

2 Zeitreihenanalyse

2.1	Problemstellung	120
2.2	Vorgehensweise	124
2.2.1	Visualisierung der Zeitreihe	125
2.2.2	Formulierung eines Modells	126
2.2.3	Schätzung des Modells	127
2.2.4	Erstellung von Prognosen	128
2.2.4.1	Punktprognose	128
2.2.4.2	Prognosefehler	128
2.2.4.3	Prognoseintervall	129
2.2.5	Prüfung der Prognosegüte	131
2.3	Nichtlineare Zeitreihenmodelle	132
2.3.1	Nichtlineare Trendmodelle	132
2.3.1.1	Das Quadratwurzel-Modell	133
2.3.1.2	Das Logarithmische Modell	135
2.3.1.3	Das Multiplikative Modell	136
2.3.1.4	Das Potenz-Modell	137
2.3.1.5	Vergleich der Modelle	137
2.3.2	Berücksichtigung von Strukturbrüchen	138
2.3.3	Berücksichtigung von zyklischen Schwankungen	139
2.3.4	Umsetzung mit SPSS	140
2.3.4.1	Lineare Regression	140
2.3.4.2	Kurvenanpassung	141
2.4	Fallbeispiel	143
2.4.1	Problemstellung	143
2.4.2	Ergebnisse	144
2.4.2.1	Extrapolationsmodelle	144
2.4.2.2	Strukturmodelle	146
2.4.2.2.1	Einbeziehung der Temperatur	146
2.4.2.2.2	Einbeziehung der Bevölkerungsentwicklung	148
2.4.2.3	Zusammenfassung	150
2.4.3	SPSS-Kommandos	151
2.5	Anwendungsempfehlungen	151
2.6	Mathematischer Anhang	153
	Literaturhinweise	154

2.1 Problemstellung

Bei der Datenanalyse wird generell zwischen der Analyse von Querschnittsdaten, die zu einem Zeitpunkt bei verschiedenen Untersuchungsobjekten erhoben wurden, und Zeitreihendaten (Längsschnittdaten), die zu verschiedenen Zeitpunkten erhoben wurden, unterschieden, wobei sich diese auch kombinieren lassen. Während die in diesem Buch behandelten Methoden primär der Querschnittsanalyse dienen, soll in diesem Kapitel gesondert auf die Zeitreihenanalyse eingegangen werden. Sie bildet ein wichtiges Anwendungsgebiet der im vorstehenden Kapitel behandelten Regressionsanalyse.

Prognose Die Zeitreihenanalyse dient neben der Beschreibung und Erklärung der zeitlichen Entwicklung einer Variablen Y insbesondere auch deren Prognose, d. h. der Schätzung von Werten dieser Variablen für zukünftige Zeitpunkte oder Perioden. Jede weitreichende Entscheidung basiert auf Prognosen. Die Zeitreihenanalyse ist daher für die Stützung von Entscheidungsproblemen jeglicher Art von großer Wichtigkeit. Abb. 2.1 zeigt beispielhaft einige typische Fragestellungen, die primär wirtschaftliche Probleme betreffen. Nicht alle diese Fragestellungen können hier behandelt werden. Vielmehr soll nur ein Einblick in Grundlagen der Zeitreihenanalyse mit Hilfe des Instrumentariums der Regressionsanalyse gegeben werden.

Zeitreihe Eine *Zeitreihe* ist formal eine Menge von Werten einer Variablen Y (z. B. Käuferzahl, Absatzmenge, Volkseinkommen, Bevölkerung, Wetter), die im Zeitablauf gemessen wurden und gemäß der Zeit geordnet sind. Sie lässt sich ausdrücken durch

$$\langle y_1, y_2, y_3, \dots, y_t, \dots, y_T \rangle \quad (2.1)$$

wobei der Index t (z. B. $t = 1, 2, 3, \dots, T$) eine Abbildung der Zeit bildet. Nach dem Wertevorrat von t lässt sich zwischen stetigen und diskreten Zeitreihen unterscheiden. Hier sollen nur diskrete Zeitreihen betrachtet werden.¹ Der Index t bezeichnet bzw. zählt in diesem Fall äquidistante Zeitpunkte oder Perioden, in die die Zeit eingeteilt wird (z. B. Tage, Wochen, Monate). Durch $(1 : T)$ sei der *Beobachtungszeitraum* der Zeitreihe bezeichnet.

Zeitvariable Die *Zeit* kann als eine Variable aufgefasst werden, die sich im Gegensatz zu anderen Variablen völlig gleichförmig und unabhängig von allem anderen Geschehen entwickelt.² Sie hat eine ordnende Funktion auf Ereignisse, indem sie diese in eine feste und unabänderliche Reihenfolge bringt. Bei Querschnittsdaten dagegen spielt die Reihenfolge der Daten keine Rolle und kann beliebig verändert werden.

Nach dem zeitlichen Bezug der Variablen Y unterscheidet man

- *zeitpunkt-bezogene Variablen*: Zustands- oder Bestandsgrößen, die punktuell gemessen werden, z. B. an einem Stichtag (Lagerbestand am Monatsende, Finanzvermögen am Jahresende),
- *zeitraum-bezogene Variablen*: Strömungsgrößen, die über einen Zeitraum gemessen werden (Verbrauch pro Tag, Umsatz pro Monat, Einkommen pro Jahr).

¹Stetige Zeitreihen werden z. B. durch automatische Aufzeichnung physikalischer oder meteorologischer Prozesse (z. B. Temperatur, Luftdruck) erzeugt.

²Der klassische Zeitbegriff wurde durch Isaak Newton (1643 - 1727) geprägt: „Absolute, true, and mathematical time, from its own nature, passes equably without relation to anything external“. Seit Albert Einstein (1879 - 1955) wissen wir zwar, dass auch die Zeit nicht unabhängig und konstant verläuft, sondern sich mit zunehmender Geschwindigkeit verlangsamt und bei Lichtgeschwindigkeit sogar zum Stillstand kommt. Für die uns hier interessierenden Phänomene aber können wir dies vernachlässigen.

- Ist das Volumen des Margarinemarktes über die Zeit konstant (abgesehen von zufälligen Schwankungen) oder besteht ein Trend (Wachstum oder Schrumpfung)?³
- Gibt es im Konsum von Margarine saisonale Schwankungen?⁴
- Hat das Wetter (z. B. die Temperatur) einen Einfluss auf den Margarinekonsum?⁵
- Wurde durch die Änderung der Verpackung eines Produktes ein Strukturbruch in der Entwicklung des Absatzvolumens bewirkt?⁶
- Lässt sich aus den Absatzzahlen eines neuen Produktes über die ersten 12 Monate nach der Markteinführung erkennen, welchen Marktanteil das Produkt im zweiten oder dritten Jahr erreichen wird und ob es am Markt erfolgreich sein wird?⁷
- Ein Automobilhersteller möchte herausfinden, ob der GfK-Konsumklima-Index oder der Ifo-Geschäftsklima-Index geeignete Frühindikatoren zur Prognose seines Absatzvolumens sind.⁸
- Ein Hersteller der Kosmetik-Branche möchte herausfinden, ob seine Produktwerbung zeitlich verzögert wirkt, und wenn ja, mit welchem Time-Lag.⁹
- Anfang der 90er Jahre begann die Ausbreitung von zwei Innovationen, die seitdem unser Leben erheblich verändert haben: Das Internet (Einführung des WWW) und das Mobiltelefon (Handy). Wie hat sich die Nutzung seitdem in Deutschland entwickelt und welche zukünftige Entwicklung ist zu erwarten?¹⁰

Abbildung 2.1: Typische Fragestellungen der Zeitreihenanalyse

Ein Problem bei Zeitreihenanalysen resultiert daraus, dass viele der üblicherweise verwendeten Perioden, wie Monate, Quartale oder Jahre, nicht wirklich äquidistant sind und daher eine Bereinigung der Zeitreihe notwendig machen können.¹¹ Die Länge eines Monats variiert zwischen 28 und 31 Tagen, d. h. die Länge des Februars in einem Normaljahr beträgt nur 90,32 % der Länge eines Monats mit 31 Tagen (z. B. Juli oder August). Entsprechend variieren die Werte von monatlich oder auch quartalsmäßig gemessenen Strömungsgrößen bei konstanter Strömung (z. B. Energieerzeugung, Fördermenge von Öl). Für ökonomische Analysen ist neben der Zahl der Kalendertage auch die Zahl der Werk- oder Arbeitstage relevant, die wiederum durch gesetzliche Feiertage eingeschränkt werden. Hierdurch können sich auch Schwankungen bei

Äquidistanz

³Siehe dazu das Fallbeispiel in diesem Kapitel.

⁴Siehe dazu das Fallbeispiel in diesem Kapitel.

⁵Siehe dazu das Fallbeispiel in diesem Kapitel.

⁶z. B. Schneeberger (1994).

⁷z. B. Mahajan/Muller/Bass (1990), Mertens, P. (Hrsg.) (1994).

⁸z. B. Backhaus/Simon (1981), Niederhübner (1994).

⁹z. B. Lambin (1969), Palda (1984).

¹⁰Siehe dazu Kapitel 10.

¹¹Siehe hierzu Rinne/Specht (2002), S. 33 ff.

wöchentlich gemessenen Werten ergeben. Von geringerer Bedeutung ist dagegen die unterschiedliche Länge zwischen Normaljahren und Schaltjahren.

Quantitative
Prognoseverfahren

Prognoseverfahren, die sich auf Zeitreihenanalysen stützen, werden als *quantitative Prognoseverfahren* bezeichnet.¹² Sie beinhalten im Kern ein mathematisches Modell des Prozesses, der die Zeitreihe erzeugt hat. Im Unterschied dazu bezeichnet man Prognoseverfahren, die sich nicht auf ein mathematisches Modell stützen, als *qualitative Prognoseverfahren*. Sie basieren primär auf den subjektiven Einschätzungen von Experten (z. B. Delphi-Methode, Szenario-Technik).¹³

Qualitative
Prognoseverfahren

Nach der Anzahl der Zeitreihen, auf die sich das Prognoseverfahren stützt, lassen sich die quantitativen Prognoseverfahren in zwei Gruppen einteilen.

Quantitative Prognoseverfahren

a) Zeitreihenextrapolation

Entwicklungs-
prognose

Die Zeitreihenextrapolation basiert auf der Analyse einer einzelnen Zeitreihe. Es wird nach einem Muster in der Entwicklung der Zeitreihendaten gesucht und durch Extrapolation dieses Musters in die Zukunft werden Prognosen abgeleitet (sog. *Entwicklungsprognosen*). Dabei wird die Annahme gemacht, dass das Muster stabil bleibt. Die hierbei verwendeten Prognoseverfahren werden auch als *Extrapolationsverfahren* bezeichnet. Sie lassen sich grob in drei Gruppen einteilen:

- Zeitregression
- Glättungsmethoden (z. B. gleitende Durchschnitte, Exponentielle Glättung)
- Autoregressive Methoden (z. B. ARIMA-Prozesse nach Box/Jenkins, adaptive Filter)¹⁴

Hier soll nur die Zeitregression behandelt werden. Im Kern beinhaltet sie ein deterministisches Modell, welches von der Zeit abhängig ist und das auf Basis der Zeitreihendaten zu schätzen ist. Zwecks Erstellung von Prognosen sind entsprechende Zeitwerte in das geschätzte Modell einzusetzen. Die Modelle lassen sich daher relativ einfach handhaben und sind auch dem Verständnis des Benutzers besser zugänglich, als viele andere Verfahren, bei denen die Modellierung der stochastischen Komponente einer Zeitreihe im Vordergrund steht. Weitere Vorteile sind, dass die Zeitregression auch angewendet werden kann, wenn nur verhältnismäßig wenige Zeitreihendaten vorliegen, und dass sie sowohl für kurzfristige wie auch mittel- und langfristige Prognosen verwendet werden kann.

b) Kausale Prognoseverfahren

Kausale Prognoseverfahren basieren auf der Analyse der Beziehungen zwischen mehreren Zeitreihen. Es wird dazu ein Modell spezifiziert, welches die kausalen Beziehungen zwischen den Zeitreihenvariablen abbildet und welches sodann auf Basis der Zeitreihendaten zu schätzen ist.

¹²Zu Zeitreihenanalysen und Prognoseverfahren vgl. z. B. Hamilton (1994), Hanke/Reitsch (1998), Makridakis/Wheelwright/Hyndman (1998), Mertens, P./Rässler, S. (Hrsg.) (2004), Rinne/Specht (2002).

¹³Einen Überblick über quantitative wie qualitative Prognoseverfahren geben z. B. Armstrong, J. (Hrsg.) (2002), Hanke/Reitsch (1998), Makridakis/Wheelwright/Hyndman (1998).

¹⁴SPSS bietet unter dem Menüpunkt „Analysieren/Zeitreihen“ Prozeduren zur Schätzung von Modellen der exponentiellen Glättung und von ARIMA-Modellen (Autoregressive Integrated Moving Average Models) an, auf die hier nicht eingegangen wird.

