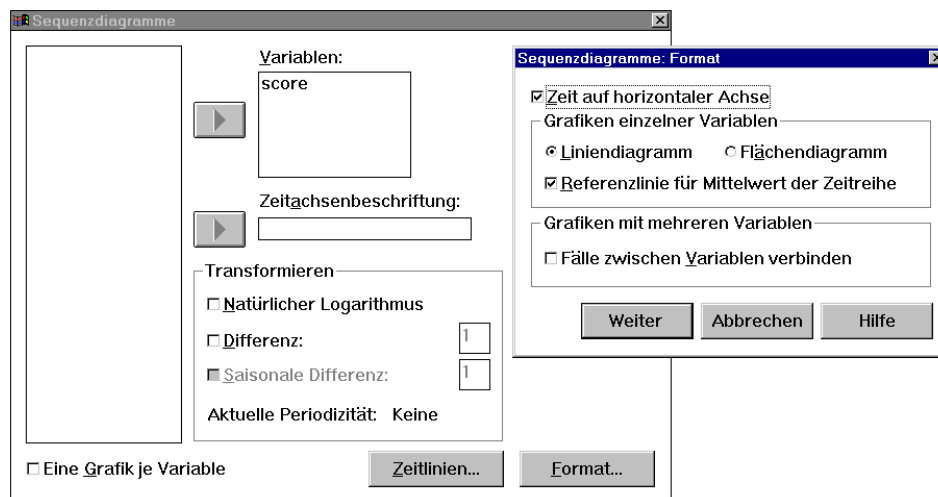


Zur Auswahl der Schätzperiode wird hier keine SPSS-Datumsvariable benutzt, sondern es wird auf die Anordnung der Beobachtungszeitpunkte (Fälle) zurückgegriffen. Zu obiger Dialogbox ist die folgende Syntax äquivalent:

```
use 1 thru 100 .
```

4.2.2 Plot der Zeitreihe zur Kontrolle auf Stationarität

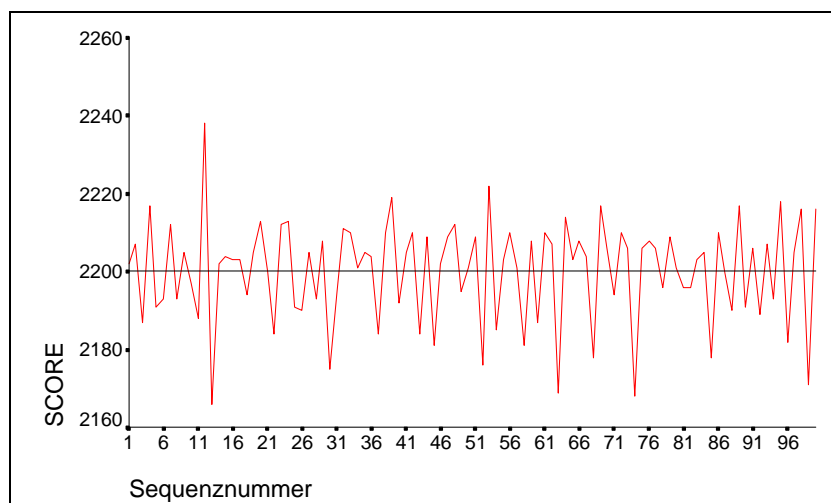
Um die Stationarität der Zeitreihe beurteilen zu können, fordern wir nach dem Menübefehl **Grafik, Sequenz** mit den folgenden Dialogboxen ein Sequenzdiagramm an:



In der Subdialogbox wird eine Referenzlinie in Höhe des Mittelwertes angefordert. Zu den beiden Dialogboxen ist das folgende Kommando äquivalent:

```
tsplot score /format reference.
```

