

### Berücksichtigung instabiler Varianzen in der Zeitreihenanalyse

Thome, Helmut

Veröffentlichungsversion / Published Version

Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:

GESIS - Leibniz-Institut für Sozialwissenschaften

#### Empfohlene Zitierung / Suggested Citation:

Thome, H. (1994). Berücksichtigung instabiler Varianzen in der Zeitreihenanalyse. *ZA-Information / Zentralarchiv für Empirische Sozialforschung*, 35, 82-109. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-201229>

#### Nutzungsbedingungen:

Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

#### Terms of use:

This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

## Berücksichtigung instabiler Varianzen in der Zeitreihenanalyse

von Helmut Thome<sup>1</sup>

### *Zusammenfassung*

*Die Modellierung von Zeitreihen - beispielsweise in Form der ARIMA-Modelle - stützt sich auf die Theorie stationärer stochastischer Prozesse. Die Stationaritätsvoraussetzung ist bei vielen sozialwissenschaftlich relevanten Zeitreihen aber nicht erfüllt. Unter anderem weisen sie häufig Varianzen auf, die entweder trend- oder zeitspezifisch schwanken. Der vorliegende Artikel erläutert ausführlich die Box/Cox-Transformation (zur Stabilisierung trendabhängiger Varianzen) und die von Tsay vorgeschlagene Adjustierung periodenspezifischer Varianzen. Außerdem wird kurz in den Ansatz der GARCH-Modelle eingeführt, einer allgemeinen Strategie zur Modellierung zeitspezifischer Varianzen.*

### *Abstract*

*The modelling of time series data is based on the theory of stationary stochastic processes. Many of the sociologically relevant time series, however, are non-stationary, not only in the mean but also in the variances which may depend on time or level. The present article discusses the **Box-Cox** transformation (as a means to stabilize trend specific variances) and a strategy proposed by Tsay to make appropriate adjustments for abrupt changes of variance. There is also a brief introduction into GARCH-modelling, a more generalized approach to deal with time specific heteroscedasticity.*

### **1. Einführende Bemerkungen**

Statistische Modelle für die Analyse von Zeitreihen setzen in der Regel voraus, daß die erzeugenden stochastischen Prozesse mindestens schwach stationär, die zeitindizierten Erwartungswerte und Varianzen also konstant und die Autokovarianzen (bzw. Autokorrelationen) unabhängig von der historischen Zeit sind. Die Varianzen und Mittelwerte sozialwissenschaftlich relevanter Zeitreihen sind aber häufig nicht stabil; es zeigen sich Trendverläufe sowie Varianzen, die zeitspezifisch oder in Abhängigkeit vom Trend schwanken. Diese Thematik, vor allem das Problem instabiler Varianzen, wird in Ein-

---

<sup>1</sup> Dr. **Helmut Thome** ist Professor für Soziologie an der Martin-Luther-Universität Halle-Wittenberg. Postanschrift: Institut für Soziologie, 06099 Halle (Saale).

führungstexten meist nur cursorisch behandelt und in der Forschungspraxis nur selten ausreichend bedacht. Differenzenbildung (zur "Herstellung" von Mittelwert-Stationarität) und logarithmische Transformationen (zur Stabilisierung der Varianzen) werden allzu mechanisch eingesetzt; häufig wird die Instabilität der Varianzen auch schlicht übergangen. Der vorliegende Artikel ist ein Versuch, anhand der Fachliteratur Problem- und Lösungskonzepte so darzustellen, daß sie sozialwissenschaftliche Forschungspraxis anleiten können. Dabei beschränke ich mich auf das Problem instabiler Varianzen im Kontext univariater Analysen, die dem *Box/Jenkins-Ansatz* folgen. Die hierzu vorgestellten Konzepte und Techniken sind aber größtenteils auch für die multivariate Zeitreihenanalyse bzw. die dynamische Regressionsanalyse relevant: Kausalitätstests und die empirische Identifikation von *Box/Jenkins*-Transferfunktionsmodellen setzen stationäre (bzw. stationär "gemachte") Zeitreihen voraus; und heteroskedastische Residuen sind bei Regressionsanalysen mit Zeitreihendaten nicht minder problematisch als bei Regressionsanalysen mit Querschnittsdaten (s. z. B. *Kmenta* 1986). Zudem ist die univariate Analyse von Zeitreihen unmittelbar relevant für die Evaluationsforschung: Der Einfluß zielgerichteter Maßnahmen (etwa zur Verminderung des Drogenkonsums) oder ungeplanter Ereignisse (z. B. Naturkatastrophen) auf die Entwicklung bestimmter sozialer Indikatoren kann nur dann zuverlässig geschätzt bzw. getestet werden, wenn für die entsprechenden Zeitreihen adäquate statistische Modelle unter Berücksichtigung eventueller Varianzheterogenität gefunden worden sind.

Wie sich instabile (heteroskedastische) Varianzen einer Zeitreihe  $X_t$  auf die Identifikation eines Prozeßmodells, die Schätzung seiner Parameter und die Qualität der daraus abgeleiteten Prognosen (das wichtigste Kriterium in der univariaten Analyse) auswirken, läßt sich nicht allgemein und eindeutig angeben. Folglich ist auch strittig, ob Homogenisierungsstrategien - wie z. B. die Transformation der Daten mittels einer Potenzfunktion - überhaupt notwendig oder wünschenswert sind. Sucht man als Praktiker Rat in der Fachliteratur, stößt man auf verwirrende Informationen und Ratschläge. *Vandaele* (1983) stellt z. B. fest, statistische Schlußfolgerungen würden durch instabile Varianzen mit Sicherheit beeinträchtigt. In einer oft zitierten Studie über die Auswirkungen varianzstabilisierender Datentransformationen auf die Prognosegüte kommen *Nelson und Granger* (1979: 66) dagegen zu dem Schluß: "The results suggest that, in practice there is very little to be gained by using transformed data rather than untransformed ...". *Hopwood, McKeown und Newbold* (1984), die eine Vielzahl von Zeitreihen über "corporate earnings" analysiert haben, kommen zu einem anderen Resümee: "... it became clear, in the early stages of our study, that, for many of these series, there was strong evidence of the desirability of a data transformation to induce homogeneity of model error variance" (S. 60; vgl. *Lenk und Tsai* 1990: 221).

Offensichtlich gibt es kein sicheres Augenschein-Kriterium, das einem Praktiker erlaubt, allein vom Plot einer Zeitreihe abzulesen, ob Art und Ausmaß einer erkennbaren Heteroskedastizität es unbedingt erforderlich machen, die Varianzen zu stabilisieren bzw. explizit zu modellieren, um ein bestimmtes Untersuchungsziel (z. B. die Optimierung der Prognosen) zu erreichen. In einer solchen Situation empfiehlt es sich nicht, die Sache auf sich beruhen zu lassen, sondern eine Strategie zu wählen, die möglichst einfach zu handhaben ist und sich nach bisherigen Erfahrungen als relativ effizient und nicht oder nur selten als schädlich erwiesen hat. Unter diesem Gesichtspunkt soll hier zunächst die meistangewandte, auch von **Box und Jenkins** (1973; 1976) empfohlene Strategie zur Varianzstabilisierung, nämlich die Transformation der Daten mit Hilfe einer bestimmten Potenzfunktion, erörtert werden. Sie ist aber nur auf niveau- bzw. trendabhängige Varianzen zugeschnitten. Strategien zur Berücksichtigung zeitspezifischer Varianzen werden im 3. Abschnitt vorgestellt.

## 2. Trendabhängige Varianzen: Box/Cox-Transformationen

Das Varianzmodell, um das es in diesem Abschnitt geht, läßt sich formal wie folgt darstellen:

$$(1) \quad \sigma_t = c \mu_t^\gamma$$

Dabei bezeichnet  $\sigma_t$  die Standardabweichung einer Zeitreihe  $X_t$  ( $t=1,2,\dots,n$ ),  $c$  eine Proportionalitätskonstante,  $\mu_t$  den aktuellen Trendwert der Zeitreihe und  $\gamma \in \mathbb{R}$  einen noch unbekanntem Parameter. Wenn sich die Varianzen im Sinne von Gleichung (1) trendabhängig entwickeln, lassen sie sich stabilisieren, indem die Daten mit einer geeigneten Potenzfunktion transformiert werden. Die von **Box und Cox** (1964: 214) hierzu vorgeschlagene Potenzfunktion lautet:<sup>2</sup>

$$(2) \quad Z_t = \begin{cases} \frac{(X_t + c)^\lambda - 1}{\lambda} & \text{für } \lambda \neq 0 \\ \ln(X_t + c) & \text{für } \lambda = 0 \end{cases}$$

<sup>2</sup> Ober Extensionen und Modifikationen dieser Potenzfunktion siehe **Cook und Weisberg** (1982: 60 f.) und insbesondere **Emerson und Stoto** (1983: 98 - 104). **Box und Cox** haben diese Potenzfunktion nicht spezifisch zur Varianzstabilisierung in der univariaten Zeitreihenanalyse vorgeschlagen, sondern zur Transformation heteroskedastischer Response- bzw. Fehlervariablen, um drei Zwecke gleichzeitig zu erreichen: eine lineare Beziehung, normalverteilte und homoskedastische Fehler. **Box und Jenkins** (1973; 1976) - siehe auch **Jenkins** (1979) - haben diesen Ansatz dann in die Zeitreihenanalyse übernommen.

Die Additionskonstante  $c$  soll lediglich sicherstellen, daß die zu transformierenden Zeitreihenwerte alle positiv sind, damit eine Potenzbildung bzw. Logarithmierung in jedem Fall sinnvoll bzw. möglich ist (Mohr 1983: 178)<sup>3</sup>. Nelson und Granger (1979: 57) empfehlen, ein möglichst kleines  $c$  zu wählen (s. auch Mosteller und Tukey 1977: 83).