Man spricht hier von *Strukturmodellen* oder *ökonomischen Modellen*.¹⁵ Man unterscheidet:

- Eingleichungsmodelle
- Mehrgleichungsmodelle

In Eingleichungsmodellen wird die Abhängigkeit der Prognosevariablen Y von einer oder mehreren Prädiktorvariablen modelliert (z. B. die Abhängigkeit der Pkw-Nachfrage vom Preis, den Werbeausgaben, dem verfügbaren Einkommen der Bevölkerung und dem Konsumklima). Im Gegensatz zu Entwicklungsprognosen spricht man hier auch von *Wirkungsprognosen*. Soweit es sich bei den Prädiktorvariablen um nicht kontrollierte Variablen handelt (wie hier dem verfügbaren Einkommen und dem Konsumklima), erfordert die Prognose von Y , dass zunächst Prognosen der Prädiktorvariablen zu erstellen sind, was die Anwendung erschwert. Dies kann u.U. erleichtert werden, wenn die Entwicklung der Prädiktorvariablen gegenüber der abhängigen Variable zeitlich voraus läuft. So ist z. B. der GfK-Konsumklima-Index ein *vorauseilender Indikator* für die Pkw-Nachfrage oder die Zahl der Neuzulassungen von Pkws für die Nachfrage nach Autozubehör.¹⁶

Wirkungsprognosen

Mehrgleichungsmodelle werden erforderlich, wenn wechselseitige Abhängigkeiten zwischen den Variablen bestehen und modelliert werden sollen. So bestehen z. B. Wechselwirkungen zwischen dem Volkseinkommen und privatem Konsum oder zwischen der Absatzmenge eines Produktes und seinem Bekanntheitsgrad.

Wir wollen uns hier auf Eingleichungsmodelle beschränken. In Eingleichungsmodellen lassen sich Entwicklungs- und Wirkungsprognosen auch kombinieren. Zur Schätzung der Modelle kann die im vorstehenden Kapitel behandelte Regressionsanalyse verwendet werden.

Anwendungsbeispiel

Die Analyse einer einzelnen Zeitreihe soll zunächst an einem kleinen Beispiel demonstriert werden. Der Manager des hier betrachteten Margarineherstellers ist etwas überrascht, aber auch erfreut über den hohen Absatz im Verkaufsgebiet 1 (vgl. Abbildung 1.6 in Kapitel 1). Falls das hohe Niveau anhalten oder sich gar weiter erhöhen sollte, müsste er allerdings eine Änderung hinsichtlich der Belieferung dieses Gebietes vornehmen. Für seine Entscheidungsfindung benötigt er eine Analyse und Prognose der Absatzentwicklung in diesem Verkaufsgebiet. Zu diesem Zweck besorgt er sich die Absatzdaten der letzten 10 Jahre, die in Abbildung 2.2 wiedergegeben sind.

Anwendungsbeispiel

¹⁵Vgl. hierzu die ökonomische Literatur, z. B. Greene, Hanssens/Parsons/Schultz (2001), Kmenta (1997), Ramanathan (1998), Schneeweiß (1990), Studenmund (2001), Wooldridge (2005).

¹⁶Vgl. z. B. Backhaus/Simon (1981), Niederhübner (1994).

Zeit (Periode t)	Menge (Kartons pro Periode)
1	1.657
2	1.864
3	1.950
4	2.204
5	2.288
6	2.410
7	2.414
8	2.534
9	2.739
10	2.785

Abbildung 2.2: Zeitreihe der Absatzmenge eines Verkaufsgebietes

2.2 Vorgehensweise

Die allgemeine Vorgehensweise bei der Analyse und Prognose einer Zeitreihe lässt sich in fünf Schritte gliedern (vgl. Abbildung 2.3).



Abbildung 2.3: Ablaufschritte der Analyse und Prognose einer Zeitreihe

Am Beginn der Analyse sollte immer die Visualisierung der Zeitreihendaten stehen, die unerlässlich für die Formulierung eines geeigneten Modells ist. Es ist wichtig, dass sich der Untersucher vom Verlauf der Zeitreihe ein „Bild“ macht, was bei Betrachtung einer langen Zahlenkolonne kaum möglich ist.

Eine Visualisierung der zu analysierenden Daten ist zwar für jede Art der Datenanalyse von Vorteil, doch stößt man gerade bei der multivariaten Datenanalyse schnell an Grenzen, da sich maximal drei Dimensionen darstellen lassen. Daher ist es ein Ziel vieler multivariater Analysemethoden, durch Analyse und Komprimierung der Daten eine Visualisierung zu ermöglichen. Bei Zeitreihendaten dagegen bietet sich immer eine Visualisierung über die Zeit an.

Die Formulierung des Modells sollte sich auf das „Bild“ stützen, welches sich der Untersucher von der Zeitreihe gemacht hat, aber auch auf weitere sachlogische oder theoretische Überlegungen. Es existiert ein umfangreiches Arsenal von Modelltypen, auf welches man bei der Modellformulierung zurückgreifen kann. Oft ist es zweckmäßig, alternative Modelle zu erproben oder miteinander zu kombinieren.

Im Zentrum des Ablaufschemas steht die Durchführung einer Regressionsanalyse, bezüglich derer wir uns hier auf das vorstehende Kapitel stützen können. Hieran

schließt sich die Verwendung des geschätzten Modells zur Erstellung von Prognosen an. Durch Gegenüberstellung dieser Prognosen mit der Realität erfolgt schließlich die Prüfung der Prognosegüte des Modells bzw. die Modellvalidierung.

2.2.1 Visualisierung der Zeitreihe

- 1 **Visualisierung der Zeitreihe**
- 2 Formulierung eines Modells
- 3 Schätzung des Modells
- 4 Erstellung von Prognosen
- 5 Prüfung der Prognosegüte

Für die Visualisierung von Zeitreihendaten bieten Programme wie Excel oder SPSS vielfältige Funktionen an (Streudiagramme, Liniendiagramme, Balkendiagramme). Abbildung 2.4 zeigt die Absatzdaten aus Abbildung 2.2 als Streudiagramm. Häufig werden die Streupunkte durch Linien verbunden, sodass sich ein Polygonzug wie in Abbildung 2.18 ergibt. Die Darstellung einer

Visualisierung von
Zeitreihendaten

Zeitreihe als Balkendiagramm zeigt z. B. Abbildung 2.19.

Das Streudiagramm zeigt einen deutlich ansteigenden Verlauf der Absatzmenge über die Zeit. Es lässt aber auch erkennen, dass der Verlauf nicht ganz linear, sondern leicht konkav gebogen ist. Das Wachstum der Absatzmenge scheint sich im Verlauf der Zeit etwas abzuschwächen.

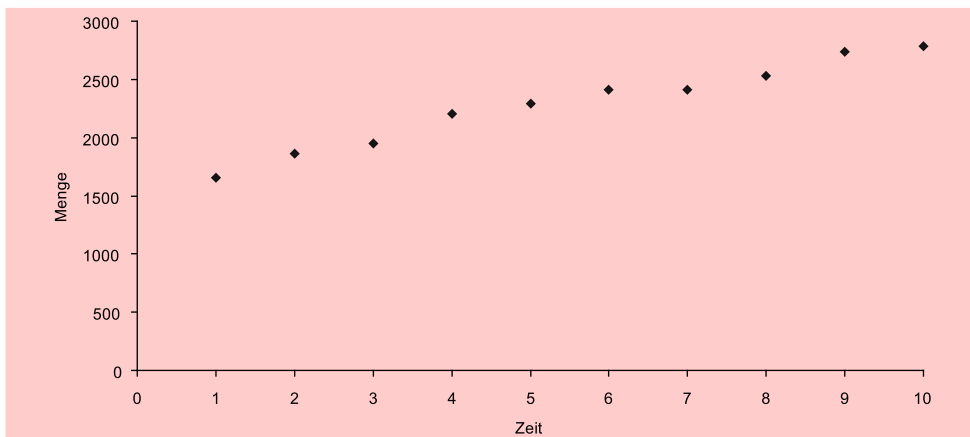
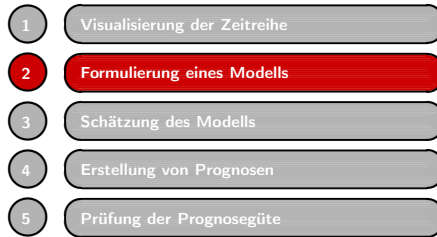


Abbildung 2.4: Streudiagramm der Zeitreihendaten

2.2.2 Formulierung eines Modells



Zeitreihenverläufe können höchst unterschiedliche Formen annehmen und es existiert eine zahllose Menge von Modelltypen, die sich an den Verlauf einer Zeitreihe mehr oder weniger gut anpassen lassen. Die Wahl des „richtigen“ Modells bildet sicherlich die wichtigste, aber auch schwierigste Entscheidung bei der Anwendung quantitativer Prognoseverfahren.

Zeitreihenzerlegung

Ein Grundprinzip der Modellierung von Zeitreihen bildet die *Zeitreihenzerlegung*, d. h. die Zerlegung einer Zeitreihe in unterschiedliche Komponenten. Man unterscheidet:

$$\text{Additive Zeitreihenzerlegung:} \quad Y = A + K + S + u \quad (2.2)$$

$$\text{Multiplikative Zeitreihenzerlegung:} \quad Y = A \cdot K \cdot S \cdot u \quad (2.3)$$

mit

Y = Prognosevariable mit den Werten $\langle y_1, y_2, y_3, \dots, y_t, \dots, y_T \rangle$
 A = Trendkomponente
 K = Konjunkturkomponente
 S = Saisonkomponente
 u = zufällige Komponente (Störgröße)

Systematische Komponente

Die Trendkomponente A repräsentiert die langfristige Entwicklung der Größe Y . Sie kann positiv (Wachstum) oder negativ (Schrumpfung), linear oder nichtlinear sein. K und S sind zyklische Schwankungen, die im Zeitraum von mehreren Jahren (Konjunktur) oder einem Jahr (Saison) periodisch wiederkehren. Sie verlaufen zwangsläufig nichtlinear. A , K und S werden gemeinsam auch als *systematische Komponente* bezeichnet, im Gegensatz zur zufälligen Komponente u .

Die vorliegende Zeitreihe lässt keine zyklische Komponente erkennen. Modell (2.2) reduziert sich damit auf

$$Y = A + u \quad (2.4)$$

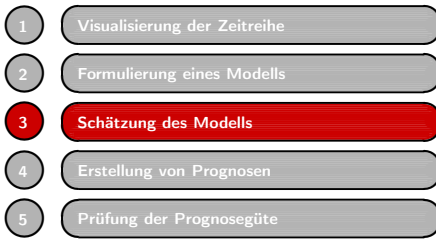
Die Trendkomponente A muss nun spezifiziert werden. Wir wollen mit einem einfachen Modell beginnen und nehmen einen linearen Trendverlauf an, obgleich wir einen leicht nichtlinearen Verlauf der Zeitreihe erkannt haben. Wir erhalten damit das *lineare Trendmodell*:

Lineares Trendmodell

$$Y = \alpha + \beta \cdot t + u \quad (2.5)$$

α und β sind unbekannte Parameter, die auf Basis der vorliegenden Daten zu schätzen sind. Der *Trendparameter* β gibt den Zuwachs von Y pro Periode an. Im linearen Trendmodell wird dieser Zuwachs als konstant angenommen.

2.2.3 Schätzung des Modells



Zur Schätzung der unbekannt Parameter α und β auf Basis der vorliegenden Daten verwenden wir die Regressionsanalyse. Dazu ist der Zeitindex mit den Werten $t = 1, 2, 3, \dots$ als unabhängige Variable zu verwenden. Anstelle der Indexzahlen 1, 2, 3, ... könnten hier auch Jahreszahlen (z. B. 1991, 1992, 1993, ...) stehen, was auf den Schätzwert für den Trendparameter β keinen Einfluss hat (es würde sich nur der Schätzwert für das konstante Glied α verändern).

Regressionsanalyse

Die Durchführung der Regressionsanalyse liefert die folgende Regressionsfunktion:

$$\hat{Y} = a + b \cdot t = 1.619,5 + 120,9 \cdot t \quad (R^2 = 0,972)$$

Der Schätzwert $b = 120,9$ besagt hier, dass die Absatzmenge im Verkaufsgebiet pro Periode um rund 121 Kartons anwächst. Die geschätzte Funktion ist in Abbildung 2.5 zusammen mit den beobachteten Werten dargestellt. Die beobachteten Werte weichen nur geringfügig von der geschätzten Funktion ab.