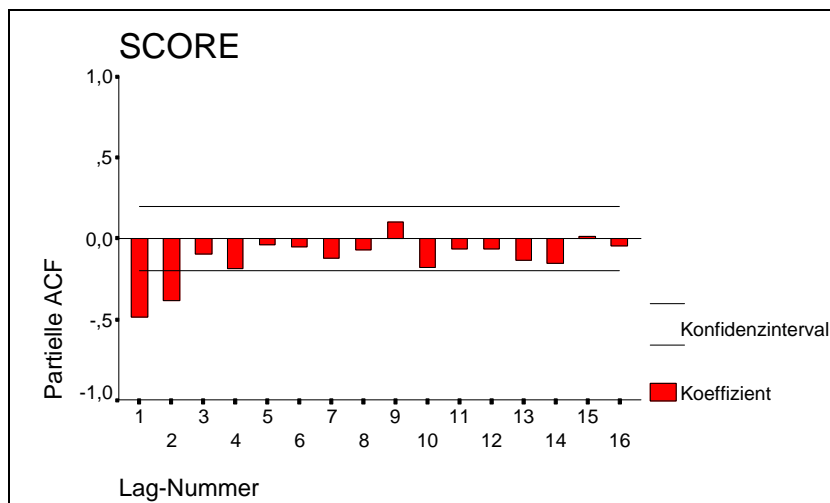
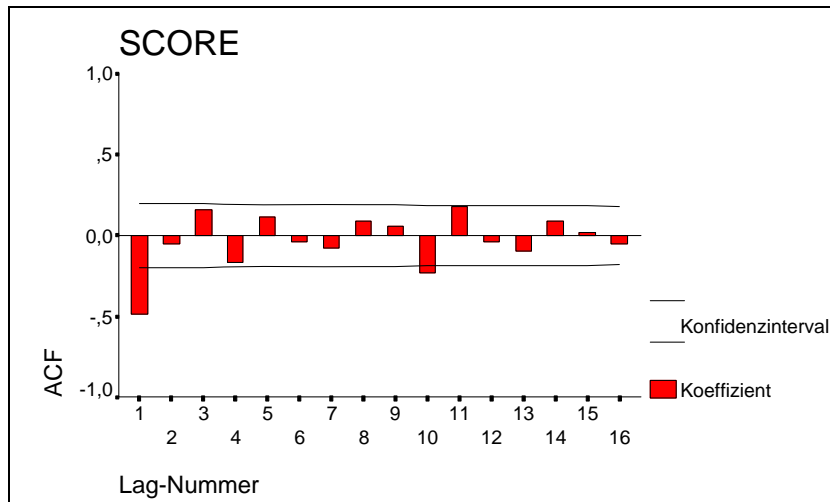
Der Plot zeigt Lage- und Streuungsstabilität über alle Zeitpunkte hinweg:



Daher wird die Stationaritätsannahme akzeptiert. Da keine Differenzenbildung erforderlich ist, liegt für den ARIMA-Parameter d der Wert Null fest.

4.2.3 Untersuchung der ACF und der PACF

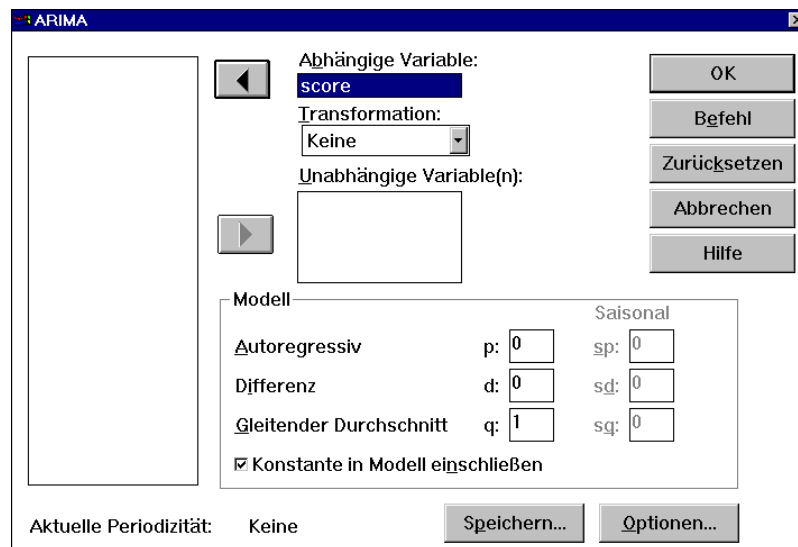
Wie die zentralen Hilfsmittel zur ARIMA-Modellidentifikation (ACF und PACF) in SPSS angefordert werden, braucht nach wiederholter Anwendung in diesem Kurs nicht mehr erwähnt zu werden. Als hochauflösende Grafiken sehen die beiden Funktionen folgendermaßen aus:



Die ACF bricht nach einem lag ab, während die PACF exponentiell abfällt. Aus diesem Befund kann man relativ sicher auf einen MA(1)-Prozeß bzw. auf einen ARIMA(0,0,1)-Prozeß schließen. Damit ist die Phase der Modell-Identifikation beendet.