In der Praxis begnügt man sich in der Regel mit einer einfacheren Transformation

$$(3) \quad Z_t = \begin{cases} (X_t + c)^\lambda & \text{für } \lambda \neq 0 \\ \ln(X_t + c) & \text{für } \lambda = 0, \end{cases}$$

die Box und Cox (1964: 214 f.) als "äquivalent" zu (1) bezeichnen; Form (1) sei "slightly preferable for theoretical analysis because it is continuous at  $\lambda = 0$ ".<sup>4</sup>

Wieso Potenzfunktionen der Form (2) oder (3) geeignet sind, trendabhängige Varianzen zu stabilisieren, erläutern u. a. Mills (1990: 6 f.), Wei (1990: S. 82 f.) und Abraham und Ledolter (1983). Es läßt sich z. B. zeigen, daß (a) im Falle einer proportionalen Beziehung zwischen der Varianz und dem Trendniveau,  $\text{Var}(X_t) = c\mu_t$ , eine Quadrat-Wurzeltransformation ( $\lambda = 0.5$ ), (b) im Falle einer proportionalen Beziehung zwischen der Standardabweichung und dem Trendniveau,  $\sigma_t = c\mu_t$  bzw.  $\text{Var}(X_t) = c^2\mu_t^2$ , eine logarithmische Transformation ( $\lambda = 0$ ) und (c) im Falle einer proportionalen Beziehung zwischen der Standardabweichung und dem quadrierten Trendniveau,  $\sigma_t = c\mu_t^2$ , eine reziproke Transformation ( $\lambda = -1$ ) die Varianzen stabilisieren.<sup>5</sup> Der Transformationsparameter  $\lambda$  ergibt sich somit aus Gleichung (1) mit  $\lambda = 1-\gamma$ :

$$(4) \quad \sigma_t = c \cdot \mu_t^{1-\lambda} = c \cdot \mu_t^\gamma$$

Falls die Varianz von  $X_t$  bereits konstant ist:  $\sigma_t = c$ , erhalten wir aus Gleichung (4) die Information, daß  $\lambda = 1$  und somit keine Transformation nötig ist:

$$(5) \quad \sigma_t = c = c \cdot \mu_t^{1-\lambda}, \quad \lambda = 1$$

Die Frage ist nun, wie ein geeigneter Transformationsparameter  $\hat{\lambda}$  (einschließlich  $\hat{\lambda} = 1$ )

<sup>3</sup> Siehe dort auch eine Erläuterung zur Einbeziehung der Logarithmustransformation in die Potenzfunktion.

<sup>4</sup> Dieser Kontinuitätsgesichtspunkt wird erläutert in Emerson und Stoto (1983: 98 - 103); s. auch Johnston (1984: 61 ff.).

<sup>5</sup> Wie bereits erwähnt, hofft man außerdem, mittels Potenzfunktionen die Verteilung der Variablen näher an die Normalverteilung heranzubringen. Die Annahme der Normalverteilung wird nicht nur bei der Ableitung der Maximum-Likelihood-Schätzfunktionen benötigt, sondern soll auch sicherzustellen, "daß der Minimum-Mean-Square-Error Prognoseschätzer von linearer Bauart ist" (Mohr 1980: 167).

aus den beobachteten Daten ermittelt werden kann. Grundsätzlich lassen sich zwei Verfahrensgruppen unterscheiden: (a) Schätzverfahren, mit denen  $\hat{\lambda}$  und die übrigen Parameter eines ARIMA-Modells simultan ermittelt werden,<sup>6</sup> (b) sog. Vorschalttechniken, mit denen  $\hat{\lambda}$  zunächst allein bestimmt wird, bevor das ARIMA-Modell anhand der transformierten Daten identifiziert und geschätzt wird (s. *Mohr* 1983). Unter theoretischen Gesichtspunkten ist die simultane Schätzung vorzuziehen,<sup>7</sup> weil andernfalls bei der Schätzung der Modellparameter bzw. ihrer Standardfehler die zusätzliche Variabilität, die mit der Schätzung von  $\lambda$  gegeben ist, nicht berücksichtigt werden kann (s. *Carroll und Ruppert* 1988: 93). Ein Standardprogramm, das die simultane Schätzung von Transformations- und Modellparametern in der univariaten Zeitreihenanalyse erlaubt, ist mir jedoch nicht bekannt.<sup>8</sup> Bei Untersuchungen von *Mohr* (1983), in denen er einige alternative Schätzmethoden an simulierten und empirischen Reihen erprobt, schneidet jedoch eine Vorschalttechnik recht gut ab, die auch *Jenkins* (1979: 95 - 97) für die praktische Zeitreihenanalyse empfohlen hat. Sie soll hier etwas näher erläutert werden.

Ausgangspunkt ist die Gleichung (4). Zunächst ist festzustellen, in welcher Weise die Varianz vom Trend abhängt. Dazu kann man sich eines einfachen Streudiagramms bedienen, in dem die geschätzten Standardabweichungen  $\hat{\sigma}_t = s_t$  (oder ein anderes Dispersionsmaß) auf der Ordinate und die geschätzten Trendwerte  $\hat{\mu}_t = \bar{x}_t$  als Abszissenwerte eingetragen sind. Da sich aus jeweils einer Realisierung  $X_t = x_t$  für jeden Beobachtungzeitpunkt  $t$  keine Schätzgrößen für  $\sigma_t$  und  $\mu_t$  ermitteln lassen, hat *Jenkins* (1979: 95 -

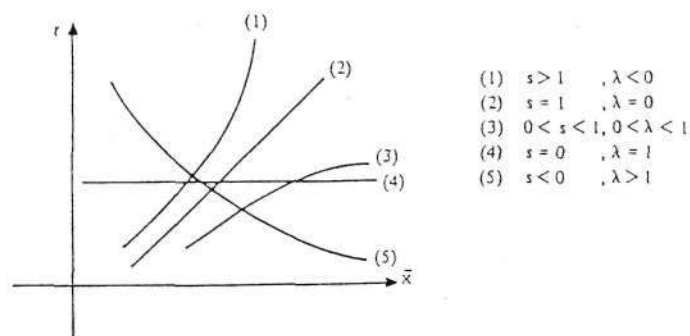
<sup>6</sup> *Box und Cox* (1964) geben die Likelihoodfunktion als Zielkriterium vor und zeigen, daß die ML-Schätzung darauf hinausläuft, die Summe der Fehlerquadrate (Modellresiduen) zu minimieren. *Mohr* (1983: 180) weist darauf hin, daß der ML-Schätzer für  $\lambda$  nicht unbedingt derjenige ist, der zu den besten Prognoseergebnissen führt (vgl. *Nelson und Granger* 1979: 67): "Neben der Likelihoodfunktion könnte man auch das Akaike-Informationskriterium, das Bestimmtheitsmaß oder spezielle Prognoseeigenschaften sowie die Güte der Restgrößen als Maßstab verwenden. Vom praktischen Standpunkt aus betrachtet sind diese Transformationen nur sinnvoll, wenn dadurch die Prognoseergebnisse verbessert werden können" (ebd.).

<sup>7</sup> Dies setzt voraus, daß bereits ein (vorläufiges) Modell identifiziert wurde. Somit ergibt sich die Frage, ob eine fehlende oder falsche Datentransformation die Autokorrelationsfunktion soweit beeinträchtigt, daß das Modell nicht korrekt identifiziert werden kann. *Anderson* (1976: 287) meint hierzu, die Struktur der Autokorrelationen sei "not very sensitive to the actual choice of  $\lambda$  ... Thus one can identify the model using a rough transformation, or sometimes no transform at all; and then, at the estimation stage, initiate a search to find simultaneously the best parameter estimates and the necessary transformation". Vorsichtiger urteilt *Mohr* (1980: 171): "Die kritische Stelle dieses Ansatzes ist die Unabhängigkeit des Modelltyps von  $X$ . Hierzu lassen sich gegensätzliche Argumente aufführen. Einerseits ist schon oft festgestellt worden, daß sich innerhalb eines geeigneten  $\lambda$ -Bereichs die Kenngrößen (z. B. die Autokorrelationen, H. T.) nur wenig ändern und daher derselbe Modelltyp gewählt werden kann. Andererseits sind - wengleich weniger häufig - gegenteilige Feststellungen gemacht worden" (vgl. hierzu auch *Mohr* 1983: 179, 183; *Granger und Newbold* 1976: 192, 195). Während *O. D. Anderson* (1979: 288) eine grobe Spezifikation des  $\lambda$ -Parameters für ausreichend hält, warnt *Poirier* (1980: 279) davor, eben diesen Parameter nicht exakt zu bestimmen.

<sup>8</sup> Im Kontext der Regressionsanalyse ermöglicht dies LIMDEP.

97) vorgeschlagen, die nach  $t = 1, 2, \dots, k$  geordneten Zeitreihenwerte in  $k$  gleich große Gruppen einzuteilen<sup>9</sup> und für jede Gruppe das arithmetische Mittel und das Dispersionsmaß zu berechnen. Die Schnittpunkte ihrer Koordinaten bilden einen Punkteschwarm, aus dessen Lage der Transformationsparameter  $X$  in ungefährer Größe abzulesen ist (siehe Abb. 1).

**Abbildung 1:** Charakteristische Fälle der Beziehung zwischen Streuung und Mittelwert  
(Quelle: *Mohr* 1980: 169)



Falls die Punktwolke eine *lineare* Beziehung zwischen  $s$ , und  $\bar{x}$ , repräsentiert, ist  $\gamma = 1$  bzw.  $\lambda = 0$  und damit eine logarithmische Transformation angezeigt. Falls die Punktwolke parallel zur Abszisse verläuft, ist die Varianz nicht vom arithmetischen Mittel abhängig, d. h.  $\lambda = 1$  (keine Transformation). Nicht-lineare Verlaufsformen weisen auf Transformationsparameter ungleich 0 und ungleich 1 hin (s. Abb. 1). Es empfiehlt sich, die transformierte Reihe nochmals mit dem "spread versus level plot" (SLP) zu überprüfen. Wenn nach einer ersten Transformation die Koordinatenschnittpunkte im SLP um eine negative Gerade bzw. Kurve streuen, muß  $\hat{\lambda}$  gegenüber der ersten Wahl erhöht werden. Wenn die Punkte des Streudiagramms nach einer ersten Transformation um eine

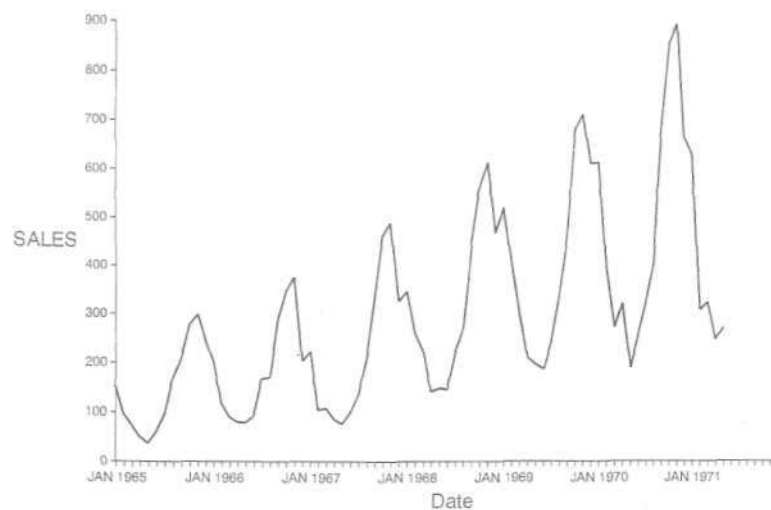
<sup>9</sup> Bei saisonalen Zeitreihen mit Monatsdaten bietet sich an,  $k$  so zu wählen, daß  $n/k = m = 12$  ist. Falls  $n$  durch  $k$  nicht ohne Rest teilbar ist, schließt man entsprechend viele Fälle am Anfang oder am Ende der Reihe aus der Berechnung aus. Zur Wahl von  $m$  siehe weitere Überlegungen unten. Bei saisonalen Zeitreihen mit Monatsdaten kann auch die Beziehung zwischen dem monatspezifischen Zuwachs (gemessen über den gesamten Beobachtungszeitraum) und den Mittelwerten der 12 Monate untersucht werden (s. *Box und Jenkins* 1973: 337 f.; kritisch hierzu *Chatfield und Prothero* 1973 b)

positive Gerade streuen, muß  $\hat{\lambda}$  kleiner werden.<sup>10</sup> **Mohr** (1983: 186) hat vorgeschlagen, dieses Verfahren zu ergänzen, indem Gleichung (4) durch Logarithmierung linearisiert und der Parameter  $\gamma$  als Steigungskoeffizient nach der Methode der kleinsten Quadrate geschätzt wird:"

$$(6) \quad \ln(\hat{\mu}_t) = \ln(c) + \gamma \ln(\hat{\mu}_t)$$

Ich möchte diese Vorgehensweise anhand einer Zeitreihe veranschaulichen, die in der Literatur intensiv diskutiert worden ist. Es handelt sich um die Verkaufszahlen der "Firma X" (siehe Abb. 2), die **Chatfield und Prothero** (1973 a) in einer Attacke gegen den **Box/Jenkins-Ansatz** präsentiert haben.