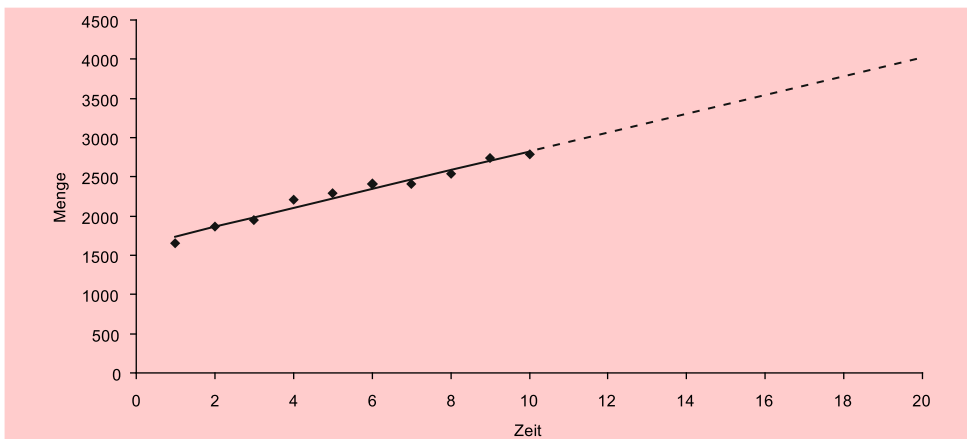


Abbildung 2.5: Lineares Trendmodell: Geschätzte Funktion und prognostizierter Verlauf

Zur Beurteilung der Güte der geschätzten Regressionsfunktion liefert die Regressionsanalyse u.a. die folgenden Statistiken (vgl. dazu die Ausführungen in Kapitel 1):

Güte der Schätzung

Bestimmtheitsmaß (R-Quadrat): $R^2 = 0,972$

F-Statistik: $F_{emp} = 276,8$

Standardfehler der Regression: $s = 66,0$

Durbin/Watson-Statistik: $d = 1,551$

Das Bestimmtheitsmaß sagt uns, dass die Zeitvariable t die Streuung in den Absatzdaten zu rund 97%, also fast vollständig, erklärt. Das lässt vermuten, dass die Marketingaktivitäten in dem betreffenden Verkaufsgebiet in der Vergangenheit relativ konstant waren. Der Wert der F-Statistik weist auf hohe Signifikanz des Modells

hin (der tabellarische F-Wert beträgt hier lediglich 5,32 bei einer Vertrauenswahrscheinlichkeit von 95 %). Der Standardfehler besagt, dass die mittlere Abweichung (Standardabweichung) zwischen der geschätzten Regressionsfunktion und den beobachteten Absatzmengen 66 Kartons beträgt. Der Wert der Durbin/Watson-Statistik, der im Idealfall $d=2$ beträgt, weist hier nicht auf das Vorliegen von Autokorrelation hin.

2.2.4 Erstellung von Prognosen



Mit Hilfe der geschätzten Funktion (2.6) ist es möglich, ohne weitere Informationen Prognosen zu erstellen. Der Prognosewert für die in der Zukunft liegende Periode $T+k$ ergibt sich durch Funktion (2.7).

Man spricht in diesem Fall von einer *Punktprognose*. Im Unterschied dazu gibt eine *Intervallprognose* einen Bereich um den Prognosewert \hat{y}_{T+k} an, in dem der zukünftige

wahre Wert mit einer bestimmten Vertrauenswahrscheinlichkeit bzw. Konfidenz (z. B. 95 %) liegen wird.

$$\hat{Y} = a + b \cdot t \tag{2.6}$$

$$\hat{y}_{T+k} = a + b \cdot (T + k) \tag{2.7}$$

2.2.4.1 Punktprognose

Mittels (2.6) erhält man für die folgende Periode $T+1 = 11$ den Prognosewert

$$\hat{y}_{11} = 1.619,5 + 120,9 \cdot 11 = 2.949$$

Analog lassen sich Prognosen für beliebige Perioden erstellen. Für Periode 20 erhält man z. B.

$$\hat{y}_{20} = 1.619,5 + 120,9 \cdot 20 = 4.038$$

In Abbildung 2.5 ist zusätzlich zur geschätzten Funktion im Beobachtungsintervall (1 : 10), der auch als *Stützbereich* des Prognosemodells bezeichnet wird, der prognostizierte Verlauf der Absatzmenge für die Perioden 11 bis 20 als gestrichelte Linie dargestellt.

2.2.4.2 Prognosefehler

Prognosen sind leider immer mit Fehlern verbunden („besonders wenn sie auf die Zukunft gerichtet sind“¹⁷). Die Basis für die Berechnung des Prognosefehlers bildet der Standardfehler der Regressions-schätzung (vgl. Abschnitt in Kapitel 1), der sich aus den Residuen $e_t = y_t - \hat{y}_t$ ergibt. Der *Standardfehler* errechnet sich durch:

$$s = \sqrt{\frac{1}{T-2} \sum_{t=1}^T e_t^2} = 66,0 \tag{2.8}$$

¹⁷Die Bemerkung stammt vermutlich von Hermann Josef Abs (1901-1994), dem legendären Bankier und langjährigen Vorstandsvorsitzenden der Deutschen Bank.

Es ist einsichtig, dass der Prognosefehler um so größer sein wird, je weiter die Prognose in die Zukunft reicht. Um dies zu berücksichtigen, ist der Standardfehler der Schätzung mit einem Faktor > 1 zu multiplizieren, dessen Größe vom Prognosehorizont $T+k$ abhängig ist. Für den Fehler einer Prognose für Periode $T+k$ gilt:¹⁸

Prognosefehler

Prognosefehler

$$s_p(T+k) = s \cdot \sqrt{1 + \frac{1}{T} + \frac{(T+k-\bar{t})^2}{\sum_t (t-\bar{t})^2}} = s \cdot \sqrt{1 + \frac{1}{T} + \frac{(T+k-\bar{t})^2}{(T-1) \cdot s_t^2}} \quad (2.9)$$

mit

s = Standardfehler der Regression

\bar{t} = Mittelwert der Zeitvariablen t

s_t = Standardabweichung der Zeitvariablen t

Mit den Werten $s = 66$, $\bar{t} = 5,5$ und $s_t = 3,028$ ergibt sich hier für den Prognosewert $\hat{y}_{11} = 2.949,4$ der Prognosefehler¹⁹

$$\begin{aligned} s_p(11) &= 66 \cdot \sqrt{1 + \frac{1}{10} + \frac{(10+1-5,5)^2}{9 \cdot 3,028^2}} = 66 \cdot \sqrt{1 + 0,1 + 0,37} \\ &= 66 \cdot 1,21 = 80 \end{aligned}$$

Interpretation

Interpretation: Gemäß dem hier zugrunde gelegten Modell (2.5) bildet die Absatzmenge Y eine Zufallsgröße, deren Verteilung durch die Störgröße u bedingt ist. Die Abweichungen zwischen dem prognostizierten Wert $\hat{y}_{11} = 2.949,4$ und dem realisierten Wert y_{11} würde, falls wiederholte Realisierungen möglich wären, im Durchschnitt 80 Kartons betragen.

Für Periode 20 hatten wir den Wert $\hat{y}_{20} = 4.037,5$ prognostiziert. Analog ergibt sich hier der Prognosefehler

$$\begin{aligned} s_p(20) &= 66 \cdot \sqrt{1 + \frac{1}{10} + \frac{(10+10-5,5)^2}{9 \cdot 3,028^2}} = 66 \cdot \sqrt{1 + 0,1 + 2,55} \\ &= 66 \cdot 1,91 = 126 \end{aligned}$$

Da die Prognose viel weiter in die Zukunft reicht, ist auch der zugehörige Prognosefehler bedeutend größer. Aus (2.9) ist ersichtlich, dass der Prognosefehler mit dem Prognosehorizont $T+k$ anwächst. Am geringsten ist der Prognosefehler im Mittelpunkt der Zeitreihe bzw. des Stützbereiches, hier im Punkt $\bar{t} = 5,5$.

Prognosehorizont

2.2.4.3 Prognoseintervall

Um anzugeben, in welchem Bereich der noch unbekannte zukünftige Wert y_{11} mit einer bestimmten Wahrscheinlichkeit liegen wird, ist ein Prognoseintervall (Konfidenzintervall) zu erstellen:

Prognoseintervall

$$y_{T+k} = \hat{y}_{T+k} \pm t_{\alpha/2} \cdot s_p(T+k) \quad (2.10)$$

bzw.

$$\hat{y}_{T+k} - t_{\alpha/2} \cdot s_p(T+k) \leq y_{T+k} \leq \hat{y}_{T+k} + t_{\alpha/2} \cdot s_p(T+k)$$

¹⁸Vgl. z. B. Kmenta (1997), S. 251.

¹⁹Die Werte von \bar{t} und s_t lassen sich in SPSS mittels „Analysieren / Deskriptive Statistiken“ berechnen.

2 Zeitreihenanalyse

Dabei bezeichnet $t_{\alpha/2}$ das Quantil der t-Verteilung (Student-Verteilung) für die Vertrauenswahrscheinlichkeit $1 - \alpha$ bei zweiseitigem Test und $T-2$ Freiheitsgraden.

Für die Vertrauenswahrscheinlichkeit $1 - \alpha = 0,95$ und $T-2 = 8$ Freiheitsgrade lässt sich aus der t-Tabelle (Anhang A.1) der Wert $t_{\alpha/2} = 2,306$ entnehmen. Für die unbekannte Absatzmenge y_{11} ergibt sich damit das *Prognoseintervall*:

$$\begin{aligned}\hat{y}_{11} - t_{\alpha/2} \cdot s_p(11) &\leq y_{11} \leq \hat{y}_{11} + t_{\alpha/2} \cdot s_p(11) \\ 2.949 - 2,3 \cdot 80 &\leq y_{11} \leq 2.949 + 2,3 \cdot 80 \\ 2.765 &\leq y_{11} \leq 3.133\end{aligned}$$

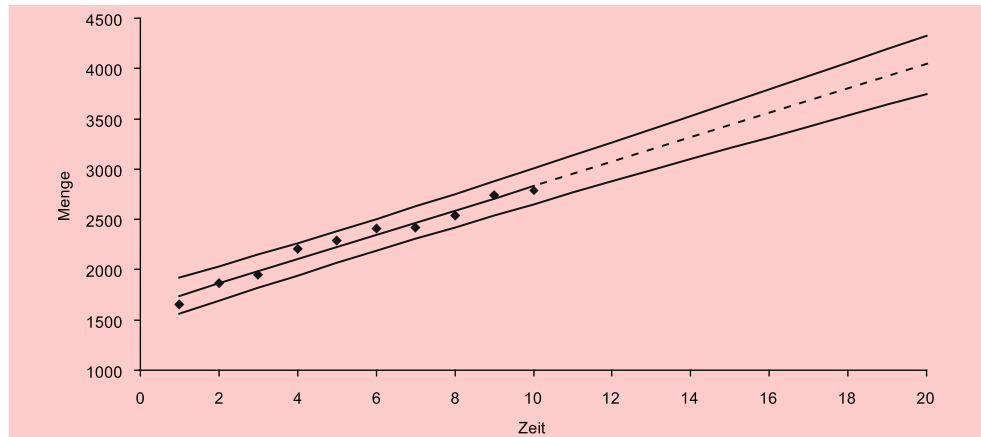


Abbildung 2.6: Lineares Trendmodell: Geschätzte Funktion, prognostizierter Verlauf und 95%-Prognoseintervall

Interpretation

Interpretation: Mit 95%iger Sicherheit ist zu erwarten, dass die Absatzmenge in der kommenden Periode zwischen 2.765 und 3.133 Kartons liegen wird. Ließe sich die Absatzmenge 100 Mal realisieren, so würde sie in 95 Fällen in diesen Bereich fallen.

Für Periode 20 hatten wir den Wert $\hat{y}_{20} = 4.037,5$ prognostiziert mit dem Prognosefehler $s_p(20) = 126$. Für das Prognoseintervall erhält man:

$$3.727 \leq y_{20} \leq 4.328$$

Abbildung 2.6 zeigt den Verlauf des Prognoseintervalls über die Zeit. Proportional zum Prognosefehler wächst auch das Prognoseintervall mit k an. Am geringsten ist es, ebenso wie der Prognosefehler, im Mittelpunkt der Zeitreihe bzw. des Stützbereiches.

2.2.5 Prüfung der Prognosegüte

- 1 Visualisierung der Zeitreihe
- 2 Formulierung eines Modells
- 3 Schätzung des Modells
- 4 Erstellung von Prognosen
- 5 **Prüfung der Prognosegüte**

Die üblichen Gütemaße der Regressionsanalyse (Bestimmtheitsmaß, F-Statistik, Standardfehler) sagen nur etwas darüber aus, wie gut die Anpassung des geschätzten Modells an die beobachteten Werte im Stützzeitraum (1:T) ist. Hierauf basieren auch die Berechnung von Prognosefehler und Prognoseintervall. Zur Prüfung der Prognosegüte und damit der Validierung des Modells

Validierung

ist es dagegen erforderlich, die prognostizierten Werte \hat{y}_{T+k} mit den realisierten Werten außerhalb des Stützbereiches im *Prognosebereich* (T+1:T+K) zu vergleichen.

Da die realisierten Werte im Zeitpunkt der Prognose nicht bekannt sind (sonst müsste man sie ja nicht prognostizieren), ist daher eine sofortige Prüfung der Prognosegüte nicht möglich. Vielmehr muss man warten, bis diese anfallen, was z. B. bei Jahresdaten recht lange dauern kann.

Eine Alternative besteht darin, die Prognosegüte durch *Ex-Post-Prognosen* zu prüfen. Dazu wird der Stützbereich für die Schätzung des Modells gegenüber dem Beobachtungszeitraum verkürzt, d. h. man verzichtet bei der Schätzung des Modells auf die letzten Zeitreihenwerte und vergleicht diese sodann mit den ex post prognostizierten Werten.