4.2.4 Modellschätzung

Nun öffnen wir mit **Statistik, Zeitreihen, ARIMA...** die Dialogbox zur ARIMA-Prozedur und wählen dort:



Alternativ können Sie auch folgende Kommandos abschicken:

```
tset newvar=all.
arima score /model=(0,0,1).
```

Das **TSET**-Kommando sorgt dafür, daß die durch nachfolgende Prozeduren berechneten Variablen (z.B. Modellprognosen oder -residuen) in der Arbeitsdatei gespeichert werden (vgl. Abschnitt 6.1). Es wird beim Abschicken einer zeitreihenanalytischen SPSS-Dialogbox automatisch ausgeführt.

Die wesentlichen Ergebnisse:

```
Conclusion of estimation phase.
Estimation terminated at iteration number 5 because:
  Sum of squares decreased by less than ,001 percent.

FINAL PARAMETERS:

Number of residuals  100
Standard error       10,265823
Log likelihood       -374,23949
AIC                  752,47899
SBC                  757,68933

      Analysis of Variance:

      DF  Adj. Sum of Squares    Residual Variance
Residuals    98                10425,658            105,38713

      Variables in the Model:

              B          SEB      T-RATIO    APPROX. PROB.
MA1           ,78105     ,06411139     12,1828     ,0000000
CONSTANT     2200,16919    ,23323983    9433,0765     ,0000000

The following new variables are being created:

Name          Label
FIT_1         Fit for SCORE from ARIMA, MOD_3 CON
ERR_1         Error for SCORE from ARIMA, MOD_3 CON
LCL_1         95% LCL for SCORE from ARIMA, MOD_3 CON
UCL_1         95% UCL for SCORE from ARIMA, MOD_3 CON
SEP_1         SE of fit for SCORE from ARIMA, MOD_3 CON
```

Obwohl die Prüfung der Modellgültigkeit noch aussteht, soll schon einmal die geschätzte Modellgleichung exzerpiert werden:

$$X_t = 2200.16919 + E_t - 0.78105 \cdot E_{t-1}, \text{ mit } \text{Var}(E_t) = 105.38713$$

4.2.5 Modellbeurteilung

Im Anschluß an die Ausgabe der Schätzergebnisse wird mitgeteilt, daß die neue Variable ERR_1 die Residualreihe zu unserem Modell enthält. Bei Modellgültigkeit muß es sich dabei um ein weißes Rauschen handeln. Daher untersuchen wir die ACF und die PACF zur Residualreihe ERR_1. Weil der wichtige Box-Ljung-Test nur in der Textgrafik-Ausgabe erscheint, wird diese Ausgabeversion wiedergegeben:

```
Autocorrelations:  ERR_1  Error for SCORE from ARIMA, MOD_3 CON
```

Lag	Auto-Corr.	Stand. Err.	-1	-.75	-.5	-.25	0	.25	.5	.75	1	Box-Ljung	Prob.
1	-.002	,099					*					,000	,985
2	,009	,098					*					,008	,996
3	,125	,098					**					1,639	,651
4	-,086	,097					**					2,426	,658
5	,040	,097					*					2,596	,762
6	-,068	,096					*					3,100	,796
7	-,102	,095					**					4,237	,752
8	,017	,095					*					4,271	,832
9	-,052	,094					*					4,573	,870
10	-,239	,094			*	***						11,053	,353
11	,030	,093					*					11,159	,430
12	-,067	,093					*					11,672	,472
13	-,090	,092					**					12,616	,478
14	,077	,092					**					13,323	,501
15	,065	,091					*					13,829	,539
16	,013	,091					*					13,851	,610

Plot Symbols: Autocorrelations * Two Standard Error Limits .

Total cases: 100 Computable first lags: 99

```
Partial Autocorrelations:  ERR_1  Error for SCORE from ARIMA, MOD_3 CON
```

Lag	Pr-Aut-Corr.	Stand. Err.	-1	-.75	-.5	-.25	0	.25	.5	.75	1
1	-.002	,100					*				
2	,009	,100					*				
3	,125	,100					**				
4	-,087	,100					**				
5	,039	,100					*				
6	-,085	,100					**				
7	-,081	,100					**				
8	,002	,100					*				
9	-,027	,100					*				
10	-,239	,100			*	***					
11	,024	,100					*				
12	-,060	,100					*				
13	-,061	,100					*				
14	,032	,100					*				
15	,106	,100					**				
16	-,035	,100					*				

Plot Symbols: Autocorrelations * Two Standard Error Limits .