**Abbildung 2:** *Chatfields*, "Sales"-Daten



**Chatfield und Prothero** haben diese Zeitreihe logarithmiert und für die logarithmierten Werte ein ARIMA-Modell geschätzt, das sehr unbefriedigende Prognosen lieferte: die Saisonamplituden wurden in absurdem Maße überdehnt. In einer nachfolgenden Debatte sind die Autoren u. a. wegen ihrer Entscheidung für die logarithmische Transformation kritisiert worden (so z. B. von **Wilson** 1973: 315). In einer Antwort auf **Chatfield und**

<sup>10</sup> Siehe das Anwendungsbeispiel mit Arbeitslosendaten in **Thome** (1992).

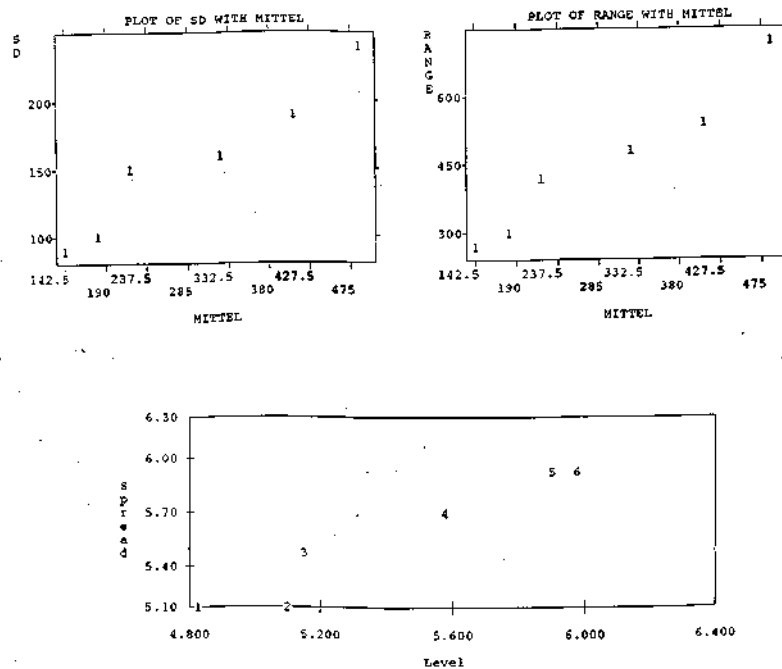
<sup>11</sup> Dieses Verfahren funktioniert allerdings nur, wenn keine Ausreißer vorkommen oder das Regressionsverfahren "robustifiziert" wird. Deshalb ist es auf jeden Fall sinnvoll, sich zuvor das Streudiagramm anzusehen.



*Prothero* (1973 a) verwendeten *Box und Jenkins* (1973) den Transformationsparameter  $\hat{\lambda} = 0.25$  und erzielten damit deutlich verbesserte Prognosen.<sup>12</sup> Verkaufszahlen eines Wirtschaftsunternehmens gehören sicher nicht zu den typischen Datensätzen, die ein Soziologe untersucht. Ihre strukturellen Merkmale (wie instabile Varianzen, Trend- u. Saisonkomponente) finden sich aber auch in Zeitreihen, die dem Soziologen bzw. dem Sozial- oder Wirtschaftsgeschichtler geläufig sind. Für eine didaktisch orientierte Reanalyse bieten sie hier den Vorteil, die eigenen Ergebnisse mit denen zu vergleichen, die in der Fachliteratur veröffentlicht sind.

Die Abbildungen 3a bis 3c liefern drei Streudiagramme, mit denen die Instabilität der Varianzen in Abhängigkeit vom Niveau der Reihe überprüft werden kann.<sup>13</sup>

**Abbildung 3:** "Spread versus level Plots" für nicht-transformierte "Sales"-Daten



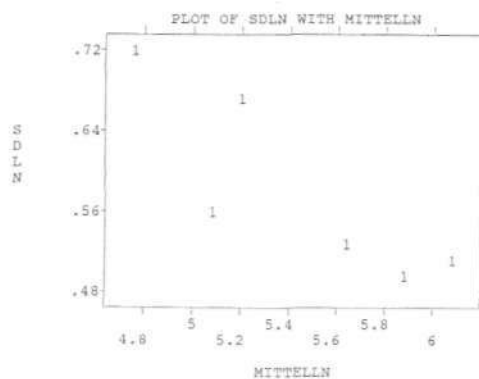
<sup>12</sup> Zum Problem der Rücktransformation der Daten in die ursprüngliche Metrik siehe unten.

<sup>13</sup> Streuung und arithmetisches Mittel wurden jeweils über 12 Monatsintervalle berechnet. Die letzten 5 Werte (Jan. bis Mai 1971) wurden eliminiert, um den Stichprobenumfang für die Dispersions- und Niveauschätzer konstant zu halten.

Abbildung 3a zeigt das Verhältnis der Standardabweichungen zu den arithmetischen Mitteln der Jahre 1965 bis 1970; in Abbildung 3b ist die Standardabweichung durch die Spannweite ("range") ersetzt; Abbildung 3c führt als Dispersionsmaß den logarithmierten Interquartilsabstand ein und ersetzt das arithmetische Mittel durch den logarithmierten Median (dies ist der Standard "spread versus level plot" in der SPSS-Routine EXAMINE). Die Schnittpunkte der Koordinaten in allen drei Diagrammen scheinen durch eine von links nach rechts aufsteigende Gerade gut repräsentierbar. Auf den ersten Blick unterstützen sie somit die Entscheidung von *Chatfield und Prothero* (1973a) für eine Logarithmus-Transformation. Die von *Mohr* (1983) vorgeschlagene Regression nach Gleichung (5) liefert jedoch die Steigungskoeffizienten  $\hat{\gamma} = 0.75$  (wenn die Standardabweichung als Dispersionsmaß herangezogen wird) und  $\hat{\gamma} = 0.76$  (bei der Spannweite) mit Standardabweichungen von 0.086 bzw. 0.087. Dies würde auf einen Transformationsparameter  $\hat{\lambda} \approx 0.25$  (bzw. 0.26) schließen lassen. Eine Regressionsschätzung mit den logarithmierten Interquartilsabständen und den logarithmierten Medianen (Abb. 3c) bestätigt dieses Ergebnis mit  $\hat{\gamma} = 0.73$ .

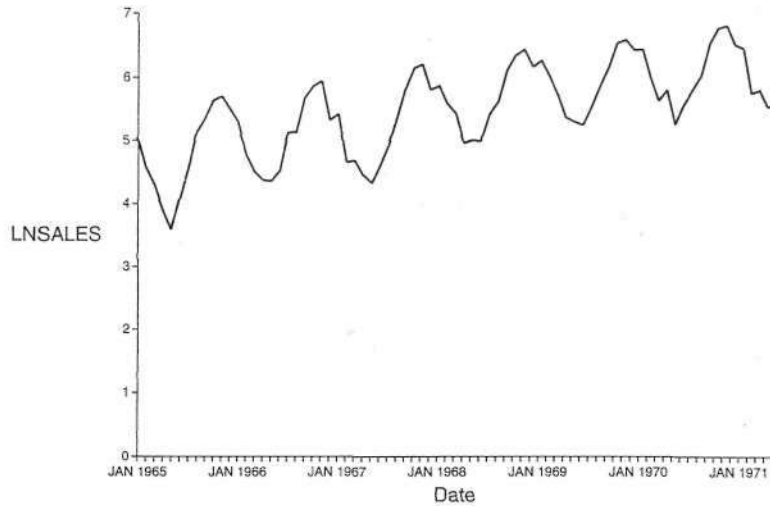
Daß der erste visuelle Eindruck, den das Streudiagramm vermittelt, täuschen kann (vor allem dann, wenn nur wenige Werte vorliegen), zeigen die entsprechenden Diagramme für die *logarithmierten* Sales-Daten. Die Koordinatenschnittpunkte im SLP zeigen nun eine abfallende Tendenz von links oben nach rechts unten (siehe Abb. 4).

**Abbildung 4:** SLP der logarithmierten "Sales"-Daten



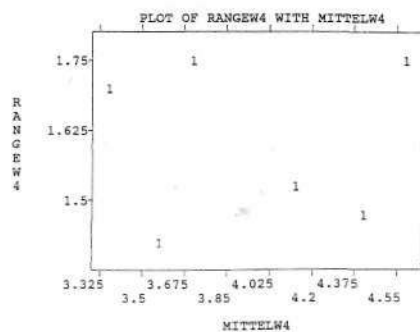
Das heißt, die Daten wurden mit dem Logarithmus zu stark transformiert; *Chatfield und Prothero* haben sich mit ihrem  $\hat{\lambda} = 0$  zu weit von  $\hat{\lambda} = 1$  (keine Transformation) wegbegeben. Das beweist auch der Verlauf der logarithmierten Verkaufszahlen in Abb. 5: die Spannweite nimmt mit steigendem Niveau leicht ab.

Abbildung 5: Logarithmierte "Sales"-Daten



Folglich muß der Transformationsparameter nun wieder erhöht werden. Läßt man die oben erwähnten Regressionsergebnisse außer acht, so bietet sich dafür zunächst ein mittlerer Wert zwischen 0 und 1 an, also die Wurzeltransformation ( $\hat{\lambda} = 0.5$ ). Alle drei "spread level plots" (analog zu den Abbildungen 3a bis 3c konstruiert, hier aber nicht gezeigt) belegen, daß diese Transformation zu schwach ist; die Streuung der wurzeltransformierten Sales-Daten nimmt mit steigendem Niveau der Reihe zu. Der "wahre" Transformationsparameter muß also zwischen  $\hat{\lambda} = 0.5$  und  $\hat{\lambda} = 0.0$  liegen. Die von *Box und Jenkins* (1973) gewählte Transformation mit  $\hat{\lambda} = 0.25$  bringt in den Streudiagrammen (von denen wir in Abb. 6 nur eines zeigen) befriedigende Ergebnisse: Eine ausgeprägte Aufwärts- oder Abwärtstendenz der Koordinatenschnittpunkte ist nun nicht mehr zu erkennen.