Ex-Post-Prognosen

Verkürzen wir hier den Stützbereich von zehn auf acht Perioden, so erhalten wir anstelle der Schätzung

$$\hat{Y} = 1.619,5 + 120,9 \cdot t \quad (R^2 = 0,953)$$

die folgende Funktion:

$$\hat{Y} = 1.610,5 + 123,3 \cdot t \quad (R^2 = 0,986)$$

Damit können wir ex post die Absatzmengen für die Perioden 9 und 10 prognostizieren. Die so prognostizierten Werte sind nachfolgend den realisierten Werten gegenübergestellt:

t	\hat{y}_t	y_t	$e_t = \hat{y}_t - y_t$
9	2.720	2.739	-19
10	2.843	2.785	58

Zur Beurteilung der *empirischen Prognosegüte* können die folgenden Fehlermaße berechnet werden:²⁰

Mittlerer absoluter Fehler (Mean Absolute Deviation)

MAD

$$MAD = \frac{1}{K} \sum_{k=1}^K |e_{T+k}| = \frac{19 + 58}{2} = 38,5 \quad (2.11)$$

²⁰Vgl. z. B. Hüttner (1994), S. 350 f., Rinne/Specht (2002), S. 134 f.

MAPE Mittlerer absoluter prozentualer Fehler (Mean Absolute Percentage Error)

$$MAPE = \frac{1}{K} \sum_{k=1}^K \left| \frac{e_{T+k}}{y_{T+k}} \right| \cdot 100 = \frac{0,0069 + 0,0208}{2} \cdot 100 = 1,4\% \quad (2.12)$$

mit

e_t = Prognoseabweichung (einfacher Prognosefehler), $e_t = \hat{y}_t - y_t$
 T = Ende des Stützbereichs, hier $T = 8$
 K = Prognosehorizont (Prognosedistanz), hier: $K = 2$

Der MAPE besitzt gegenüber dem MAD sowie weiteren gebräuchlichen Fehlermaßen den Vorteil, dass sich damit auch die Prognosegüte zwischen unterschiedlichen Zeitreihen vergleichen lässt, da er unabhängig von der Maßeinheit der Zeitreihen ist.

Theil's U Ein weiteres interessantes Fehlermaß bildet *Theil's U-Statistik*:²¹

$$U = \sqrt{\frac{\sum_{k=1}^K (\hat{y}_{T+k} - y_{T+k})^2}{\sum_{k=1}^K (y_{T+k} - y_{T+k-1})^2}} \quad (2.13)$$

mit

$$U = \begin{cases} < 1 & : \text{besser als naive Prognose} \\ = 1 & : \text{naive Prognose} \\ > 1 & : \text{schlechter als naive Prognose} \end{cases}$$

Die U-Statistik vergleicht die Prognosegüte eines geschätzten Modells mit der naiven Prognose, die darin besteht, den jeweils letzten Beobachtungswert als Prognose der folgenden Periode zu verwenden. Im Zähler von (2.13) steht der quadrierte Fehler der Prognose für Periode $T+k$ und im Nenner der entsprechende quadrierte Fehler der naiven Prognose. Man erhält hier:

$$\begin{aligned} U &= \sqrt{\frac{(\hat{y}_9 - y_9)^2 + (\hat{y}_{10} - y_{10})^2}{(y_9 - y_8)^2 + (y_{10} - y_9)^2}} = \sqrt{\frac{(2.720 - 2.739)^2 + (2.843 - 2.785)^2}{(2.739 - 2.534)^2 + (2.785 - 2.739)^2}} \\ &= \sqrt{\frac{(-19)^2 + (58)^2}{(205)^2 + (46)^2}} = \sqrt{\frac{3.725}{44.141}} = 0,29 \end{aligned}$$

Der erhaltene Wert besagt, dass das lineare Trendmodell die Absatzmenge deutlich besser prognostiziert als die naive Vorgehensweise.

2.3 Nichtlineare Zeitreihenmodelle

2.3.1 Nichtlineare Trendmodelle

Reale Trendverläufe sind nur selten linear. Das lineare Trendmodell bildet daher nur eine Approximation des wahren Verlaufs, die aber meist ausreichend ist, wenn es um kurzfristige Prognosen geht. Für langfristige Prognosen wird dagegen meist ein nichtlineares Trendmodell geeigneter sein.

²¹Vgl. z. B. Rinne/Spacht (2002), S. 137.

2.3.1.1 Das Quadratwurzel-Modell

Wie oben schon bemerkt wurde, lässt das Streudiagramm der Zeitreihe erkennen, dass sich das Wachstum der Absatzmenge im Verlauf der Zeit etwas abschwächt, d. h. der Verlauf ist nicht ganz linear, sondern leicht konkav gebogen. Dies gilt für zahlreiche ökonomische Zeitreihen oder andere Prozesse (auch Bäume wachsen bekanntlich nicht in den Himmel). Ein einfaches und häufig verwendetes nichtlineares Trendmodell, das ein sich abflachendes Wachstum bzw. einen degressiv ansteigenden Verlauf aufweist, ist das Quadratwurzel-Modell, dessen Verlaufsform in Abbildung 2.7 dargestellt ist.

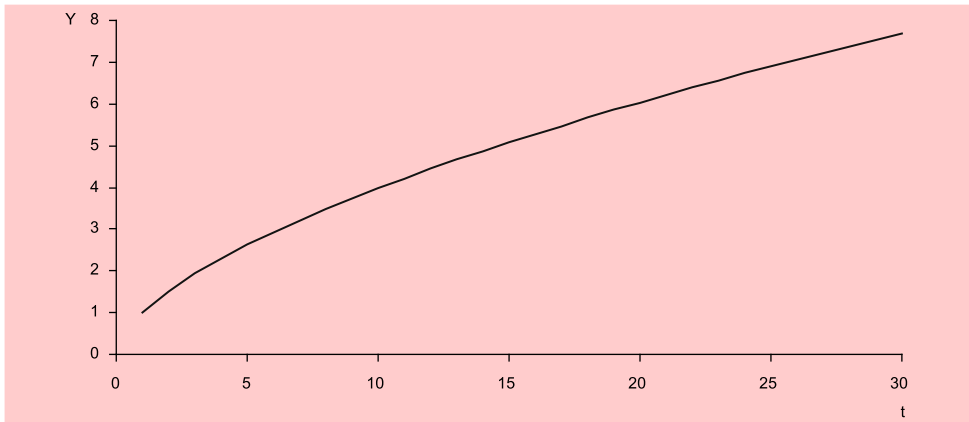


Abbildung 2.7: Quadratwurzel-Funktion $Y = \sqrt{t}$

Das *Quadratwurzel-Modell* lautet:

$$Y = \alpha + \beta \cdot \sqrt{t} + u \quad (2.14)$$

Das Modell ist zwar nichtlinear in Bezug auf die Zeitvariable t , aber linear in Bezug auf die zu schätzenden Parameter. Es handelt sich um ein sog. *intrinsisch lineares Modell*, das sich mit Hilfe der linearen Regressionsanalyse schätzen lässt. Zur Schätzung des Quadratwurzel-Modells kreieren wir eine neue Variable $X = \sqrt{t}$, deren Werte in Abbildung 2.8 aufgelistet sind.

Intrinsisch linear

Die Regression der Menge auf die transformierte Zeitvariable $X = \sqrt{t}$ liefert:

$$\hat{Y} = 1.117,1 + 519,57X = 1.117,1 + 519,57\sqrt{t} \quad (R^2 = 0,981)$$

Das Bestimmtheitsmaß lässt erkennen, dass die Anpassungsgüte des Quadratwurzel-Modells an die beobachteten Daten besser ist als die des linearen Modells. Der gute Fit ist auch in Abbildung 2.9 zu erkennen.

Die Punktprognose für Periode $T+1 = 11$ ergibt:

$$\hat{y}_{11} = 1.117,1 + 519,57\sqrt{11} = 2.840$$

und liegt um mehr als 100 Einheiten niedriger als beim linearen Modell. Noch erheblich größer ist der Unterschied bei der Prognose für Periode 20. Man erhält hier den Wert

$$\hat{y}_{20} = 1.117,1 + 519,57\sqrt{20} = 3.441$$

der fast 600 Einheiten niedriger liegt als der des linearen Modells.

Zeit (Periode t)	Menge (Kartons pro Periode)	$X = \sqrt{t}$
1	1.657	1,000
2	1.864	1,414
3	1.950	1,732
4	2.204	2,000
5	2.288	2,236
6	2.410	2,449
7	2.414	2,646
8	2.534	2,828
9	2.739	3,000
10	2.785	3,162

Abbildung 2.8: Zeitreihe der Absatzmenge und transformierte Zeitvariable

Prognosefehler

Analog zu (2.9) lässt sich der Prognosefehler nach nichtlinearer Transformation der Zeitvariablen t wie folgt berechnen:

$$s_p(T+k) = s \cdot \sqrt{1 + \frac{1}{T} + \frac{(x_{T+k} - \bar{x})^2}{\sum_t (x_t - \bar{x})^2}} = s \cdot \sqrt{1 + \frac{1}{T} + \frac{(x_{T+k} - \bar{x})^2}{(T-1) \cdot s_x^2}} \quad (2.15)$$

mit

- x_{T+k} = Wert der transformierten Zeitvariablen für Periode $T+k$
- s = Standardfehler der Regression
- \bar{x} = Mittelwert der transformierten Zeitvariablen
- s_x = Standardabweichung der transformierten Zeitvariablen

Es gilt hier $x_{T+k} = \sqrt{T+k}$ und damit $x_{11} = \sqrt{11} = 3,317$. Mit den Werten $s = 51,55$, $\bar{x} = 2,247$ und $s_x = 0,708$ erhält man hier für den Prognosewert $\hat{y}_{11} = 2.840$ den Prognosefehler

$$\begin{aligned} s_p(11) &= 51,55 \cdot \sqrt{1 + \frac{1}{10} + \frac{(3,317 - 2,247)^2}{9 \cdot 0,708^2}} \\ &= 51,55 \cdot \sqrt{1 + 0,1 + 0,254} = 51,55 \cdot 1,163 = 60 \end{aligned}$$

Mit $x_{20} = \sqrt{20} = 4,472$ erhält man für den Prognosewert $\hat{y}_{20} = 3.441$ den Prognosefehler

$$\begin{aligned} s_p(20) &= 51,55 \cdot \sqrt{1 + \frac{1}{10} + \frac{(4,472 - 2,247)^2}{9 \cdot 0,708^2}} \\ &= 51,55 \cdot \sqrt{1 + 0,1 + 1,097} = 51,55 \cdot 1,482 = 76 \end{aligned}$$

Die Prognosefehler des Quadratwurzel-Modells sind deutlich kleiner als die des linearen Modells.

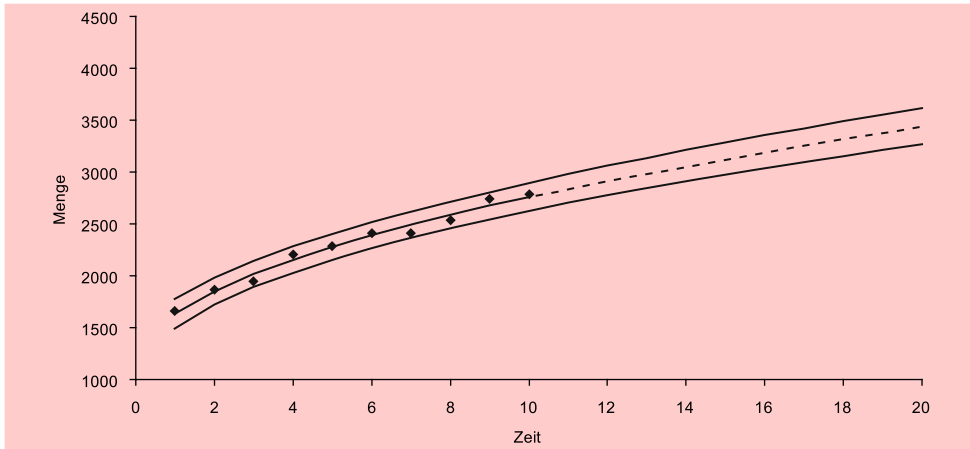


Abbildung 2.9: Quadratwurzel-Modell: Geschätzte Funktion, prognostizierter Verlauf und 95 %-Prognoseintervall

Prognoseintervall

Mit Hilfe der obigen Prognosefehler und $t_{\alpha/2} = 2,306$ (für die Vertrauenswahrscheinlichkeit $1 - \alpha = 0,95$ und $T-2=8$ Freiheitsgrade) erhält man die folgenden Prognoseintervalle für die Perioden 11 und 20:

$$\begin{aligned} 2.702 &\leq y_{11} \leq 2.979 \\ 3.265 &\leq y_{20} \leq 3.617 \end{aligned}$$

Abbildung 2.9 zeigt für das Quadratwurzel-Modell den Verlauf der Prognose und des Prognoseintervalls über die Zeit.