Total cases: 100 Computable first lags: 99

Nach Abschnitt 1.6.2 folgt die Box-Ljung-Prüfstatistik zum lag m bei der Residualzeitreihe aus einem ARIMA(0,0,1)-Modell unter der Nullhypothese, also bei Modellgültigkeit, für $m > 1$ einer χ^2_{m-1} -Verteilung. Die SPSS-Prozedur ACF berücksichtigt leider die Herkunft einer Residualzeitreihe **nicht**, unterläßt daher die erforderliche Freiheitsgrad-Korrektur und berechnet die Überschreitungswahrscheinlichkeiten anhand der χ^2_m -Verteilung.¹ Die Korrektur muß nun vom Anwender durchgeführt werden, z.B. mit dem folgenden SPSS-Programm, das für $m = 1, \dots, 16$ die korrekten Überschreitungswahrscheinlichkeiten errechnet und relativ leicht anhand aus der ACF-Ausgabe erstellt werden kann:

¹ Der SPSS-Konkurrent SAS gibt zusammen mit den ARIMA-Ergebnissen unaufgefordert die korrekten Testergebnisse aus (SAS-Version 6.12 auf der RZSUN00, die Autokorrelationen weichen durch Verwendung anderer Algorithmen minimal ab):

Autocorrelation Check of Residuals

To Lag	Chi Square	DF	Prob	Autocorrelations						
6	3.06	5	0.690	0.001	0.009	0.124	-0.085	0.040	-0.068	
12	11.73	11	0.384	-0.102	0.016	-0.050	-0.242	0.030	-0.066	
18	14.69	17	0.618	-0.090	0.078	0.065	0.013	0.069	-0.038	
24	19.28	23	0.685	-0.082	-0.028	-0.142	0.086	0.013	-0.022	

```

data list free /lagg bl.
begin data.
 1      ,000
 2      ,008
 3      1,639
 4      2,426
 5      2,596
 6      3,100
 7      4,237
 8      4,271
 9      4,573
10     11,053
11     11,159
12     11,672
13     12,616
14     13,323
15     13,829
16     13,851
end data.
compute pchi = 1-cdf.chisq(bl,lagg-1).
list var=lagg pchi.
    
```

Die Eingabedaten sind aus der ACF-Ausgabe durch Löschen bzw. Kopieren von rechteckigen Blöcken, die in SPSS erfreulicherweise mit Hilfe der rechten Maustaste markiert werden können, gewonnen worden. Sie finden das Programm am Ende der Datei **QCTR.SPS**, die zusammen mit dem Manuskript verfügbar ist (siehe Vorwort).

Wir erhalten folgende Ergebnisse:

LAGG	PCHI (BOX-LJUNG)
2	,93
3	,44
4	,49
5	,63
6	,68
7	,64
8	,75
9	,80
10	,27
11	,35
12	,39
13	,40
14	,42
15	,46
16	,54

Wie es sich für ein Lehrbuchbeispiel gehört, sind die Testergebnisse gut verträglich mit der Annahme, daß es sich bei den Residuen um ein weißes Rauschen handelt. Damit kann das ARIMA(0,0,1)-Modell akzeptiert werden. Wir sehen in obigen (P)ACF-Plots zwar einige Überschreitungen der 95%-Vertrauensgrenzen, die jedoch in Anbetracht der zahlreichen Beurteilungen mit der globalen Nullhypothese verträglich sind.

Folglich dürfen die Schätz- und Testergebnisse zu den Modellparametern interpretiert werden. Für den zentrale Parameter ϑ_1 erhalten wir den Schätzwert 0.78105. Die zugehörige Nullhypothese:

$$H_0 : \vartheta_1 = 0$$

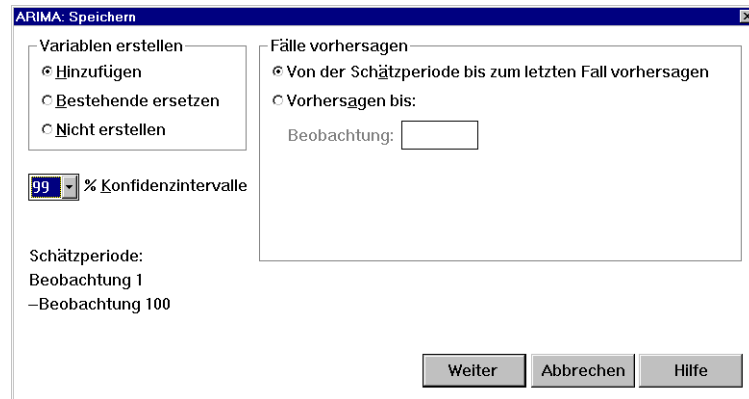
darf aufgrund der verschwindend kleinen Überschreitungswahrscheinlichkeit verworfen werden.