Abbildung 6: SLP der zweifach wurzeltransformierten "Sales"-Daten



Auch der Plot der zweifach wurzeltransformierten Sales-Daten (hier nicht gezeigt) läßt keine Instabilität der Varianzen mehr erkennen.

Graphische Darstellungen wirken in der Regel besonders überzeugend. Die Suche nach dem wahren Transformationsparameter kann jedoch, wie eben vorgeführt, ziemlich mühsam werden, wenn man sich allein auf die subjektive Interpretation der Streudiagramme verläßt. Für die Praxis empfiehlt sich ein Verfahren, das mit dem SLP beginnt, die Wahl eines ersten Schätzers für  $\lambda$  aber sogleich an dem Ergebnis der Regressionsanalyse nach Gleichung (5) ausrichtet.<sup>14</sup> Obwohl die zeitlich geordneten Beobachtungswerte in der Regel nicht unabhängig voneinander sind, kann man auch in diesem Kontext den Standard-Tests auf Varianzhomogenität nützliche Hinweise entnehmen. So weist der in SPSS (EXAMINE) verfügbare *Levene-Test* (s. *Hartung et. al.* 1985: 617) bei den untransformierten Verkaufszahlen ein Fehlerrisiko von  $\alpha = 0.0007$ , bei den wurzeltransformierten Daten ein  $\alpha = 0.45$  und bei den zweifach wurzeltransformierten Daten ein  $\alpha = 0.84$  für die Zurückweisung der Nullhypothese ("Die Varianzen sind homogen") aus.

Die Menge  $m = n/k$  der Zeitpunkte, die zur Berechnung der Dispersions- und Lokationsmaße zu aggregieren sind, stellt ein gewisses Problem dar. *Guerrero* (1993:45) weist darauf hin, daß die Schätzgröße  $\hat{\lambda}$  "sometimes heavily" von  $m$  abhängt. Im Falle saisonaler Zeitreihen ist  $m$  durch die Saisonperiode vorgegeben: 4 Zeitpunkte bei Quartalsdaten, 12 Zeitpunkte bei Monatsdaten. Für nicht-saisonale Zeitreihen empfiehlt *Guerrero* ein  $m = 2$ , um den durch die Gruppierung entstehenden Informationsverlust so gering wie möglich zu halten (ebd., S. 40).

Kehren wir zu unserem Analysebeispiel zurück. Nach der Schätzung von  $\lambda$  wird die Zeitreihe  $X_t$  gemäß Gleichung (3) zu  $Z_t$  (mit  $c = 0$ ) transformiert. Anhand der transformierten Daten wird ein (saisonales) ARIMA-Modell mit den üblichen Verfahren identifiziert und geschätzt. Ich erspare mir eine erneute Identifikationsanalyse und greife auf das von *Chatfield und Prothero* (1973 a) vorgeschlagene "Modell A":

$(1 - \phi_1 B)Z_t(1 - B)(1 - B^{12}) = (1 - \theta_{12} B^{12})a_t$  zurück, das später auch von *Box und Jenkins* (1973) und weiteren Autoren (so z. B. *Guerrero* 1993: 45) übernommen wurde.<sup>15</sup> Während *Box und Jenkins* (1973) die entsprechenden Modellparameter für die zweifach wurzeltransformierte Reihe mit  $\hat{\phi}_1 = -0.5$  und  $\hat{\theta}_{12} = 0.8$  auf der Basis aller Beobach-

<sup>14</sup> Wie in der Regressionsanalyse allgemein erforderlich, ist dabei die Ausreißer-Problematik zu beachten. - In der SPSS-Routine EXAMINE liefert die Option SPREADLEVEL den Steigungskoeffizienten automatisch mit dem entsprechenden Plot.

<sup>15</sup> Die Notation entspricht der in der Box/Jenkins-Literatur üblichen Symbolik, wobei  $B$  den Verschiebe- oder "Lag"-Operator angibt:  $(1 - B^m)Z_t = Z_t - Z_{t-m}$   
 $(1 - B^m)Z_t = Z_t - Z_{t-m}$

tungswerte schätzen<sup>16</sup> und die Güte der Prognosen zur Basis Mai 1971 beurteilen, folge ich hier *Guerrero* (1993) u. a., die eine Schätzperiode bis (einschließlich) Mai 1970 festlegen und die dann noch folgenden 12 Beobachtungswerte zur "Validierung" einer 12-Schritt-Prognose verwenden. Neben dem natürlichen Logarithmus und der zweifachen Wurzelbildung verwende ich als weiteren Transformationsparameter  $\hat{\lambda} = 0.33$ , der verschiedentlich als Ergebnis einer exakten (simultanen) ML-Schätzung genannt worden ist (s. z. B. *Wilson* 1973, *Guerrero* 1993:45). Für die unterschiedlich transformierten Reihen und die Originaldaten wurden die in Tabelle 1 ausgewiesenen Parameter (mit SPSS-PC) geschätzt (Standardfehler in Klammern).

**Tabelle 1:** Parameterschätzer und RMS bei unterschiedlichen *Box/Cox-Transformationen*

	$\lambda = 1$	$\lambda = 0$	$\lambda = 0.25$	$\lambda = 0.33$
$\varphi_1$	-.496	-.489	-.52	-.52
$\Theta_{12}$	-.183 <sup>a)</sup>	.462	.61	.57
<b>RMS</b>	<b>82.82</b>	<b>141.65</b>	<b>81.51</b>	<b>58.54</b>
RMS nach Korrektur	-	187.31	94.44	63.04
a) nicht signifikant ( $p > 0.10$ )				

Gemäß dem Kriterium, den "Mean Square Forecast Error" (MSFE) zu minimieren, werden die bedingten Erwartungswerte  $\mu_j = E(Z_{n+j} | Z_n, Z_{n-1}, \dots)$ ,  $j = 1, 2, \dots$ , als Prognosewerte eingesetzt. Der Index "n" bezeichnet den Prognoseursprung, hier Mai 1970. Die bedingten Erwartungswerte werden zunächst für die transformierte Reihe  $Z_t$  berechnet und müssen durch die inverse Operation

$$(7) \quad \mu_j(x) = [\mu_j(z)]^{\frac{1}{\lambda}}$$

in die Originalmetrik der Reihe  $X_t$  zurücktransformiert werden. Da die ursprüngliche *Box/Cox*-Transformation die Verteilung der  $Z_t$  symmetrischer gestaltet als die Verteilung der  $X_t$ , wird bei der Rücktransformation aus dem arithmetischen Mittel der Median der entsprechenden Verteilung. Um die mit dem Erwartungswert verbundenen Optimalitäts-

<sup>16</sup> Erheblich davon abweichende Theta-Schätzer liefern sowohl SCA ( $\hat{\Theta}=0.986$ ) als auch SPSS-PC-TREND ( $\hat{\Theta}=0.973$ ). Während SCA dafür trotz der Nähe zur Stationaritätsgrenze einen sinnvollen Schätzer für den Standardfehler ausweist (0.093), liefert SPSS-PCATREND in dieser Extremsituation einen sinnlosen Wert von 2.18.

eigenschaften bezüglich des minimalen durchschnittlichen Fehlerquadrats (MMSFE) wieder zurückzugewinnen, sind verschiedene Korrekturfaktoren vorgeschlagen worden.<sup>17</sup> Der neueste Vorschlag für einen "approximate debiasing factor" stammt von **Guerrero** (1993: 41). Ich zitiere ihn hier nur in der Version, die auf die einfache Potenzfunktion der Gleichung (3) abgestellt ist<sup>18</sup>

$$\begin{aligned}
 E(X_{n,j}) &= \mu_j(x) \cdot C_\lambda(j) \quad j = 1, 2, \dots \\
 (8) \quad C_\lambda(j) &= \left[ 0.5 + 0.5 \left( 1 + 2(\lambda^{-1} - 1) \frac{\sigma_j^2}{\mu_j(z)^2} \right)^{0.5} \right]^{\frac{1}{\lambda}}, \quad \lambda \neq 0 \\
 &= \exp\left(\frac{\sigma_j^2}{2}\right), \quad \lambda = 0
 \end{aligned}$$

Dabei bezeichnet  $\sigma_j^2$  die Varianz des Prognosefehlers, der in jedem Standard-Programmausdruck mitgeliefert wird. Die Konfidenzintervalle können in der gleichen Weise korrigiert werden, indem man ihre Enden jeweils mit  $C_\lambda(j)$  multipliziert. Die Verzerrungskorrektur führt nicht dazu, daß bei einer geringen Zahl prognostizierter Werte die Prognosefehler unbedingt geringer ausfallen als dies ohne Verzerrungskorrektur der Fall wäre.<sup>19</sup> Tabelle 1 (oben) zeigt für die Validierungsperiode Juni 1970 bis Mai 1971 die Wurzel (RMS) des MSFE bei unterschiedlichen Transformationen sowohl mit als auch ohne Verzerrungskorrektur. (Bei der Berechnung wurde ein Ausreißerwert - Febr. 1971 - eliminiert.)

Der geringste Prognosefehler (gemessen am RMS-Wert) ist bei einer Transformation mit dem ML-Schätzer  $\hat{\lambda} = 0.33$  zu verzeichnen, das mit Abstand schlechteste Ergebnis resultiert aus der falschen Logarithmus-Transformation. Überraschend gut mit RMS = 82.8 fallen die Prognosen mit den untransformierten Daten aus. Falls man den Ausreißerwert nicht eliminiert und den RMS für alle 12 Monate der Validierungsperiode berechnet, übertrifft die Prognosegüte, die mit den untransformierten Daten erzielt wird, die Qualität der Prognosen mit den zweifach wurzeltransformierten Daten.

<sup>17</sup> Bei der "naiven" Rücktransformation ohne Korrektur wird nicht der mittlere quadratische Fehler, sondern der mittlere absolute Fehlerbetrag minimiert, was u. U. auch erwünscht sein kann (s. **Mills** 1990: 338).

<sup>18</sup> Wenn zur Transformation die Gleichung (2) benutzt wurde, muß die Korrekturformel leicht modifiziert werden (s. **Guerrero** 1993: 41).

<sup>19</sup> **Mills** (1990: 339) referiert aus einer Arbeit von **Pankratz und Dudley** Bedingungen, unter denen die Verzerrung bezüglich des MMSFE größer oder kleiner ausfällt.