Prognoseintervall

2.3.1.2 Das Logarithmische Modell

Es existiert eine Vielzahl weiterer nichtlinearer Trendmodelle. Zu den Modellen, die wie das Quadratwurzel-Modell ein sich abflachendes Wachstum (einen degressiv steigenden Verlauf) aufweisen, gehört das logarithmische Modell

$$Y = \alpha + \beta \cdot \ln(t) + u \tag{2.16}$$

dessen Verlauf in Abbildung 2.10 dem Verlauf des Quadratwurzel-Modells gegenübergestellt ist. Das logarithmische Modell verläuft erheblich flacher als das Quadratwurzel-Modell. Die Schätzung des logarithmischen Modells liefert hier:

$$\hat{Y} = 1.540,6 + 492,56 \cdot \ln(t) \tag{R^2 = 0,945}$$

Die Anpassungsgüte an die beobachteten Daten ist in diesem Fall schlechter als die des linearen Modells.

Degressiv steigend

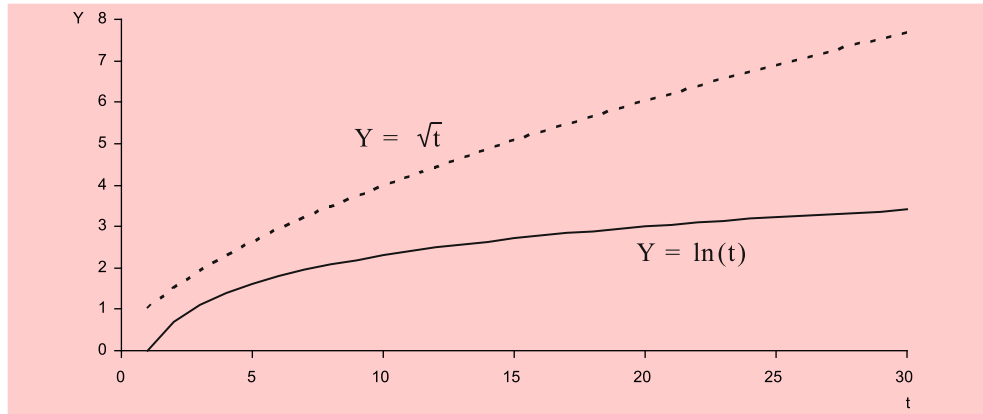


Abbildung 2.10: Logarithmische Funktion $Y = \ln(t)$ und Quadratwurzel-Funktion (gestrichelte Linie)

2.3.1.3 Das Multiplikative Modell

Ein sehr flexibles und ebenfalls häufig verwendetes nichtlineares Modell bildet das *multiplikative Modell* (vgl. Formel 1.26):

$$Y = \alpha \cdot t^\beta \cdot u \tag{2.17}$$

Das multiplikative Modell kann, je nach der Größe von β , unterschiedliche Verläufe annehmen (vgl. Abbildung 2.11):

- (a) $\beta > 1$: progressiv steigender Verlauf (konvexer Verlauf)
- (b) $0 < \beta < 1$: degressiv steigender Verlauf (konkaver Verlauf)
- (c) $\beta < 0$: degressiv fallender Verlauf

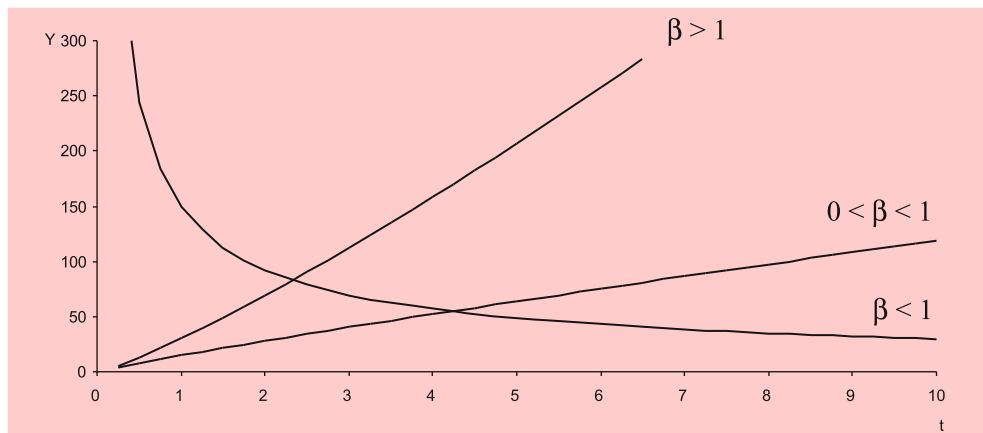


Abbildung 2.11: Multiplikatives Modell: Funktionsverläufe für unterschiedliche Werte von β

Linearisierte Form

Durch Logarithmierung des multiplikativen Modells erhält man die linearisierte Form:

$$\ln(Y) = \ln(\alpha) + \beta \cdot \ln(t) + \ln(u)$$

Die Schätzung mittels linearer Regression für die vorliegenden Absatzmengen liefert:

$$\ln(\hat{Y}) = 7,378 + 0,228 \cdot \ln(t) \quad (R^2 = 0,971)$$

Nach Delogarithmieren erhält man damit die folgende Funktion:

$$\hat{Y} = 1.600 \cdot t^{0,228} \quad (R^2 = 0,971)$$

Der Verlauf des multiplikativen Modells liegt zwischen dem des Quadratwurzel-Modells und dem des logarithmischen Modells.

2.3.1.4 Das Potenz-Modell

Ähnlich flexibel, wie das multiplikative Modell, ist das Potenz-Modell, das allerdings drei unbekannte Parameter enthält:

$$Y = \alpha + \beta \cdot t^\gamma + u \quad (2.18)$$

Wie das multiplikative Modell nimmt es für unterschiedliche Werte von γ verschiedene Verlaufsformen an:

- (a) $\gamma > 1$: progressiv steigender Verlauf (konvexer Verlauf)
- (b) $0 < \gamma < 1$: degressiv steigender Verlauf (konkaver Verlauf)

Für $\gamma = 0,5$ ergibt sich das Quadratwurzel-Modell, das einen Spezialfall des Potenzmodells bildet.

Das Potenz-Modell ist ein echtes bzw. *intrinsisch nichtlineares Modell*. Es ist nicht-linear in Bezug auf den unbekannt Parameter γ und lässt sich daher nicht mittels linearer Regressionsanalyse schätzen. Für die Lösung des Kleinstquadratprinzips existiert keine analytische Lösung. Vielmehr bedarf es hierfür der Anwendung iterativer Methoden (siehe dazu Kapitel 10: Nichtlineare Regression). Man erhält hier:

$$\hat{Y} = 1.298 + 362,3 \cdot t^{0,611} \quad (R^2 = 0,984)$$

Der Verlauf des Potenzmodells ist in diesem Fall nur geringfügig steiler als der des Quadratwurzel-Modells und die Anpassungsgüte nur marginal besser.

2.3.1.5 Vergleich der Modelle

In der Abbildung 2.12 sind die fünf obigen Trend-Modelle mit ihren Gütemaßen sowie Prognosewerten und Prognosefehlern für Periode 11 zusammengestellt. Den besten Fit weisen hier das Potenzmodell und das Quadratwurzel-Modell auf. Während das Bestimmtheitsmaß des Potenzmodells noch etwas größer ist als das des Quadratwurzel-Modells, ist der F-Wert des Potenzmodells bedeutend niedriger, da ein Parameter mehr zu schätzen ist als beim Quadratwurzel-Modell und den anderen Modellen.

Betrachtet man die prognostizierten Werte für Periode 11, so fällt auf, dass diese beim logarithmischen Modell und beim multiplikativen Modell unter dem letzten Beobachtungswert (2.785 für Periode 10) liegen. Im Beobachtungszeitraum kommt dies nicht vor, sondern die Absatzmenge hat sich kontinuierlich jeweils gegenüber der Vorperiode erhöht. Damit erscheint es naheliegend, dass der Manager des Margarineherstellers diese Modelle nicht auswählen wird. Da auch das lineare Modell nicht adäquat erscheint, verbleiben zur Erstellung seiner Absatzprognose hier nur das Quadratwurzel-Modell oder das Potenz-Modell. Diese beiden Modelle weisen auch den geringsten Prognosefehler auf.

Intrinsisch
nichtlinear

Fit

Modell	geschätzte Funktion	R^2	F	\hat{y}_{11}	$s_p(11)$
Linear	$\hat{Y} = 1.620 + 120,9t$	0,972	277	2.949	80
Quadratwurzel	$\hat{Y} = 1.117 + 519,6\sqrt{t}$	0,983	459	2.840	60
Logarithmisch	$\hat{Y} = 1.542 + 492,6 \cdot \ln(t)$	0,945	139	2.722	104
Multiplikativ	$\hat{Y} = 1.600 \cdot t^{0,228}$	0,971	265	2.761	96
Potenz	$\hat{Y} = 1.298 + 362,3 \cdot t^{0,611}$	0,984	216	2.866	59

Abbildung 2.12: Vergleich der Modelle

2.3.2 Berücksichtigung von Strukturbrüchen

Die in Abbildung 2.13 dargestellten Trendverläufe weisen jeweils einen Strukturbruch auf. Derartige Strukturbrüche findet man häufig bei Zeitreihenanalysen, z. B. wenn durch Änderung der wirtschaftlichen Rahmenbedingungen eine Änderung in der zeitlichen Entwicklung einer betrachteten Variablen Y bewirkt wird. Strukturbrüche lassen sich durch eine Dummy-Variable berücksichtigen, deren Werte vor dem Strukturbruch in Periode t' Null sind und danach Werte größer als Null annehmen.

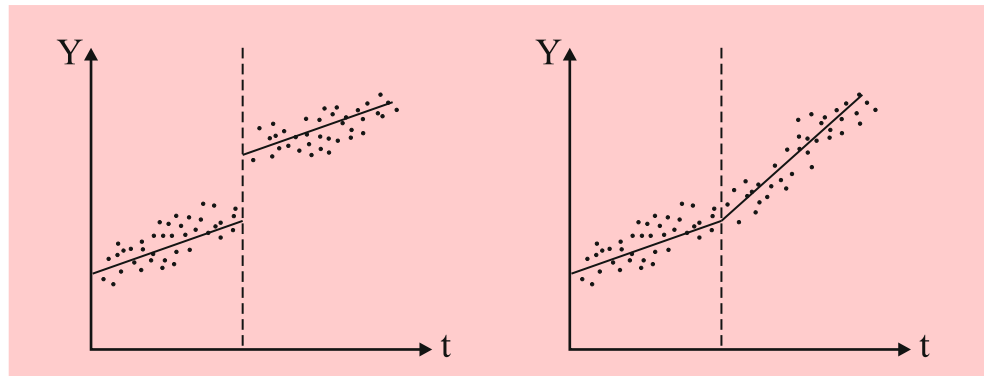


Abbildung 2.13: Strukturbrüche

Niveauänderung a) Niveauänderung in Periode $t = t'$:

$$Y = \alpha + \beta_1 \cdot t + \beta_2 \cdot q + u \quad \text{mit } q = \begin{cases} 0 & \text{für } t < t' \\ 1 & \text{für } t \geq t' \end{cases} \quad (2.19)$$

Die Werte der Dummy-Variablen q über die Zeit t sind nachfolgend dargestellt:

t	1	2	3	...	$t'-1$	t'	$t'+1$	$t'+2$...	$t'+k$
q	0	0	0	...	0	1	1	1	...	1

Vor dem Strukturbruch gilt damit:

$$t < t' : Y = \alpha + \beta_1 \cdot t + u \quad (2.20)$$

Nach dem Strukturbruch erhöht sich das Niveau um β_2 :

$$t \geq t' : Y = \alpha + \beta_1 \cdot t + \beta_2 \cdot q + u = (\alpha + \beta_2) + \beta_1 \cdot t + u \quad (2.21)$$

Durch lineare Regression von Y auf die Variablen t und q lassen sich die drei Parameter α , β_1 und β_2 in (2.19) schätzen.

b) Trendänderung in Periode $t = t'$:

Trendänderung

$$Y = \alpha + \beta_1 \cdot t + \beta_2 \cdot q + u \quad \text{mit } q = \begin{cases} 0 & \text{für } t < t' \\ t - t' + 1 & \text{für } t \geq t' \end{cases} \quad (2.22)$$

Die Werte der Dummy-Variablen q über die Zeit t sind nachfolgend dargestellt:

t	1	2	3	...	t'-1	t'	t'+1	t'+2	...	t'+k
q	0	0	0	...	0	1	2	3	...	k+1

Vor dem Strukturbruch gilt damit:

$$t < t' : \quad Y = \alpha + \beta_1 \cdot t + u \quad (2.23)$$

Ab dem Strukturbruch erhöht sich die Steigung um β_2 :

$$\begin{aligned} t \geq t' : \quad Y &= \alpha + \beta_1 \cdot t + \beta_2 \cdot (t - t' + 1) + u \\ &= (\alpha - \beta_2 \cdot t' + \beta_2) + (\beta_1 + \beta_2) \cdot t + u \\ &= \alpha' + (\beta_1 + \beta_2) \cdot t + u \end{aligned} \quad (2.24)$$

Für den rechten Ast der Trendfunktion, falls man diesen nach links verlängern würde, ergibt sich der fiktive Schnittpunkt α' mit der Y-Achse, der auch negativ sein kann.