4.2.6 Erprobung des ARIMA-Modells als Störungsdetektor

Unsere Zeitreihe enthält noch 87 weitere Zeitpunkte, darin auch eine Phase mit einem bekannten Störfall (ab Zeitpunkt 160). Aufgrund einer Voreinstellung in der Subdialogbox **ARIMA: Speichern** (s.u.) hat SPSS diesen Zeitraum als Validierungsperiode verwendet und aufgrund des Modells zur korrekt funktionierenden

Anlage Prognosen sowie deren Vertrauensschranken berechnet. Es ist zu hoffen, daß der Störfall am Ausbrechen der Zeitreihe aus dem Vertrauenskorridor erkennbar wird.

Wir wollen allerdings noch eine kleine Verbesserung vornehmen und als statistische Sicherheit zu den Konfidenzintervallen die bei einer Prozeßkontrolle üblichen 99 Prozent einstellen. Dazu aktivieren wir erneut die Dialogbox **ARIMA** mit **Statistik, Zeitreihen, ARIMA...** und wechseln in die Subdialogbox **Speichern**:



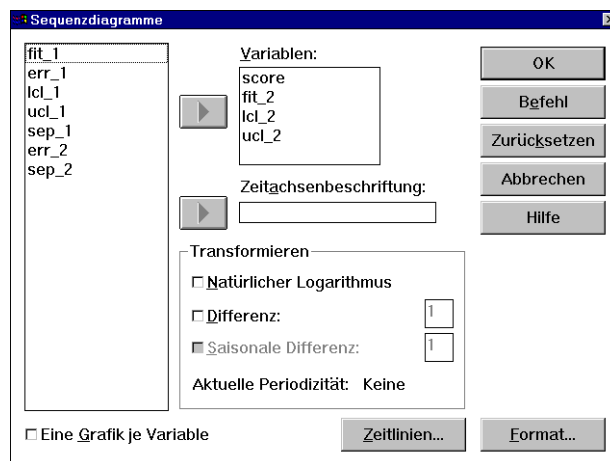
Nun sollen für alle Beobachtungen folgende Variablen gemeinsam geplottet werden:

- die SCORE-Zeitreihe selbst
- die FIT_2-Zeitreihe mit den Modellprognosen
- die Zeitreihen LCL_2 bzw. UCL_2 mit der unteren bzw. oberen Konfidenzschranke zum Niveau $\alpha = 1\%$.

Damit die Zeitpunkte im aktuellen Validierungsbereich bei dem Plot berücksichtigt werden, müssen sie in die historische Periode aufgenommen werden. Aktivieren Sie dazu mit dem Menübefehl **Daten, Fälle auswählen...** die zuständige Dialogbox und wählen Sie dort die Option: **Alle Fälle**. Denselben Effekt erreichen Sie mit folgendem Kommando:

```
use all.
```

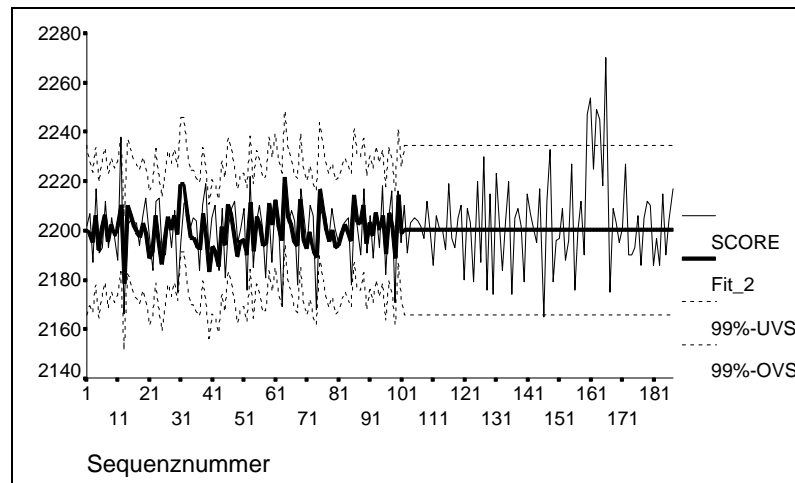
Das Sequenzdiagramm wird mit dem Menübefehl **Grafik, Sequenz...** und folgender Dialogbox angefordert:



Per Syntax erhalten Sie denselben Plot mit:

```
tsplot score fit_2 lcl_2 ucl_2.
```

Unser ARIMA-Modell zur Prozeßkontrolle reagiert nicht nervös auf Zufallsspitzen, entdeckt aber zuverlässig den echten Defekt:



An dem Sequenzdiagramm können wir folgende Eigenschaften von ARIMA-Prognosen erkennen:

- Das Vertrauensintervall wird nach Verlassen der Schätzperiode schnell breiter.
- Ist die zeitliche Distanz größer als ein Takt, wird konstant der geschätzte Zeitreihenmittelwert prognostiziert, weil offenbar die mehr als einen Takt zurückliegenden Beobachtungen keinen prognostischen Wert besitzen.