Die hier erzielten Ergebnisse könnten dadurch beeinflusst sein, daß zwischen Schätz- und Validierungsperiode möglicherweise ein (in der von mir durchgesehenen Literatur nicht diskutierter) Strukturbruch aufgetreten ist oder/und daß das Modell nicht adäquat spezifiziert wurde. Auf diese Möglichkeit weist die Tatsache hin, daß die saisonalen MA-Schätzer für die Schätzperiode bis Mai 1970 mit  $\hat{\Theta}_{12} = .61$  erheblich von dem Schätzer  $\hat{\Theta}_{12} = .973$  abweichen, der bei Ausdehnung der Schätzperiode bis Mai 1971 erzielt wird. Auffällig ist, daß mit den von **Box und Jenkins** (1973) eingesetzten Schätzern  $\hat{\Theta}_{12} = .8$  und  $\hat{\lambda} = 0.25$  der unbereinigte mittlere Prognosefehler mit  $RMS = 72.67$  (statt  $RMS = 81.51$ ) deutlich gesenkt werden kann. Zu einem ähnlich verbesserten Prognoseergebnis führt auch "Modell B" von **Chatfield und Prothero** (1973 b), das den saisonalen MA-Parameter durch einen saisonalen AR-Parameter ersetzt:  $(1+.557B)(1+.189B^{12})Z_t(1-B)(1-B^{12}) = a_t$ . Die Frage, ob ungefähr im Mai 1970 ein gewisser Strukturbruch eingetreten ist, läßt sich aber ohne zusätzliche Daten (über Mai 1971 hinaus) nicht beantworten.

Zur Frage, ob das von **Box und Jenkins** vorgeschlagene Verfahren der empirischen Modellidentifikation mittels Autokorrelationsfunktionen und partiellen Autokorrelationsfunktionen in bestimmten Fällen versagt, stellen **Bowerman et al** (1990) einige interessante Überlegungen an. Dabei beziehen sie sich auf eine Empfehlung von **Chatfield und Prothero** (1973 b), überhaupt keine Potenztransformation vorzunehmen (also auch nicht die von ihnen ursprünglich vorgeschlagene Log-Transformation), sondern: "If the seasonal effect is approximately multiplicative (as for our data), then we would use the Operator  $\nabla_{12}^2$ , since this removes a linear trend and multiplicative seasonal pattern. If the seasonal effect is approximately additive, we would use  $\nabla\nabla_{12}$ , since this removes a linear trend and additive seasonal pattern" (ebd.: 347)<sup>20</sup>. Die Notwendigkeit, den saisonalen Differenzenoperator *zweifach* anzuwenden, wäre nach den Regeln der *Box/Jenkins-Modellidentifikation* angezeigt, wenn in der Autokorrelationsfunktion der Zeitreihe  $Y_t = \nabla\nabla_{12}X_t$  - also nach einer einfachen saisonalen Differenzierung - bei den saisonalen "lags"  $k = 12, 24, 36$  usw. immer noch allmählich gegen Null tendierende Koeffizienten aufträten. Ein solches Muster ist aber auch dann häufig nicht zu beobachten, wenn Trend- und Saisonkomponente multiplikativ verknüpft sind. **Bowerman et al.** (1990: 425 f.) begründen das damit, daß die meisten Zeitreihen mit wachsender Saisonamplitude starke deterministische Prozeßkomponenten enthalten. Allgemein gilt, daß ein (stochastisches) ARIMA-Modell sich in dem Maße einem in der Prognosefunktion deterministischen Regressionsmodell annähert, je geringer die Theta-Parameter des MA-Operators

$$\Theta^*(B) = \Theta(B)\Theta(B^*)$$

von den Phi-Parametern des generalisierten autoregressiven

<sup>20</sup> Ein multiplikatives Trend-Saison-Modell bedeutet bei monatlichen Daten a) daß die für jeden fixierten Monat (z. B. Januar oder Juli) im Verlauf der Jahre beobachteten Werte einen linearen Trend aufweisen und b) daß diese Trendverläufe unterschiedlich sind und die saisonale Variation folglich zunimmt (s. **Bowerman et al** 1990: 423 f.). Bei identischen Trendverläufen könnte die Saisonkomponente schon durch einfache (saisonale) Differenzenbildung in eine stationäre Reihe überführt werden.

Operators  $\varphi(B) = U(B)\phi(B)\Phi(B^s)$  im Modell  $\varphi(B)z_t = \theta'(B)a_t$  abweichen<sup>21</sup>. Dabei ist  $\theta(B) = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q)$ ;  $\Theta(B) = (1 - \Theta_1 B^s - \Theta_2 B^{2s} - \dots - \Theta_r B^{rs})$ . Die  $\phi(B)$  und  $\Phi(B)$  sind analog hierzu definierte Polynome;  $U(B)$  ist der sog. "simplifying Operator", dessen Wurzeln alle auf dem Einheitskreis liegen. Spezialfälle sind die Differenzenoperatoren  $(1-B)^d$  und  $(1-B^s)^p$ , wobei  $s$  die Länge der Saisonperiode nennt und  $d$  bzw.  $D$  angeben, wie oft der jeweilige Differenzenoperator anzuwenden ist. Beispielsweise wird der Prozeß

$$(1-B^s)y_t = (1-\theta_s B^s)a_t$$

bei  $\theta_s \rightarrow 1$  für die Reihe  $Y$ , Autokorrelationskoeffizienten produzieren, die sich auch bei den saisonalen "lags"  $s, 2s$  usw. dem Wert 0 nähern.<sup>22</sup>

**Bowerman et al.** (1990) geben folgende Empfehlungen: Wenn die Saison-Amplitude nicht-linear wächst, sollte eine Datentransformation nach dem Box/Cox-Modell gewählt werden (ebd., S. 430, 434). Bei linearem Anwachsen der Saisonkomponente ist dagegen ein Modell mit dem saisonalen Differenzenfaktor  $(1-B^s)^2$  für die untransformierten Daten in der Regel angebracht. Solche Modelle sind allerdings nur sehr schwer zu identifizieren, da sie sich, wie eben erläutert, in der empirischen Autokorrelationsfunktion häufig nicht zu erkennen geben. Als brauchbare Alternative stellen **Bowerman et al.** (1990) Modelle mit additiven oder - wenn die deterministische Komponente stark wird - multiplikativen Saison-Dummies vor.

Der Plot der Sales-Daten, der nur sechs volle Saison-Perioden umfaßt (s. oben, Abb. 2) läßt nicht eindeutig erkennen, ob die Saison-Amplituden linear oder nicht-linear zunehmen. Die mittlere Periode über vier "Gipfel" erscheint linear im Zuwachs, nimmt man den ersten und letzten Gipfel hinzu, wird ein leicht kurvenförmiger Verlauf erkennbar. Es ist also schwer antizipierbar, ob ein Modell mit transformierten Daten oder ein double seasonal difference model in diesem Fall zu besseren Ergebnissen führt. Das von **Abraham und Ledolter** (1986: 64) für die untransformierten Sales-Daten vorgeschlagene Modell  $(1-B^{12})^2 Z_t = (1-\theta_1 B - \theta_2 B^2)(1-\Theta_1 B^{12})a_t$  führt in der Validierungsperiode zu Prognosen, deren Qualität (RMS = 93.7) schlechter ist als diejenige, die mit **Chatfields** "Modell A" (s. oben) und einer vorherigen Datentransformation  $\hat{\lambda} = 0.33$  erzielt wurde. Wenn der oben erwähnte Ausreißer nicht eliminiert wird, führen die beiden Modelle jedoch zu etwa gleich guten Prognosen.

<sup>21</sup> Siehe hierzu die Erläuterungen in **Abraham und Box** (1978) sowie **Abraham und Ledolter** (1986).

<sup>22</sup> Eine genauere Erläuterung müßte auf die Lösungsformeln der jeweiligen Differenzgleichungen eingehen, die die Autokorrelationsfunktionen erzeugen und die "eventual forecastfunction" ergeben. Dies würde offenkundig den Rahmen der vorliegenden Arbeit überschreiten.



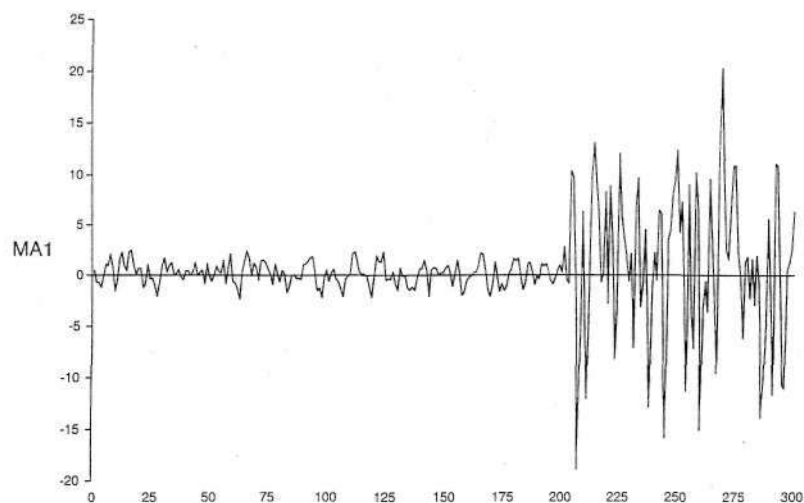
Damit schließen wir unsere Betrachtungen über die Berücksichtigung trendabhängiger Varianzen ab. Für zeitspezifische Varianzen benötigt man andere Strategien als die Box/Cox-Transformation. Zwei dieser Strategien werden im nächsten Abschnitt vorgestellt.

### 3. Zeitspezifische Varianzen

#### a) Abrupte Varianzänderungen (Tsay-Modell)

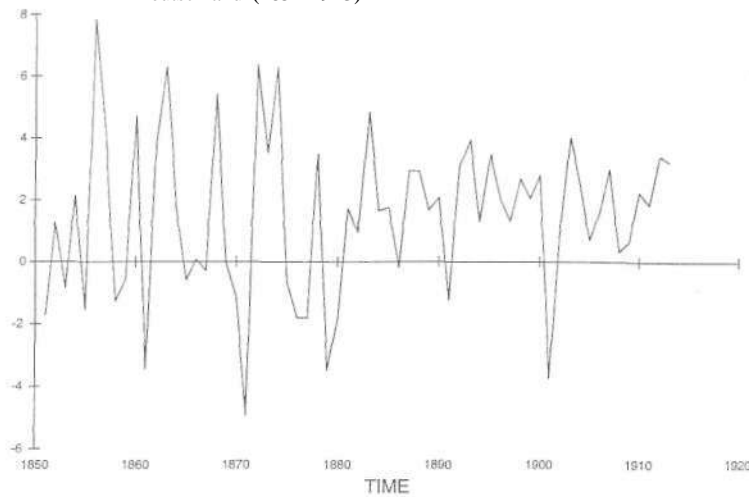
Abbildung 7 verdeutlicht, was mit einer abrupten Varianzänderung "idealtypisch" gemeint ist. Sie repräsentiert einen MA(1)-Prozeß  $X_t = (1-\theta B)a_t$ , der mit  $\sigma_t = 1.0$  für  $t = 1, 2, \dots, 200$  und  $\sigma_t = 6.0$  für  $t = 201, 202, \dots, 300$  simuliert wurde.