Mittels linearer Regression von Y auf die Variablen t und q lassen sich wiederum die drei unbekannt Parameter α , β_1 und β_2 in (2.22) schätzen.

2.3.3 Berücksichtigung von zyklischen Schwankungen

Bei diskreten Zeitreihen lassen sich auch zyklische Schwankungen, wie Saisoneffekte oder Konjunkturschwankungen, mit Hilfe von Dummy-Variablen berücksichtigen. Als Beispiel diene das folgende Modell mit Trend und Saisoneffekten:

Saisoneffekte und Konjunkturschwankungen

$$Y = \alpha + \beta_1 \cdot t + \gamma_1 \cdot q_1 + \gamma_2 \cdot q_2 + \gamma_3 \cdot q_3 + u \quad (2.25)$$

$$\text{mit } q_1 = \begin{cases} 1 & \text{für Frühjahr} \\ 0 & \text{sonst} \end{cases}, q_2 = \begin{cases} 1 & \text{für Sommer} \\ 0 & \text{sonst} \end{cases}, q_3 = \begin{cases} 1 & \text{für Herbst} \\ 0 & \text{sonst} \end{cases}$$

Für den Fall, dass Monatsdaten vorliegen, ergibt sich für die Dummy-Variablen die in Abbildung 2.14 dargestellte Datenstruktur.

Die vierte Dummy-Variable für Winter wird in obiger Modellformulierung nicht benötigt. Das vierte Quartal bildet quasi die Basis und die Koeffizienten der drei Dummy-Variablen geben die Abweichungen von dieser Basis an. Der Effekt des Basisquartals ist im konstanten Glied α enthalten. Würde man auch die vierte Dummy-Variable in das Modell aufnehmen, ließe es sich wegen perfekter Multikollinearität nicht mehr schätzen, da die Addition der vier Dummy-Variablen immer den konstanten Wert Eins ergibt.

Multikollinearität

Entfernt man das konstante Glied aus der Regressionsgleichung, so lassen sich alle vier Dummy-Variablen einbeziehen und man erhält das folgende Modell:

$$Y = \beta_1 \cdot t + \gamma_1 \cdot q_1 + \gamma_2 \cdot q_2 + \gamma_3 \cdot q_3 + \gamma_4 \cdot q_4 + u \quad (2.26)$$

Durch Wegfall des konstanten Gliedes wird hier die Multikollinearität vermieden.

Trend		Dummy-Variablen				
		t (Monat)	q ₁	q ₂	q ₃	q ₄
1. Zyklus (Jahr)	1. Saison (Quartal)	1	1	0	0	0
		2	1	0	0	0
		3	1	0	0	0
	2. Saison (Quartal)	4	0	1	0	0
		5	0	1	0	0
		6	0	1	0	0
	3. Saison (Quartal)	7	0	0	1	0
		8	0	0	1	0
		9	0	0	1	0
	4. Saison (Quartal)	10	0	0	0	1
		11	0	0	0	1
		12	0	0	0	1
2. Zyklus (Jahr)	1. Saison (Quartal)	13	1	0	0	0
		14	1	0	0	0
		15	1	0	0	0

Abbildung 2.14: Datenstruktur der Dummy-Variablen für saisonale Schwankungen

2.3.4 Umsetzung mit SPSS

Zur Durchführung von Zeitreihenanalyse auf Basis regressionsanalytischer Modelle lassen sich in SPSS die Prozeduren „Lineare Regression“, „Kurvenanpassung“ und „Nichtlineare Regression“ verwenden.

2.3.4.1 Lineare Regression

Die Schätzung der obigen Zeitreihenmodelle lässt sich mit Hilfe der Prozedur „Lineare Regression“ durchführen. Eine Ausnahme bildet das Potenz-Modell (2.18), bei dem es sich um ein intrinsisch nichtlineares Modell handelt, für dessen Schätzung die Prozedur „Nichtlineare Regression“ erforderlich ist (siehe dazu Kapitel 10: Nichtlineare Regression). Alle anderen Modelle lassen sich mit Hilfe nichtlinearer Transformationen oder der Erzeugung von Dummy-Variablen auf eine lineare Form bringen. Die erforderlichen nichtlinearen Transformationen können mit Hilfe des Menüpunktes „Transformieren / Variable berechnen“ vorgenommen werden. Damit lassen sich neue Variablen berechnen, mittels derer sodann die lineare Regression durchgeführt werden kann. Die Prozedur „Lineare Regression“ enthält allerdings keine Option zur Durchführung von Prognoserechnungen. Die Erstellung von Prognosen muss der Untersucher auf Basis der geschätzten Modelle selber vornehmen. Um den Verlauf der geschätzten Funktion in die Zukunft zu extrapolieren, kann wiederum der Menüpunkt „Transformieren / Variable berechnen“ oder ein Taschenrechner verwendet werden. Zur Berechnung des Prognosefehlers gemäß (2.15) liefert die Prozedur „Lineare Regression“ den Standardfehler der Regression s . Den Mittelwert \bar{x} und die Standardabweichung s_x der transformierten Prädiktorvariablen kann man mit der Prozedur „Deskriptive Statistik“ berechnen. Somit sind für die Berechnung des Prognosefehlers keine Berechnungen auf Basis der individuellen Zeitreihendaten erforderlich, was bei langen Zeitreihen sehr aufwändig sein könnte.

Transformation

Prognose

2.3.4.2 Kurvenanpassung

Die Arbeit wird wesentlich erleichtert, wenn man die Prozedur „Kurvenanpassung“ (Curve Fit) nutzen kann. In dieser Prozedur werden dem Benutzer das lineare Trendmodell sowie zehn weitere nichtlineare Trendmodelle zur Verfügung gestellt (siehe Abbildung 2.15). Es werden globale Gütemaße (Bestimmtheitsmaß und F-Wert) sowie die Standardfehler und t-Werte der Regressionsparameter ermittelt, und die beobachteten und geschätzten Werte der Zeitreihenvariablen Y werden grafisch dargestellt. Außerdem lassen sich durch Extrapolation des geschätzten Modells auch Punkt- und Intervallprognosen durchführen.²²

Trendmodell

Für die Erstellung von Prognosen ist es erforderlich, dass die unabhängige Variable „Zeit“ nicht aus der Datendatei ausgewählt wird, sondern durch die Prozedur „Kurvenanpassung“ erzeugt wird. Weiterhin ist in der Dialogbox die Option „Speichern“ (Save) aufzurufen. Dort lassen sich eine Vertrauenswahrscheinlichkeit (90 %, 95 % oder 99 %) sowie die Periode, bis zu der Prognosen erstellt werden sollen (Prognosehorizont), wählen. Die prognostizierten Werte sowie die zugehörigen Prognoseintervalle werden anschließend in der Datendatei gespeichert. Hierzu werden drei neue Variablen erzeugt: „FIT“ (geschätzte bzw. prognostizierte Werte), „LCL“ (Lower Confidence Limit) und „UCL“ (Upper Confidence Limit).²³

Prognose

In Abbildung 2.15 sind die elf Modelle zusammengestellt. Alle Modelle sind intrinsisch linear, d. h. sie lassen sich, soweit erforderlich, mittels nichtlinearer Transformationen auf eine lineare Form bringen und somit mittels linearer Regression schätzen. Im rechten Teil von Abbildung 2.15 sind die linearisierten Formen dargestellt. Auf die Einbeziehung der Störgröße wurde hier verzichtet.

Nr.	Name	Formel	Linearisierung
1	Linear	$Y = \alpha + \beta \cdot t$	
2	Logarithmisch	$Y = \alpha + \beta \cdot \ln(t)$	
3	Inverse	$Y = \alpha + \beta/t$	$Y = \alpha + \beta \cdot 1/t$
4	Quadratisch (Quadratic)	$Y = \alpha + \beta_1 \cdot t + \beta_2 \cdot t^2$	
5	Kubisch (Cubic)	$Y = \alpha + \beta_1 \cdot t + \beta_2 \cdot t^2 + \beta_3 \cdot t^3$	
6	Power	$Y = \alpha \cdot t^\beta$	$\ln(Y) = \ln(\alpha) + \beta \cdot \ln(t)$
7	Zusammengesetzt (Compound)	$Y = \alpha \cdot \beta^t$	$\ln(Y) = \ln(\alpha) + \ln(\beta) \cdot t$
8	S-Kurve (S-curve)	$Y = e^{\alpha + \beta/t}$	$\ln(Y) = \alpha + \beta \cdot 1/t$
9	Logistisch	$Y = \frac{1}{1/M + \alpha \cdot \beta^t}$	$\ln(1/Y - 1/M) = \ln(\alpha) + \ln(\beta) \cdot t$
10	Aufbau (Growth)	$Y = e^{\alpha + \beta \cdot t}$	$\ln(Y) = \alpha + \beta \cdot t$
11	Exponentiell (Exponential)	$Y = \alpha \cdot e^{\beta \cdot t}$	$\ln(Y) = \ln(\alpha) + \beta \cdot t$

Abbildung 2.15: Modelle der SPSS-Prozedur „Kurvenanpassung“

Die Modelle 1, 2 und 6 wurden oben behandelt. Das Modell 6 (Power) entspricht dem multiplikativen Modell (2.17).

Die Modelle 4 (Quadratisch) und 5 (Kubisch) sind Polynome zweiten und dritten Grades. Sie erzielen zwar für die Zeitreihendaten der Absatzmenge einen guten Fit, sind aber für mittel- oder langfristige Prognosen ungeeignet, da sie sich wegen der

²²Vgl. SPSS Inc. (2007), S. 372 ff.

²³In der deutschen Version von SPSS funktioniert diese Option allerdings noch nicht.

großen Exponenten sehr schnell vom Wertebereich der Beobachtungsdaten entfernen und auch negative Werte annehmen können.

Die übrigen Modelle sind zur Modellierung der obigen Zeitreihe, die einen recht flach gekrümmten, degressiv steigenden (konkaven) Verlauf besitzt, nicht geeignet. Das Inverse Modell 3 besitzt einen sehr stark gekrümmten Verlauf und würde viel zu niedrige Prognosen erbringen. Die Schätzung der Modelle 7 bis 11 führt hier zu progressiv steigenden (konvexen) Verläufen.

Die Modelle 5 (Kubisch) und 9 (Logistisch) können auch s-förmige Verläufe annehmen.²⁴ Das logistische Modell ist ein sog. Wachstumsmodell (Growth Model), das für wachsendes t gegen eine Wachstumsgrenze, den Maximalwert M , konvergiert.²⁵ Sein Verlauf ist symmetrisch um den Wendepunkt (vgl. Abbildung 2.16), der bei $y_w = M/2$ liegt. Es findet in vielen Bereichen Verwendung, so z. B. in der Innovations- und Diffusionsforschung²⁶, in der Epidemiologie sowie auch als Approximation der Verteilungsfunktion der Gauss'schen Normalverteilung, für die kein analytischer Ausdruck existiert (vgl. dazu auch Kapitel 5: Logistische Regression). In der Marktforschung wird das logistische Modell u.a. zur Modellierung der Absatzmenge bei der Einführung von neuen Produkten verwendet.

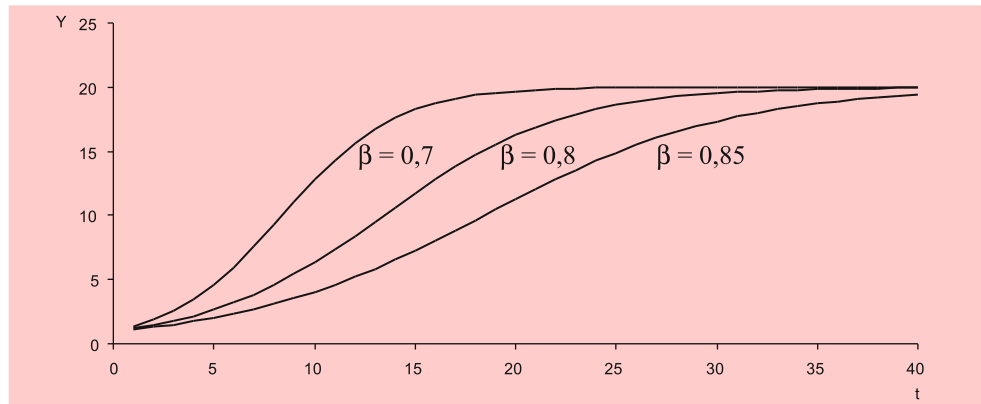


Abbildung 2.16: Logistisches Modell für verschiedene Werte von β

Das logistische Modell ist eigentlich ein intrinsisch nichtlineares Modell. Es lässt sich nur linearisieren, wenn die Wachstumsgrenze M bekannt ist. In der Prozedur „Kurvenanpassung“ muss daher M durch den Benutzer spezifiziert werden. Der anzugebende Wert muss größer sein als der größte Wert der vorliegenden Zeitreihe. Zur Schätzung des logistischen Modells siehe auch Kapitel 1 im Buch *Fortgeschrittene Multivariate Analysemethoden: Nichtlineare Regression*.

Die Formulierung des logistischen Modells in SPSS ist etwas ungewöhnlich. Üblicher

²⁴Das als „S-Kurve“ bezeichnete Exponentialmodell 8 kann, ähnlich dem Inversen Modell 3, nur konvexe oder konkave Verläufe annehmen, je nachdem, ob $\beta > 0$ oder $\beta < 0$ gilt. Die Bezeichnung als „S-Kurve“ ist daher irreführend.