Oft wird von der Modelprognose erwartet, daß sie sich ähnlich dem beobachteten Zeitreihenverlauf endlos zuckend fortbewegen müsse. Hier wird offenbar *Prognose* mit *Simulation* verwechselt. Während von der Simulation ein naturgetreues Benehmen zu erwarten ist, hat die Prognose ein Fehlerkriterium zu minimieren, das in der Regel nach dem Prinzip der kleinsten Quadrate definiert wird. Wenn aber keine verwertbare Information vorliegt, ist sicherlich der geschätzte Reihenmittelwert die fehlervarianz-minimierende Prognose.

- Wenn als optimale Prognose nur mehr der geschätzte Mittelwert angeboten werden kann, bleiben auch die Konfidenzgrenzen konstant, insbesondere verbreitert sich das Konfidenzintervall nicht mehr. Die Varianz der wahren um die prognostizierten Werte ist jetzt identisch mit der Varianz der Zeitreihe.

Für das im Beispiel favorisierte MA(1)-Modell sollen die Eigenschaften der Prognose näher betrachtet werden. Die optimalen Prognosen $\hat{X}_{t,h}$ für Schrittweiten $h > 0$ lauten (vgl. Schlittgen & Streitberg 1989, S. 343):

$$\hat{X}_{t,h} = \begin{cases} -\vartheta_1 E_t & \text{für } h = 1 \\ 0 & \text{für } h > 1 \end{cases}$$

Beachten Sie bitte, daß wir in Abschnitt 4 grundsätzlich für die betrachteten theoretischen Zeitreihen der Einfachheit halber annehmen, daß ihr Mittelwert Null ist. Bei einer allgemeineren Formulierung erhalten wir im MA(1)-Prozeß als optimale Prognose für $h > 1$ also den Zeitreihenmittelwert.

Nach obiger Formel geht die Störgröße E_t in die optimale Prognose ein. Natürlich sind in der Praxis die Werte der Störgrößen unbekannt, so daß wir noch eine Formel zur optimalen Einschnitt-Prognose aufgrund der beobachteten Zeitreihen-Vergangenheit benötigen (vgl. Schlittgen & Streitberg 1989, S. 343):

$$\hat{X}_{t,1} = \sum_{u=0}^{\infty} \vartheta_1^{u+1} X_{t-u}$$

In die optimale Einschnitt-Prognose geht also die gesamte Zeitreihen-Vergangenheit ein, während diese schon bei der Zweischritt-Prognose völlig nutzlos ist. Dieser scheinbare "Widerspruch" ist durch die ACF und die PACF der MA(1)-Reihe zu erklären: Während z.B. die Autokorrelation zweiter Ordnung verschwindet, ist die partielle Autokorrelation zweiter Ordnung von Null verschieden. Die um zwei Takte zurückliegende Variable ist also nur dann prognostisch relevant, wenn auch die um einen Takt zurückliegende Variable beteiligt ist, also bei der Einschnitt-Prognose.

In Übereinstimmung mit den obigen Formeln liefert SPSS in unserem Beispiel für den ersten Zeitpunkt im Prognosebereich (Nr. 101) eine individuelle Einschritt-Prognose und für alle weiteren Zeitpunkte den geschätzten Zeitreihenmittelwert als konstante h -Schritt-Prognose:

	score	fit_2	err_2	lcl_2	ucl_2	sep_2
96	2182	2190,70044	-8,70044	2163,58802	2217,81286	10,32095
97	2205	2206,96470	-1,96470	2179,85228	2234,07712	10,32095
98	2216	2201,70373	14,29627	2174,59131	2228,81615	10,32095
99	2171	2189,00302	-18,00302	2161,89060	2216,11544	10,32095
100	2216	2214,23052	1,76948	2187,11811	2241,34294	10,32095
101	2195	2198,78713	-3,78713	2171,67471	2225,89955	10,32095
102	2210	2200,16919	9,83081	2165,94515	2234,39322	13,02814
103	2191	2200,16919	-9,16919	2165,94515	2234,39322	13,02814
104	2203	2200,16919	2,83081	2165,94515	2234,39322	13,02814
105	2205	2200,16919	4,83081	2165,94515	2234,39322	13,02814
106	2204	2200,16919	3,83081	2165,94515	2234,39322	13,02814
107	2202	2200,16919	1,83081	2165,94515	2234,39322	13,02814
108	2197	2200,16919	-3,16919	2165,94515	2234,39322	13,02814