**Abbildung 7:** Simulierter MA(1)-Prozeß mit Varianzsprung bei  $n=200$



Die Situation ist insofern einfach, als sich die Varianzen nicht ständig, sondern nur zwischen zwei relativ klar abgrenzbaren Perioden ändern, innerhalb derer sie stabil sind. In der Realität treten gelegentlich Zeitreihen wie die in Abbildung 8 auf:

**Abbildung 8:** Jährliche Wachstumsrate des realen Bruttoinlandsprodukts (BIP) in Deutschland (1851-1913)



Die Varianzen sind in der ersten Hälfte deutlich größer als in der zweiten Hälfte; aber die Varianzänderung tritt weniger abrupt ein, sondern vollzieht sich über mehrere Intervalle; sie ist zudem weniger ausgeprägt als in dem simulierten Prozeß. Ob sie die Modellidentifikation und die Parameterschätzung relevant beeinflusst, läßt sich nicht ohne besondere Prüfung feststellen. *Tsay* (1988) hat ein leicht handhabbares Verfahren vorgeschlagen, derartige Varianzsprünge zu lokalisieren und bei der Identifikation des Prozeßmodells zu berücksichtigen:

Das ARMA-Modell einer Reihe  $X_t$ , die einen Varianzsprung zum Zeitpunkt  $t=d$  aufweist, läßt sich in der üblichen *Box/Jenkins*-Notation wie folgt definieren:

$$(9) \quad b_t = \frac{\phi(B)}{\theta(B)} Y_t$$

$$b_t = \begin{cases} a_t & \text{für } t < d \\ a_t[1 + \omega_t] & \text{für } t \geq d \end{cases}$$

Dabei sind die  $a_t$ , mit  $t = 1, 2, \dots$  unabhängig normalverteilte Zufallsgrößen mit  $E(a_t) = 0$  und der Varianz/Kovarianzmatrix  $\sigma^2 I_r$ . Während die Parameter des ARMA-Modells - die Theta- und Phi-Polynome - konstant sind, wächst der *white noise* Input zu einem

bestimmten Zeitpunkt  $t=d$  um den Betrag  $\omega_v$ , der zu schätzen ist. Das Verhältnis der Varianzen vor und nach  $t=d$  ist mit

$$(11) \quad \hat{r}_d = \frac{(d-1) \sum_{t=d}^n b_t^2}{(n-d+1) \sum_{t=1}^{d-1} b_t^2}$$

gegeben, wobei  $(d-1) > 0$  und  $(n-d+1) > 0$ . Diese Verhältnisgröße ist ein Schätzer für  $(1 + \omega_v)^2$ . Unter der Nullhypothese (keine Varianzänderung) ist  $\hat{r}_d$   $F$ -verteilt mit den Freiheitsgraden  $(n-d+1)$  und  $(d-1)$ . Wenn  $d$  unbekannt ist, lassen sich verschiedene Tests konstruieren, aber keine Verteilungsfunktionen analytisch ableiten. Tsay (1988: 5 ff.) schlägt vor, wie folgt zu verfahren:

Im 1. Schritt wird für die beobachtete Reihe  $X_t$  nach dem üblichen Verfahren ein ARI-MA-Modell identifiziert und geschätzt (so, als ob die Varianzen stabil wären). Die Residuen dieses Modells dienen als Schätzer für die  $b_t$ 's in Gleichung (10). Auf dieser Basis werden im 2. Schritt die Varianzverhältnisse  $\lambda_{\min}$  und  $\lambda_{\max}$  über eine Menge von  $d$ 's berechnet:

$$(12) \quad \begin{aligned} \lambda_{\min} &= \text{Min}\{\hat{r}_d \mid h \leq d \leq n - h\} \\ \lambda_{\max} &= \text{Max}\{\hat{r}_d \mid h \leq d \leq n - h\} \end{aligned}$$

Dabei bezeichnet  $h$  die minimale Anzahl der Meßzeitpunkte am Anfang und am Ende einer Zeitreihe, auf deren Basis das Varianzverhältnis  $\hat{r}_d$  gemäß Gleichung (11) ermittelt wird. Tsay (1988: 18) schlägt ein  $h$  zwischen 20 und 30 vor. Neben dem maximalen ist auch das minimale Varianzverhältnis festzustellen, weil die größere Varianz im Zähler oder im Nenner des Quotienten stehen kann. Im 3. Schritt wird der reziproke Wert von  $\lambda_{\min}$  berechnet und  $\lambda_v = \text{Max}[\lambda_{\max}, (\lambda_{\min})^{-1}]$  ermittelt. Das Varianzverhältnis  $\lambda_v$  wird dann mit einem a priori spezifizierten kritischen Wert  $C$  (s. unten) verglichen. Falls  $\lambda_v < C$  ist, liegt keine signifikante Varianzänderung vor; die Iteration stoppt. Falls  $\lambda_v \geq C$ , wird der entsprechende Zeitpunkt  $d = d_0$  notiert und im 4. Schritt die Zeitreihe entsprechend adjustiert:

$$(13) \quad Y_t = \begin{cases} X_t & \text{für } t < d_0 \\ \bar{X} + \frac{X_t - \bar{X}}{\sqrt{\lambda_v}} & \text{für } t \geq d_0 \end{cases}$$

Dabei ist  $\bar{X}$  das Stichprobenmittel der Zeitreihe  $X_t$ . Die Schritte eins bis drei werden mit der so transformierten Reihe wiederholt, um festzustellen, ob auch noch zu anderen Zeitpunkten  $d \neq d_0$  sinifikante Varianzsprünge stattgefunden haben. Wie bereits erwähnt,

können die kritischen Werte  $C$  nicht analytisch abgeleitet werden. Auf der Basis von Monte-Carlo-Studien hat *Tsay* die Prozentquantile für *white noise* und autoregressive Prozesse 1. Ordnung (mit  $\phi = 0.8$ ) ermittelt (s. Tabelle 2).

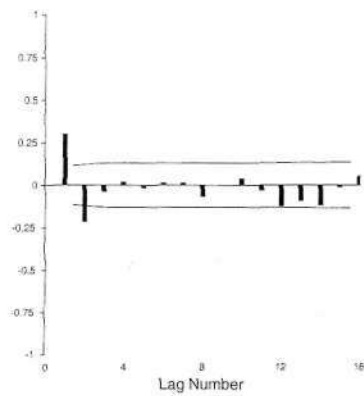
**Tabelle 2:** Empirische Perzentile der  $\lambda$ -Statistik, basierend auf 5000 Wiederholungen

Perzentile	Stichprobenumfang					
	100		200		300	
	$\lambda_{\max}$	$1/\lambda_{\min}$	$\lambda_{\max}$	$1/\lambda_{\min}$	$\lambda_{\max}$	$\lambda_{\min}$
<b>(a) White Noise</b>						
90%	2.33	2.53	2.19	2.42	2.22	2.38
95%	2.67	2.93	2.52	2.97	2.60	2.78
97.5%	3.14	3.36	2.91	3.26	3.01	3.24
99%	3.89	4.07	3.48	3.98	3.61	3.92
99.5%	4.30	4.63	3.96	4.44	4.19	4.44
<b>(b) AR(1) Modell mit <math>\phi = 0.8</math></b>						
90%	2.34	2.51	2.25	2.41	2.24	2.36
95%	2.77	2.93	2.67	2.77	2.62	2.70
97.5%	3.21	3.46	3.14	3.23	3.13	3.10
99%	3.93	4.24	3.95	3.87	3.90	3.68
99.5%	4.49	4.81	4.69	4.44	4.38	4.08

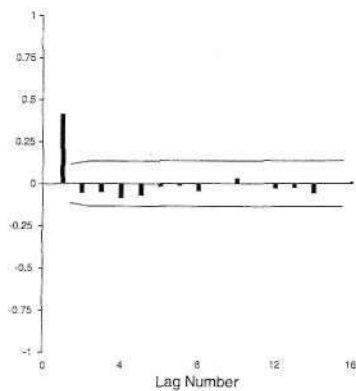
Für die simulierten Daten in Abb. 7 ist der Zeitpunkt  $t = d_0$  des Varianzsprungs bekannt. Ohne die Daten zu adjustieren, wird für sie eine empirische Autokorrelationsfunktion (AKF) ermittelt, die nicht einen MA(1), sondern einen MA(2) Prozeß anzeigt (s. Abb. 9a). Adjustiert man dagegen die Daten gemäß Gleichung (13) wird in der AKF (s. Abb. 9b) das korrekte MA(1)-Modell klar erkennbar.<sup>23</sup>

<sup>23</sup> Weitere Simulationsbeispiele zeigen, daß auch relativ große Varianzsprünge die Autokorrelationsfunktion nicht unbedingt verzerren müssen.

**Abbildung 9a:** Autokorrelationsfunktion des simulierten MA(1)-Prozesses in Abb. 7



**Abbildung 9b:** Autokorrelationsfunktion des simulierten MA(1)-Prozesses nach Varianzadjustierung

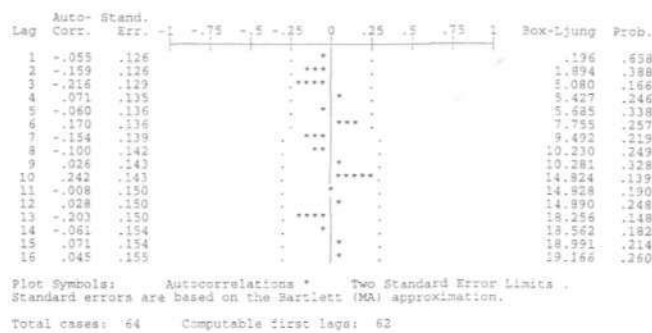


Das von *Tsay* vorgeschlagene Verfahren ist vor allem bei kurzen Zeitreihen vorteilhaft, weil es erlaubt, bei der Modellschätzung sämtliche Meßzeitpunkte in der Stichprobe zu belassen.

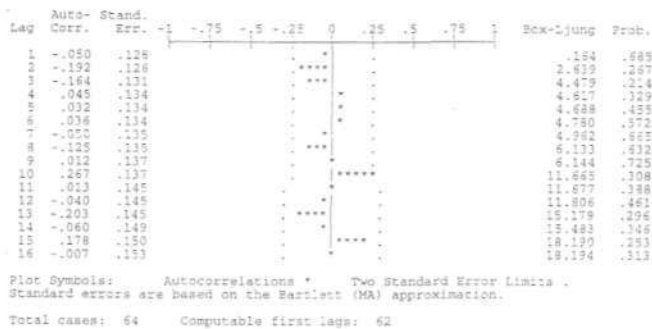
Die Modellidentifikation für die in Abb. 8 abgebildeten Wachstumsraten des BIP von 1851 bis 1913 ist nicht eindeutig. Die *Ljung-Box* Statistik legt zwar die Beibehaltung der Nullhypothese nahe, es handele sich um einen *white noise* Prozeß (minimales Feh-

lerisiko für die Zurückweisung der Nullhypothese immerhin  $\alpha = 0.139$ ), aber die relativ starken Koeffizienten bei Lag 2 und Lag 3 der SAKF (s. Abb. 10a) lassen auch ein MA(2)- oder MA(3)-Modell als möglich erscheinen.