²⁵Bei dem Modell 10 („Aufbau“ bzw. „Growth“) handelt es sich nicht um ein Wachstumsmodell im engeren Sinne, da es nicht gegen einen oberen Grenzwert konvergiert. Einen Überblick über Wachstumsmodelle geben z. B. Mertens, P. (Hrsg.) (1994), S. 157 ff., Hammann/Erichson (2000), S. 446 ff.

²⁶Vgl. z. B. Meffert/Steffenhagen (1977), S. 70 ff., Weiber (1992).

ist die folgende Formulierung:

$$Y = \frac{M}{1 + e^{\alpha - \beta \cdot t}} \quad (2.27)$$

die sich für bekanntes M wie folgt linearisieren lässt:

$$\ln\left(\frac{M}{Y} - 1\right) = \alpha - \beta \cdot t \quad (2.28)$$

Abbildung 2.16 zeigt den Verlauf für $M = 20$ und verschiedene Werte von β ($0 < \beta < 1$). Für größere Werte von β wird der Verlauf flacher. Der Parameter α dagegen verändert die Form des Modells nicht, sondern bewirkt mit zunehmender Größe eine Verschiebung der Funktion nach rechts. In Modell 9 muss $\alpha > 0$ gelten, während in der Formulierung gemäß (2.27) der Parameter α beliebige Werte annehmen kann.

2.4 Fallbeispiel

2.4.1 Problemstellung

Während in dem oben betrachteten Absatzgebiet der Margarineabsatz über die letzten Jahre kontinuierlich angestiegen ist, mussten dagegen in den meisten anderen Absatzgebieten z. T. erhebliche Einbußen hingenommen werden. Der Manager unseres Margarineherstellers fragt sich daher, ob dies an unzureichenden Marketingbemühungen des Unternehmens liegt oder ob sich die Marktbedingungen geändert haben. Überdies kam es im Jahresablauf des öfteren zu Lieferengpässen, da die Absatzmengen stark schwankten. Für die Produktionsplanung wäre daher sowohl eine kurzfristige Prognose des Absatzvolumens für die jeweils nächsten Monate wie auch eine mittelfristige Prognose für die kommenden Jahre von großer Wichtigkeit.

Prognose des
Absatzvolumens

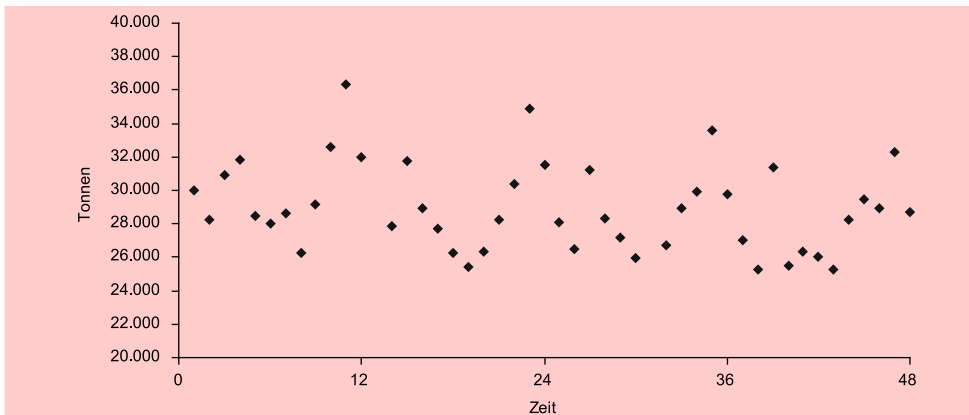


Abbildung 2.17: Zeitreihe des Marktabsatzvolumens von Margarine in Tonnen (Monatliche Daten von Jan. 2004 - Dez. 2007, Quelle: GfK ConsumerScan)

Um dem Problem auf den Grund zu gehen, beabsichtigt der Manager unseres Margarineherstellers, zunächst den Gesamtmarkt für Margarine in Deutschland zu analysieren und zu prognostizieren. Hierzu besorgt er sich von der Gesellschaft für Konsumforschung (GfK AG) in Nürnberg die monatlichen Absatzdaten der letzten vier Jahre,

die in Abbildung 2.17 als Streudiagramm dargestellt sind.²⁷ Auch die Absatzmengen des Gesamtmarktes weisen eine erhebliche Streuung auf. Es stellt sich die Frage, ob diese Streuung eher zufällig bedingt ist, oder ob sich dahinter eine systematische Entwicklung verbirgt, die sich modellieren und für Prognosen nutzen lässt.

2.4.2 Ergebnisse

2.4.2.1 Extrapolationsmodelle

Der Manager beginnt zunächst mit der Anpassung eines linearen Trendmodells gemäß (5) an die Daten.

Linearer Trend

Modell 1: Linearer Trend

$$\hat{Y} = 29.836 - 41,7 \cdot \text{Zeit} \quad (R^2 = 0,046)$$

Das lineare Trendmodell zeigt eine negative Entwicklung des Marktabsatzvolumens von Margarine in Deutschland über die Jahre 2004 bis 2007. Durch den Trend, der in Abbildung 2.18 grafisch dargestellt ist, lässt sich die Streuung in den Absatzdaten allerdings nur in geringem Umfang erklären. Das Bestimmtheitsmaß beträgt lediglich 4,6 %.

Schwankungen

In Abbildung 2.18 sind die Streupunkte des Marktabsatzvolumens durch Linien verbunden. Man erkennt so, dass die starken Schwankungen eine gewisse Regelmäßigkeit aufweisen, die sich im Abstand von 12 Monaten wiederholt. Daraus ist zu folgern, dass sie primär saisonal bedingt sind.

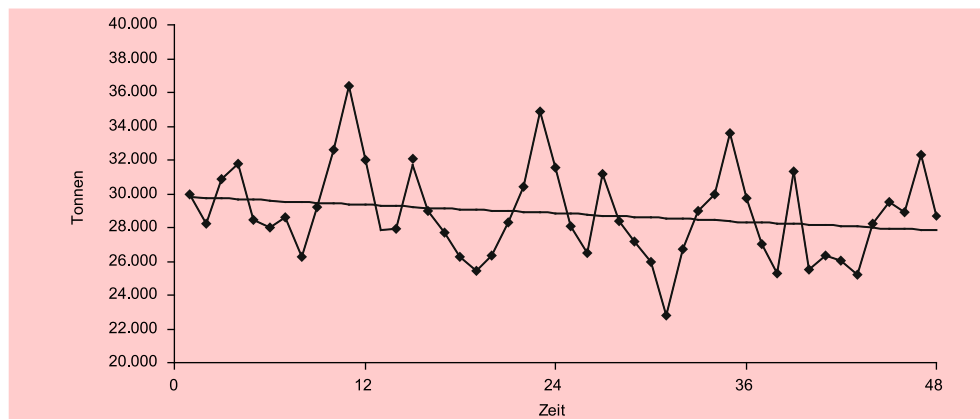


Abbildung 2.18: Entwicklung des Marktabsatzvolumens von Margarine (Jan. 2004 - Dez. 2007)

²⁷Der GfK AG sei an dieser Stelle für die Überlassung der Daten gedankt.

Modell 2: Trend + saisonale Dummies

Saisonale Dummies

$$\hat{Y} = 29.439 - 64,0 \text{ Zeit } (-5,3) ** \quad (R^2 = 0,873)$$

$$- 1.194 \text{ Februar } (-1,5)$$

$$+ 3.198 \text{ März } (+4,0) **$$

$$+ 634 \text{ April } (+0,8)$$

$$- 547 \text{ Mai } (-0,7)$$

$$- 1.343 \text{ Juni } (-1,7)$$

$$- 2.319 \text{ Juli } (-2,9) **$$

$$- 892 \text{ August } (-1,1)$$

$$+ 1.270 \text{ September } (+1,6)$$

$$+ 2.816 \text{ Oktober } (+3,5) **$$

$$+ 6.706 \text{ November } (+8,3) **$$

$$+ 2.981 \text{ Dezember } (+3,7) **$$

Durch die Einbeziehung von monatlichen Dummy-Variablen zur Erfassung der saisonalen Schwankungen lässt sich der Erklärungsanteil des Modells auf 87,3 % erhöhen.

Erklärungsanteil

Die t-Werte der Regressionskoeffizienten sind (hier und im Folgenden) jeweils in Klammern angegeben. Werte mit einem Signifikanzniveau von 5 % sind durch einen Stern (*) und Werte mit einem Signifikanzniveau von 1 % sind durch zwei Sterne (**) gekennzeichnet.

Signifikanzniveau

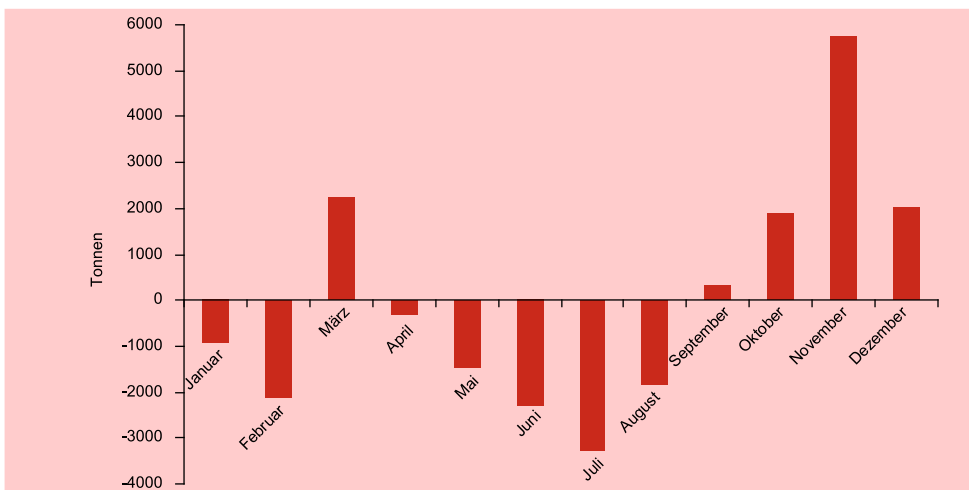


Abbildung 2.19: Saisonfigur im Margarinekonsum (2004 - 2007)

Aus den Regressionskoeffizienten der monatlichen Dummy-Variablen lässt sich die in Abbildung 2.19 dargestellte Saisonfigur des Marktabsatzvolumens von Margarine ableiten. Die Werte wurden hierzu um den Mittelwert zentriert. Auffällig sind die starken positiven Ausschläge im März und ganz besonders im November, die die umsatzstärksten Monate bilden. Der Rückgang in den Sommermonaten hängt vermutlich mit den höheren Temperaturen und dem dadurch bedingten geringeren Kalorienbedarf zusammen. Dies soll im Folgenden näher untersucht werden.

Saisonfigur

Prognose

Die zwei vorstehenden Regressionsmodelle machen ein Muster in den Schwankungen der Zeitreihe deutlich (Trend- und Saisonkomponente) und trennen dieses von den zufälligen Schwankungen. Mittels Extrapolation dieses Musters lassen sich Prognosen für beliebige zukünftige Perioden erstellen. Die Prognosen implizieren allerdings die Annahme, dass das Muster stabil ist, sich also in der Zukunft nicht verändert.

Extrapolationsmodell

Derartige Extrapolationsmodelle beschreiben lediglich den Verlauf der Zeitreihe, können ihn aber nicht erklären, d. h. sie berücksichtigen nicht die Faktoren, die den Verlauf der Zeitreihe beeinflussen bzw. ihr kausal zugrunde liegen. Prognosen lassen sich damit recht einfach erstellen, da man keine weiteren Informationen benötigt. Nachteilig ist, dass theoretische Überlegungen bezüglich kausaler Zusammenhänge, die die betrachtete Zeitreihe tangieren, außer acht bleiben.

2.4.2.2 Strukturmodelle

Strukturmodelle

Ein Vorteil der Regressionsanalyse gegenüber vielen anderen Verfahren der Zeitreihenanalyse ist darin zu sehen, dass sich die Modelle leicht durch Einbeziehung weiterer Variablen, die den Verlauf der Zeitreihe möglicherweise erklären, erweitern lassen und somit theoretische Überlegungen in die Modellbildung einbezogen werden können. Dies erfordert allerdings, dass für diese erklärenden Variablen entsprechende Zeitreihendaten vorliegen müssen. Die Prognose mittels derartiger *Strukturmodelle* wird überdies erschwert, da die Prognose der interessierenden Zeitreihenvariablen Y i. d. R. zunächst eine Prognose der erklärenden Variablen (Prädiktoren) erforderlich macht.

Relevante Einflussfaktoren der Nachfrage nach Konsumgütern bilden z. B. die Größe der Bevölkerung und in vielen Fällen auch das Wetter. Ein Indikator des Wetters ist z. B. die mittlere Temperatur pro betrachteter Periode, in diesem Fall pro Monat. Nachfolgend soll zunächst der Einfluss der Temperatur auf den Margarinekonsum und sodann der Einfluss des Bevölkerungswachstums untersucht werden.