Für die Varianz des Prognosefehlers im MA(1)-Modell erhalten wir folgenden Ausdruck (vgl. Schlittgen & Streitberg 1989, S. 367):

$$\text{Var}(X_{t+h} - \hat{X}_{t,h}) = \begin{cases} \text{Var}(E_t) & \text{für } h = 1 \\ \text{Var}(E_t)(1 + \theta_1^2) = \text{Var}(X_t) & \text{für } h > 1 \end{cases}$$

Der von SPSS für Einschritt-Prognosen berechnete Standardfehler (siehe Variable SEP_2) liegt mit 10,32095 aus nicht nachvollziehbaren Gründen etwas höher als der vom selben ARIMA-Lauf ausgegebenen Standardfehler für das weiße Rauschen (10,265823).¹

¹ SAS liefert exakt das erwartete Ergebnis:
 ARIMA Procedure
 Forecasts for variable SCORE

Obs	Forecast	Std Error	Lower 95%	Upper 95%
101	2198.7963	10.2656	2178.6761	2218.9165
102	2200.1692	13.0329	2174.6252	2225.7133
103	2200.1692	13.0329	2174.6252	2225.7133

5 Literatur

- Box, G.E.P. & Jenkins, G.M. (1976). *Time series analysis*. Forecasting and control. San Francisco: Holden-Day.
- Brockwell, P.J. & Davis, R.A. (1987). *Time series: Theory and methods*. New York: Springer.
- Hartung, J. (1989), *Statistik*. München: Oldenbourg.
- König, H. & Wolters, J. (1972). *Einführung in die Spektralanalyse ökonomischer Zeitreihen*. Meisenheim: Hain.
- Leiner, B. (1982). *Einführung in die Zeitreihenanalyse*. München: Oldenbourg.
- Pindyck, R.S. & Rubinfeld, D.L. (1990). *Econometric models and econometric forecasts*. Auckland: McGraw-Hill.
- Schlittgen, R. & Streitberg, B.H.J. (1989). *Zeitreihenanalyse*. München: Oldenbourg.
- SPSS Inc. (1988). *SPSS-X Trends*. Chicago, IL: SPSS.
- SPSS Inc. (1993b). *SPSS/PC+ Trends. Release 5.0*. Chicago, IL: SPSS.
- SPSS Inc. (1994). *SPSS Trends 6.1*. Chicago, IL: SPSS.

Ausführliche Literaturangaben zu speziellen Methoden der Zeitreihenanalyse finden sich z.B. im Referenzteil der SPSS-Handbücher zu SPSS-Trends, jeweils hinter den Erläuterungen des zugehörigen Trends-Kommandos. Die Trends-Handbücher sind in mehreren Exemplaren in der Lehrbuchsammlung der Universität Trier vorhanden.

6 Anhang

6.1 Trends-Systemparameter

Wie das Basispaket von SPSS verfügt auch das Zusatzmodul Trends über zahlreiche Systemparameter, über die das Verhalten der Software beeinflusst werden kann. Sie können mit dem Kommando **TSHOW** angezeigt und mit dem Kommando **TSET** verändert werden.

Nach dem SPSS-Start liefert **TSHOW** das folgende Protokoll der aktuellen Systemparameter-Einstellungen:

Time Series Parameters					
Parameter	Setting	Parameter	Setting	Parameter	Setting
-----	-----	-----	-----	-----	-----
ACFSE	IND	ID VAR	none	MXNEWVAR	60
CIN	95,0000	MISSING	EXCLUDE	NEWVAR	CURRENT
CNVERGE	,0010	MXPREDICT	1000	PRINT	DEFAULT
CONSTANT	YES	MXAUTO	16	TOLERANCE	,0001
		MXCROSS	7		

Es gibt zwar keine Dialogbox, die sich speziell mit Trends-Systemparametern befaßt, doch enthalten manche Dialogboxen zu Trends-Prozeduren Kontrollmöglichkeiten, bei deren Verwendung im Hintergrund **TSET**-Kommandos ausgeführt und somit Trends-Systemparameter gesetzt werden.