**Abbildung 10a:** Autokorrelationsfunktion des BIP (s. Abb. 8)



**Abbildung 10b:** Autokorrelationsfunktion des BIP nach Varianzadjustierung



Wir akzeptieren zunächst die Nullhypothese eines ARMA(0,0)-Modells und berechnen auf dieser Basis (d. h. unmittelbar aus den beobachteten Werten) ein maximales Varianzverhältnis  $\hat{\lambda}_v = 4,33$  zum Zeitpunkt  $\hat{d}_0 = 1880$ . Nach den Angaben in Tabelle 2 dürfte dieser Wert mindestens auf dem 5-Prozent-Niveau statistisch signifikant sein. Die SAKF der nach Gleichung (13) adjustierten Zeitreihenwerte verstärkt die Evidenz für einen *white noise* Prozeß (s. Abb. 10b). Schätzt man mit den adjustierten Daten ein MA(3)-Modell mit der Restriktion  $\theta_1 = 0$ , erhält man für die Schätzer von  $\theta_2$  und  $\theta_3$  t-Werte von 1.72 (entspricht  $\alpha=0.09$ ) und 1.58 ( $\alpha=0.12$ ). Ohne diese Varianzadjustierung lagen die Alpha-Werte mit 0.07 und 0.06 näher an dem üblichen Signifikanzkriterium von  $\alpha=0.05$ , das zur Annahme des MA(3)-Modells geführt hätte. Für die ökonomische

Wachsturstheorie ist die Unterscheidung eines *white noise* Prozesses von einem MA-Modell aber keineswegs trivial (s. Metz 1992).

Die BIP-Daten in Abb. 8 dürften einen Grenzfall für die Anwendbarkeit des *Tsay-Modells* darstellen. Eine flexiblere Strategie für die Modellierung zeitabhängiger Varianzen beinhalten die sog. (*Generalized*) *Autoregressive Conditional Heteroscedasticity* - (G)ARCH - Modelle.

#### b) ARCH-Modelle

Eine ausführliche Darstellung dieser Modellklasse würde den Rahmen dieses Artikels sprengen;<sup>24</sup> der Grundgedanke soll aber kurz erläutert werden. So wie die ARIMA-Modelle die **bedingten** Erwartungswerte eines stochastischen Prozesses ausdrücken, so sollen die (G)ARCH-Modelle die **bedingten** Varianzen darstellen; ARIMA- und (G)ARCH-Komponente bilden ein integriertes Modell, dessen Parameter gemeinsam geschätzt werden. Das sei anhand eines autoregressiven Prozesses 1. Ordnung - AR(1) bzw. AR-MAC 1,0) Modell - erläutert. In dem AR(1)-Modell

$$(14) \quad Y_t = \varphi Y_{t-1} + \varepsilon_t; E(\varepsilon_t) = 0; V(\varepsilon_t) = \sigma^2; Cov(\varepsilon_t, \varepsilon_{t-k}) = 0, k > 0$$

ist der nicht-bedingte Erwartungswert  $E(Y_t) = 0$ , der bedingte Erwartungswert dagegen

$$(15) \quad E(Y_t | y_{t-1}) = \varphi y_{t-1}$$

Die nicht-bedingte Varianz ist mit

$$V(Y_t) = \frac{\sigma^2}{1 - \varphi}$$

gegeben (s. *Box und Jenkins* 1976: 49, 58), die bedingte mit

$$(17) \quad \begin{aligned} V(Y_t | y_{t-1}) &= E[(Y_t - E(Y_t | y_{t-1}))^2] \\ &= E[(Y_t - \varphi y_{t-1})^2] \\ &= E(\varepsilon_t^2) \\ \sigma_t &= \sigma^2 \end{aligned}$$

Bei dieser Modellierung ist die bedingte Varianz (Moment 2. Ordnung) im Unterschied zum bedingten Erwartungswert (Moment 1. Ordnung) ebenso eine Konstante wie die nicht-bedingte Varianz. Im ARCH-Modell wird dagegen die zeitabhängige Varianz in

<sup>24</sup> Die ARCH-Modelle wurden ursprünglich von Engle (1982) zur Diskussion gestellt. Kurze Einführungen geben Mills (1990) und Greene (1993). Eine ausführliche Übersicht über Theorie und praktische Anwendungen geben Bollerslev et al. (1992).

ähnlicher Weise modelliert wie der zeitabhängige Mittelwert. Nimmt man z. B. an, daß überdurchschnittlich große Fehlerquadrate zum Zeitpunkt  $t-1$  überdurchschnittlich große Varianzen zum Zeitpunkt  $t$  nach sich ziehen, läßt sich diese Vorstellung mit Hilfe der Gleichung

$$(18) \quad V(\varepsilon_t | \varepsilon_{t-1}) = V(Y_t | y_{t-1}) = h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2$$

ausdrücken. Das kombinierte AR(1)-ARCH(1)-Modell wird somit in drei Gleichungen dargestellt:

$$(19) \quad \begin{aligned} Y_t &= \phi Y_{t-1} + \varepsilon_t \\ \varepsilon_t | \varepsilon_{t-1} &\sim N(0, h_t) \\ h_t &= \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 \end{aligned}$$

(siehe *Greene* 1993: 568)

Um den veränderten Modellkontext anzudeuten, ist  $\sigma^2$  als Kürzel für "bedingte Varianz" durch  $h_t$  ersetzt worden. Um sicherzustellen, daß die bedingte Varianz  $h_t$  positiv ist, muß  $\alpha_0 > 0$  und  $\alpha_1 \geq 0$  sein. Damit die nicht-bedingte Varianz von  $\varepsilon_t$  endlich ist, muß  $\alpha_1 < 1$  sein. Unter dieser Voraussetzung gilt für die nicht-bedingte Varianz (s. *Mills* 1990: 327)

$$(20) \quad V(\varepsilon_t) = \sigma^2 = \frac{\alpha_0}{1 - \alpha_1} \quad (\text{vgl. Gleichung (16)}).$$

Aus den Gleichungen (18) und (19) folgt

$$(21) \quad \hat{h}_t - \sigma^2 = \alpha_1 (\varepsilon_{t-1}^2 - \sigma^2)$$

Das heißt, die bedingte Varianz zum Zeitpunkt  $t$  wird die "durchschnittliche" (nicht-bedingte) Varianz immer dann überschreiten, wenn der quadrierte Fehler zum Zeitpunkt  $t-1$  seinen nicht bedingten Erwartungswert  $\sigma^2$  übersteigt. Man beachte, daß die Fehler, obwohl nicht autokorreliert, nicht unabhängig voneinander, sondern über ihre 2. Momente miteinander verbunden sind. Es ist wahrscheinlich, daß hohen (niedrigen) Fehlerquadraten  $\varepsilon_t^2$  hohe (niedrige) Fehlerquadrate ( $\varepsilon_{t+1}^2$ ) folgen, bis sie auf Grund irgendeines Zufallseinflusses für eine gewisse Zeit unter (über) ihr Durchschnittsniveau fallen (steigen). Im äußeren Erscheinungsbild treten also Varianzcluster auf. Wenn ein ARCH-Effekt vorliegt, ist die Varianz der SAKF größer als mit der üblichen *Bartlett*-Formel berechnet (s. *Tsay* 1987: 598; *Mills* 1990: 332); d. h., ohne Korrektur oder explizite Modellierung der instabilen Varianzen kommt man leicht zu einer Überparametrisierung des ARMA-Modells.



Das ARCH(1)-Modell lässt sich zu einem Modell höherer Ordnung erweitern:

$$(22) \quad h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-2}^2 + \dots + \alpha_q \varepsilon_{t-q}^2$$

Bei großem  $q$  empfiehlt **Bollerslev** (1986) die Umwandlung in ein Mischmodell GARCH(p,q) mit autoregressiver und *moving average* Komponente:<sup>25</sup>

$$(23) \quad h_t = \alpha_0 + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^p \beta_i h_{t-i}$$

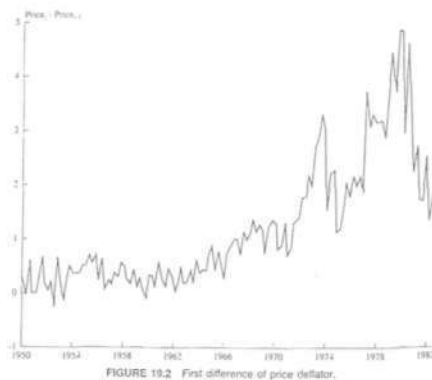
Wenn man für die Differenzen  $\varepsilon_t^2 - h_t = v_t$ , ( $t=1,2,\dots,n$ ) einen *white noise* Prozeß annimmt, läßt sich diese Gleichung umformen zu

$$(24) \quad \begin{aligned} \varepsilon_t^2 &= h_t + v_t \\ \varepsilon_t^2 &= \alpha_0 + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^p \beta_i \varepsilon_{t-i}^2 - \sum_{i=1}^p \beta_i v_{t-i} + v_t \end{aligned}$$

(s. **Bollerslev** 1986: 310). Somit kann der GARCH(p,q)-Prozeß als ARMA(m,p)-Prozeß in den Fehlerquadraten  $\varepsilon_t^2$  interpretiert werden, wobei die Ordnung der autoregressiven Komponente mit  $m = \max(p,q)$  gegeben ist. **Bollerslev** (1986) führt weiter aus, daß das übliche Instrumentarium der empirischen Identifikation von ARMA-Modellen mittels SAKF und SPAKF analog auf die Zeitreihen der (geschätzten) Fehlerquadrate angewandt werden kann. Somit ist auch die Ljung/Box-Statistik, bezogen auf die Autokorrelation der Fehlerquadrate, eine erste Möglichkeit, das Vorliegen (irgendwelcher) ARCH-Effekte zu testen. Auf die Test- und Schätzmethodik kann ich im Rahmen dieses Artikels nicht weiter eingehen (s. hierzu beispielsweise **Mills** 1990: 331 ff.; **Greene** 1993: 440 ff., 570 ff.) Um wenigstens einen Eindruck zu vermitteln, für welche Art von Zeitreihen GARCH-Modelle in Frage kommen, sei hier die Entwicklung des US-Preisindex (Deflator) abgebildet (s. Abb. 11). Die ARCH-Modellierung mit diesen Daten wird in **Mills** (1990: 336) und **Greene** (1993: 576) erläutert.

<sup>25</sup> Auch exogene Variablen lassen sich in dem Bedingungs- bzw. Informationsset berücksichtigen (s. **Bollerslev** 1986).