Der wichtige Trends-Systemparameter **NEWVAR** regelt z.B. die Behandlung der Variablen mit Residuen, prognostizierten Werte etc., die gemäß Abschnitt 2.5.2.2 von den Trends-Prozeduren **AREG**, **ARIMA**, **EXSMOOTH**, **SEASON** und **X11ARIMA** neu erzeugt werden. In den Dialogboxen zu den genannten Trends-Prozeduren ist über den Schalter **Speichern** eine Subdialogbox verfügbar, in der Sie die Behandlung neuer Variablen festlegen können (siehe Abschnitt 2.5.2.2). Mit dem folgenden **TSET**-Kommando wird für alle nachfolgenden Prozeduren festgelegt, daß keine neuen Variablen erzeugt werden sollen:

```
tset newvar=none.
```

7 Stichwortverzeichnis

A	ACF 10, 11, 30, 34 ACF-Kommando 13, 16, 27 ACF-Prozedur 36 Akaike Information Criterion 32 AR(1)-Prozeß 30 AR(p)- Prozeß 29 AR(p)-Prozeß 31 AREG-Kommando 15, 19, 42 ARIMA(p,d,q)-Prozeß 29 ARIMA-Kommando 14, 15, 19, 35, 42 ARIMA-Modell Prognosen 39 ARIMA-Modelle 23, 29 ARIMA-Modellierung 14 ARMA(p,q)-Prozeß 29 Autokorrelationsfunktion 10 Autoregressiver Prozeß 29	G	Gittersuche 25
B	Beispiel Defekte Produkte 6, 32 Lichtstärke 6, 14 Nullsummenroulett 5 Verkaufszahlen 6, 16, 23 Zahnpasta-Marktanteil 7 Beobachtungsbereich 18 Box-Ljung-Test 11, 32, 36	H	Historische Periode 18 Holt-Modell 24
C	CASEPLOT-Kommando 13 CCF-Kommando 14 Census I 14 Census II X11 14 CREATE-Kommando 20 CURVEFIT-Kommando 15, 19	I	Identifikation 29, 34 Irrfahrt 5, 9, 11
D	DATE-Kommando 17 Datumsvariablen 16 Defekte Produkte 6, 32 Differenzenbildung 11, 29, 30 Differenzenzeitreihe 10, 11 Durbin-Watson-Test 28	K	Kurvenanpassung 15
E	Einschritt-Prognose 22 Empirische Zeitreihe 7 Exponentielles Glätten 14, 22 EXSMOOTH-Kommando 14, 19, 25, 42	L	Lichtstärke 6, 14 Lineare Interpolation 20 Linearer Trend an dem Punkt 21
F	Fehlende Werte ersetzen 20 FIT-Kommando 16, 27	M	MA(q) - Prozeß 29 MA(q)-Prozeß 31 Modellbeurteilung 26 Modell-Beurteilung 32, 36 Modelldiagnose 16 Modell-Identifikation 29, 34 moving average - Prozeß 29
G	Gittersuche 25	N	NEWVAR-Systemparameter 42 NPLOT-Kommando 16 Nullsummenroulett 5
H	Historische Periode 18 Holt-Modell 24	P	PACF 10, 11, 30, 34 PACF-Kommando 13 Partielle Autokorrelationsfunktion 10 Plotten (P)ACF 13 CCF 14 Zeitreihe 13 PREDICT-Kommando 19, 25 Prognose 22, 23 Prognosebereich 18 Prognosen im ARIMA-Modell 39
I	Identifikation 29, 34 Irrfahrt 5, 9, 11	R	Random Walk 9, 11 Regressionsanalyse 15 RMV-Kommando 20
J	(Empty)	S	Saison 22

Saisonbereinigung	14
Schätzbereich.....	18, 32
Schwartz Bayesian Criterion	32
SEASON-Kommando.....	14, 42
Sequenzdiagramm	33
SPECTRA-Kommando.....	14
Spektralanalyse.....	14
Stationarität	8, 10, 33
Systemparameter	42

T

Theoretische Zeitreihe	7
Trend	22
TSET-Kommando	25, 35, 42
TSHOW-Kommando.....	19, 42
TSPLOT-Kommando	13, 24, 28, 33, 38

U

USE-Kommando	19, 28, 33, 38
--------------------	----------------

V

Validierungsbereich.....	18
Validierungsperiode.....	37
Verkaufszahlen	6, 16, 23

W

Weißes Rauschen.....	8
----------------------	---

X

X11ARIMA-Kommando	14, 42
-------------------------	--------

Z

Zahnpasta-Marktanteil.....	7
Zeitreihe	
empirische	7
theoretische	7
Zufallsexperiment	8