Abbildung 11: Erste Differenzen des Preisindex der USA (Quelle: Greene 1991: 516)



#### 4. Zusammenfassung und Schlußbemerkungen

Die Modellierung von Zeitreihen beruht auf der Theorie *stationärer* stochastischer Prozesse. Häufig weisen Zeitreihen jedoch nicht nur einen Trend auf, sondern auch trendabhängige oder zeitspezifische Varianzen. Instabile Varianzen können bestimmte Kenngrößen (wie z. B. die Autokorrelationsfunktion) verzerren, die Standardfehler aufblähen und bei Signifikanztests zu falschen Schlüssen führen. In welchem Maße das der Fall ist, läßt sich nicht durch eine rein visuelle Analyse des Zeitreihenverlaufs feststellen. Deshalb sollte man eine erkennbare Instabilität der Varianzen auf jeden Fall bei der Modellierung von Zeitreihen berücksichtigen. Trendabhängige Varianzen können in der Regel mit Hilfe einer *Box/Cox*-Transformation der Daten stabilisiert werden. Die entsprechenden Transformationsparameter lassen sich relativ einfach mit Hilfe des "spread versus level plot" und einer damit verbundenen Regressionsanalyse schätzen. Standard-Transformationen, wie z. B. das Logarithmieren der Daten, sollten nicht ungeprüft übernommen werden. Bei der Analyse von Anteilswerten, z. B. Arbeitslosenraten, ist die logistische Transformation als Alternative zu den *Box/Cox*-Transformationen zu erwägen (s. *Wallis* 1987; *Guerrero und Johnson* 1982).

Erhebliche Varianzschwankungen können auch bei mittelwert-stationären Reihen auftreten, z. B. in Form von Varianzsprüngen. *Tsay* (1988) hat vorgeschlagen, sie analog zu den *Box/Tiao*-Interventionsmodellen zu konzeptualisieren und die Varianz der untersuchten Reihe entsprechend zu adjustieren. Die von *Tsay* vorgeschlagene Strategie, die technisch einfach auszuführen ist, läßt sich nicht anwenden, wenn sich die zeitspezifischen Varianzänderungen allmählich und nicht in Form von "Sprüngen" vollziehen. In

einem solchen Falle empfiehlt sich eine ARCH- bzw. GARCH-Modellierung, die mit einer ARMA-Modellierung gekoppelt wird. Während die ARMA-Modelle die interne Beziehungsstruktur der bedingten Mittelwerte ausdrücken, leisten die (G)ARCH-Modelle das gleiche bezüglich der bedingten Varianzen, sprich der quadrierten Residuen der ARMA-Modelle, was wiederum zu einem iterativen Schätzverfahren führt, das aber in diesem Artikel nicht mehr erläutert wurde. Der besondere Vorteil dieser Strategie liegt einerseits in ihrer Flexibilität (sie läßt sich auf eine große Bandbreite unterschiedlicher Formen zeitabhängiger Varianzen anwenden), andererseits darin, daß sie eine explizite Modellierung der Varianzen beinhaltet und damit zu deren substantieller Interpretation auffordert. Instabile Varianzen sind dann nicht mehr nur ein "Ärgernis", das die Modellierung der bedingten Mittelwerte (also zeitabhängiger Niveaus) stört, sondern ein eigenständiger Indikator sozialer oder ökonomischer Prozesse (ökonomische Beispiele finden sich in der bereits erwähnten Arbeit von **Bollerslev et al.** (1992).

Sowohl die (G)ARCH-Modelle als auch die Box/Cox-Transformation werden nicht nur in der univariaten Zeitreihenanalyse, sondern auch in der Regressionsanalyse angewandt. Für diesen Kontext stellt das Programmsystem LIMDEP (**Greene** 1991) EDV-Routinen zur Verfügung. Bei der **Box/Cox**-Transformation im Rahmen der Regressionsanalyse ist darauf zu achten, daß diese Datentransformationen nicht nur die Varianz stabilisieren, sondern auch die Form der Beziehung zwischen Kriterium und Regressor verändern, was durchaus unerwünscht sein kann (s. **Judge et al.** 1985: 444 f., 839 ff); **Kmenta** 1986: 298, 520 ff). Alternativen zur **Box/Cox**-Transformation, die diese Problematik umgehen sollen, schlagen **Carroll und Ruppert** (1988); **Ruppert und Aldershof** (1989) sowie **Tibshirani** (1988) vor<sup>26</sup>.

## Literatur

**Abraham, B.; Ledoller, J.** 1983:

Statistical Methods for Forecasting. New York: Wiley.

**Anderson, O. D.** 1976:

On the Transformation of Raw Time Series Data: A Review. Statistische Hefte, 17, 285-289.

**Anderson, O. D.** 1979:

Time Series Analysis and Forecasting. The Box-Jenkins Approach. London / Boston: Butterworths.

**Bollerslev, Tim.** 1986:

Generalized Autoregressive Conditional Heteroskedasticity. Journal of Econometrics, 31, 307-327.

**Bollerslev, Tim; Chou, Ray Y.; Kroner, Kenneth F.** 1992:

ARCH Modeling in Finance. Journal of Econometrics, 52, 5-59.

<sup>26</sup>

Zu erinnern ist auch daran, daß die Koeffizienten eines nicht-linearen Modells nicht als Steigungsgrößen ("slopes") interpretiert werden können. Die (punktuelle) Steigung muß über die 1. Ableitung der entsprechenden Regressionsgleichung ermittelt werden.

- Bowerman, B. L.; Koehler, A. B.; Pack, D. J.** 1990:  
Forecasting Time Series with Increasing Seasonal Variation. *Journal of Forecasting*, 9, 419-436.
- Box, G. E. P.; Jenkins, G. M.** 1973:  
Some Comments on a Paper by Chatfield and Prothero and on a Review by Kendall. *Journal of the Royal Statistical Society A*, 135, 337-352.
- Box, G. E. P.; Jenkins, G. M.** 1976:  
*Time Series Analysis Forecasting and Control* (2nd edition). San Francisco: Holden-Day.
- Box, G. E. P.; Tiao, G. P.** 1975: Intervention Analysis with Applications to Economic and Environmental Problems. *Journal of the American Statistical Association*, 70, 70-79.
- Carroll, Raymond J.; Ruppert, David** 1984: Power Transformations When Fitting Theoretical Models to Data. *Journal of the American Statistical Association*, 79, 321-328.
- Chatfield, C; Prothero, D. L.** 1973b:  
Reply to Box/Jenkins (1973). *Journal of the Royal Statistical Society A*, 136, 345-352.
- Emerson, John D.; Stoto, Michael A.** 1983:  
Transforming Data. In: Hoaglin, D. C, Mosteller, F. & Tukey, J. W. (Hrsg.), *Understanding Robust and Exploratory Data Analysis* (S. 97-127). New York usw.: Wiley.
- Engle, Robert F.** 1982:  
Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica*, 50, 987-1007.
- Greene, William H.** 1993:  
*Econometric Analysis* (2nd. edition). New York / London: Macmillan.
- Greene, William H.** 1991:  
*LIMDEP, Version 6.0: User's Manual*. Bellport, N. Y.: Econometric Software.
- Guerrero, Victor M.; Johnson, Richard A.** 1982:  
Use of the Box-Cox Transformation with Binary Response Models. *Biometrika*, 69, 309-314.
- Guerrero, Victor M.** 1993: Time-Series Analysis supported by Power Transformations. *Journal of Forecasting*, 12, 37-48.
- Hartung, J.; Elpelt, B.; Klösener, H.-H.** 1985:  
*Statistik* (3. Auflage). München/Wien: Oldenbourg.
- Hopwood, W. S.; McKeown, J. C; Newbold, P.** 1984:  
Time Series Forecasting Models Involving Power Transformations. *Journal of Forecasting*, 3, 57-61.
- Johnston, J.** 1984:  
*Econometric Methods* (3rd. edition). Auckland u. a.: McGraw-Hill.
- Judge, George G. et al** 1985:  
*The Theory and Practice of Econometrics* (2nd. edition). New York usw.: John Wiley and Sons.
- Kmenta, Jan** 1986: *Elements of Econometrics* (2nd ed.). New York / London: Macmillan.
- Lenk, Peter J.; Tsai, Chih-Ling** 1990:  
Transformation and Dynamic Linear Models. *Journal of Forecasting*, 9, 219-232.

- Metz, Rainer** 1992:  
Über die stochastische Struktur langfristiger Wachstumsschwankungen. In: ifo Studien 38, S. 171-197.
- Mills, Terence C.** 1990: Time Series Techniques for Economists. Cambridge usw.: Cambridge University Press.
- Mohr, Walter** 1980:  
Grobidentifikation und Modellvergleich bei ARIMA-Modellen (mit einer Fallstudie für die Reihe der Arbeitslosen in der BRD). Allgemeines Statistisches Archiv, 64, 164-183.
- Mohr, Walter** 1983: Untersuchungen zur Anwendung der Box-Cox-Transformation in der Zeitreihenanalyse. Allgemeines Statistisches Archiv, 67, 177-198.
- Mosteller, Frederick; Tukey, John W.** 1977:  
Data Analysis and Regression. A Second Course in Statistics. Reading usw.: Addison-Wesley.
- Nelson, Harald L; Granger, C. W. J.** 1979:  
Experience with Using the Box-Cox Transformation When Forecasting Economic Time Series. Journal of Econometrics, 10, 57-69.
- Poirier, Dale J.** 1980:  
Experience With Using the Box-Cox Transformation When Forecasting Economic Time Series. A Comment. Journal of Econometrics, 14, 277-280.
- Ruppert, David; Aldershof, Brian** 1989:  
Transformations to Symmetry and Homoscedasticity. Journal of the American Statistical Association, 84, 437-446.
- Tibshirani, Robert** 1988: Estimating Transformations for Regression Via Additivity and Variance Stabilization. Journal of the American Statistical Association, 83, 394-405.
- Thome, Helmut** 1992: Beschreibende Zeitreihenanalyse: Komponentenzersetzung mit Hilfe gleitender Mittelwerte. In: Historical Social Research/Historische Sozialforschung 17, S. 63-105.
- Tsay, Ruey S.** 1987:  
Conditional Heteroskedastic Time Series Models. Journal of the American Statistical Association, 82, 590-604.
- Tsay, Ruey S.** 1988: Outliers, Level Shifts, and Variance Changes in Time Series. Journal of Forecasting, 7, 1-20.
- Vandaele, Walter** 1983:  
Applied Time Series and Box-Jenkins Models. Orlando usw.: Academic Press.
- Wallis, Kenneth F.** 1987: Time Series Analysis of Bounded Economic Variables. Journal of Time Series Analysis, 8, 115-123.
- Wei, William S.** 1990:  
Time Series Analysis. Univariate and Multivariate Methods. Redwood City usw.: Addison-Wesley.
- Wichern, Dean W.** 1973:  
The Behaviour of the Sample Autocorrelation Function for an Integrated Moving Average Process. Biometrika, 60, 235-239.
- Wilson, G. Tunncliffe** 1973:  
Discussion of Paper by Dr. Chatfield and Dr. Prothero. Journal of the Royal Statistical Society A, 136, 315-319.