2.4.2.2.1 Einbeziehung der Temperatur

Temperaturverlauf

In Abbildung 2.20 ist der Temperaturverlauf (monatliche Durchschnittswerte) über den Zeitraum, für den die Margarine-Daten vorliegen, dargestellt. Daraus lässt sich ein leichter Anstieg erkennen.

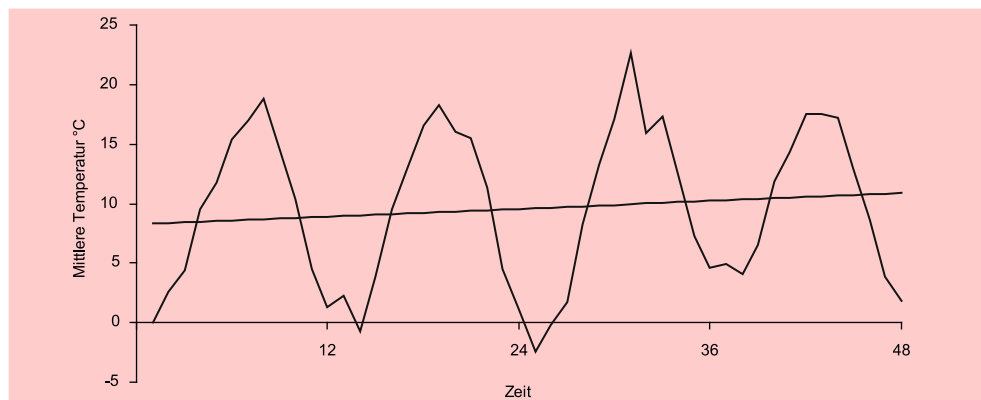


Abbildung 2.20: Temperaturverlauf Jan. 2004 - Dez. 2007 (monatliche Durchschnittswerte)

Analog zur Saisonfigur des Margarinekonsums in Abbildung 2.19 ist in 2.21 die Saisonfigur des jährlichen Temperaturverlaufs dargestellt. Ein Vergleich der Abbildungen macht deutlich, dass sich die saisonalen Schwankungen des Margarinekonsums nur unvollständig durch die Temperaturschwankungen erklären lassen. Dies zeigt genauer die nachfolgende Regressionsanalyse.

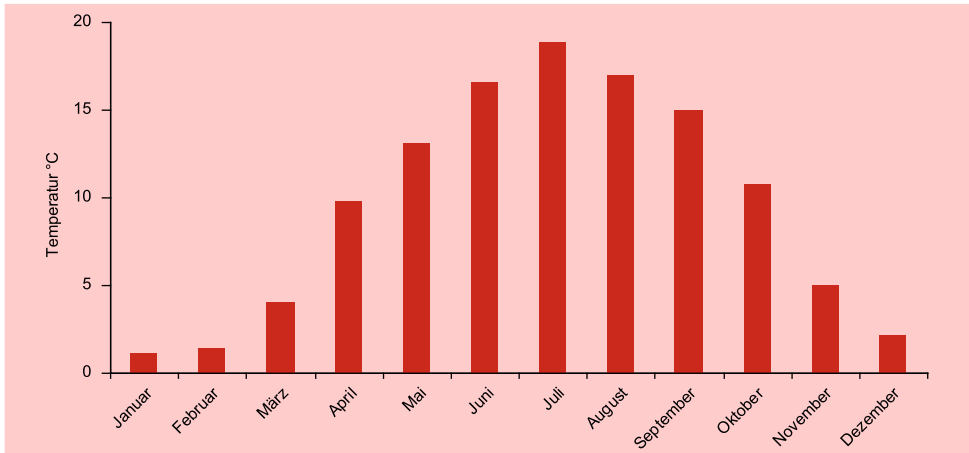


Abbildung 2.21: Saisonfigur der Temperatur (Mittel der Jahre 2004 - 2007)

Durch die Einbeziehung der Temperaturdaten erhöht sich der Erklärungsanteil des Modells von 87,3% geringfügig auf 89,1%. Die um den Temperatureinfluss bereinigten Regressionskoeffizienten der monatlichen Dummy-Variablen dagegen verändern sich z. T. erheblich. Sie sind in Abbildung 2.22 grafisch dargestellt. Dies deutet darauf hin, dass die saisonalen Schwankungen des Margarinekonsums in starkem Maße im Verbraucherverhalten begründet liegen.

Der Rückgang des Margarinekonsums in den Sommermonaten wird jetzt durch die Temperatur-Variable aufgefangen und zeigt sich in den Koeffizienten der Dummy-Variablen nur noch abgeschwächt. Dagegen wird der starke Einbruch im Januar und Februar, der nicht temperaturbedingt ist, jetzt noch deutlicher.

Modell 3: Trend + saisonale Dummies + Temperatur

$$\hat{Y} = 29.582 - 57,5 \text{ Zeit } (-4,9) ** \quad (R^2 = 0,891)$$

$$- 1.136 \text{ Februar } (-1,5)$$

$$+ 3.854 \text{ März } (+4,8) **$$

$$+ 2.579 \text{ April } (+2,3) *$$

$$+ 2.142 \text{ Mai } (+1,6)$$

$$+ 2.148 \text{ Juni } (+1,3)$$

$$+ 1.668 \text{ Juli } (+0,9)$$

$$+ 2.663 \text{ August } (+1,6)$$

$$+ 4.372 \text{ September } (+2,9) **$$

$$+ 4.938 \text{ Oktober } (+4,2) **$$

$$+ 7.520 \text{ November } (+9,0) **$$

$$+ 3.142 \text{ Dezember } (+4,1) **$$

$$- 227,8 \text{ Temperatur } (-2,4) *$$

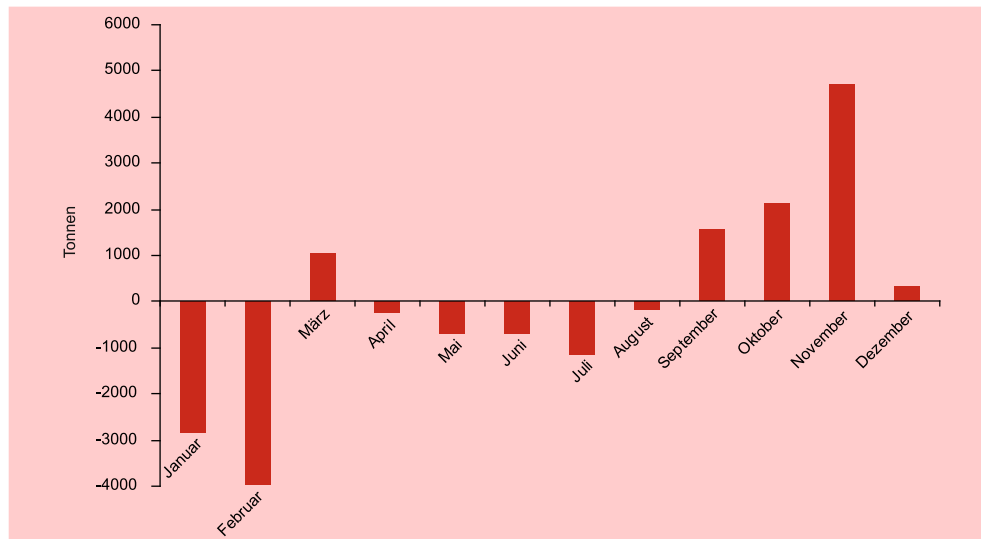


Abbildung 2.22: Saisonfigur im Margarinekonsum nach Elimination des Temperatureinflusses

2.4.2.2.2 Einbeziehung der Bevölkerungsentwicklung

Bevölkerungsgröße

Die Bevölkerungsgröße hat generell maßgeblichen Einfluss auf die Nachfrage nach Konsumgütern. Es soll hier deshalb der Einfluss auf den Margarinekonsum untersucht werden.

Abbildung 2.23 zeigt die Bevölkerungsentwicklung für den hier betrachteten Zeitraum. Während für den Margarinekonsum monatliche Daten vorliegen, sind Bevölkerungsdaten nur jährlich verfügbar, im vorliegenden Fall also nur vier Daten.

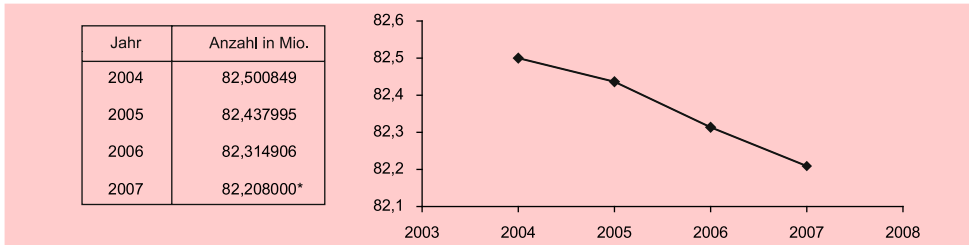


Abbildung 2.23: Bevölkerung Deutschlands

Quelle: Stat. Bundesamt, *Wert geschätzt

Im betrachteten Zeitraum von 2004 bis 2007 besteht eine sehr enge Korrelation zwischen der Variablen Zeit und der Bevölkerung ($r = -0,992$), wie sich aus Abbildung 2.23 ersehen lässt. Ersetzt man in Modell 3 die Variable Zeit durch die Bevölkerungszahl, so erhält man folgendes Modell:

Modell 4: Saisonale Dummies + Temperatur + Bevölkerung

$$\hat{Y} = -510,754 - 1,193 \text{ Februar } (-1, 5) \quad (R^2 = 0,884)$$

$$+ 3,746 \text{ März } (+4, 5) **$$

$$+ 2,428 \text{ April } (+2, 1) *$$

$$+ 1,941 \text{ Mai } (+1, 4)$$

$$+ 1,899 \text{ Juni } (+1, 1)$$

$$+ 1,367 \text{ Juli } (+0, 7)$$

$$+ 2,300 \text{ August } (+1, 3)$$

$$+ 3,947 \text{ September } (+2, 5) *$$

$$+ 4,444 \text{ Oktober } (+3, 6) **$$

$$+ 6,955 \text{ November } (+8, 1) **$$

$$+ 2,512 \text{ Dezember } (+3, 2) **$$

$$- 230,3 \text{ Temperatur } (-2, 3) *$$

$$+ 6,547 \text{ Bevölkerung } (+4, 6) *$$

Der Einfluss der Bevölkerung erweist sich als hoch signifikant. Das Bestimmtheitsmaß verringert sich nur geringfügig von 89,1 % auf 88,4 %.

Die Modelle 3 und 4 eignen sich damit annähernd gleich gut für Prognosen, allerdings nur solange, wie die enge Korrelation zwischen Zeit und Bevölkerung bestehen bleibt. Ist dies nicht der Fall, so wird Modell 4 die besseren Prognosen liefern. Für seine Anwendung werden dann allerdings Prognosen der Bevölkerungsentwicklung benötigt. Mittels Kennziffern zu Fertilität, Sterblichkeit und Migration lassen sich allerdings recht genaue Bevölkerungsprognosen erstellen.²⁸

Bevölkerungs-
prognose

²⁸Siehe hierzu „Bevölkerung Deutschlands bis 2050“, Statistisches Bundesamt, Wiesbaden 2006.

2.4.2.3 Zusammenfassung

Die Analyse der Zeitreihe des Marktabsatzvolumens zeigt, dass der Margarinekonsum in Deutschland gekennzeichnet ist durch

- starke **saisonale Schwankungen** und
- einen leicht **abfallenden Trend** (siehe Abbildung 2.18).

Konsumverhalten

Die saisonalen Schwankungen im Margarinekonsum, die sich regelmäßig wiederkehrend im Verlauf eines Jahres zeigen, lassen sich z. T. auf die jahreszeitlichen Temperaturschwankungen (erhöhter Kalorienbedarf bei niedrigen Temperaturen) zurückführen. Allerdings lassen sich mit dem Temperatureinfluss nur knapp 25 % dieser Schwankungen erklären. Sehr viel stärker wirkt offenbar der Einfluss von Weihnachten und Ostern. Im Vorfeld dieser Festlichkeiten zeigt der Margarinekonsum jeweils im November und im März eine starke Zunahme, wobei der Konsum im November der weitaus höchste ist (vgl. Abbildung 2.19). Abbildung 2.22 zeigt die um den Temperatureinfluss bereinigten Schwankungen, die insbesondere den Einfluss von Weihnachten auf den Margarinekonsum deutlich werden lassen.

Der negative Trend im Margarinekonsum lässt sich weitgehend (zu über 90 %) durch den Rückgang der Bevölkerung erklären.

Prognoseeignung

Prognoseeignung der Modelle:

- Modell 1: Linearer Trend
Infolge der starken Saisonalität im Margarinekonsum ist dieses Modell für monatliche Prognosen ungeeignet.
- Modell 2: Trend + saisonale Dummies
Dieses Modell ist einfach anzuwenden und wird kurzfristig gute Prognosen liefern.

Temperatur- entwicklung

- Modell 3: Trend + saisonale Dummies + Temperatur
Da die Saisonalität im Margarinekonsum nicht parallel zur Temperaturentwicklung im Jahresablauf verläuft (vgl. Abbildung 2.19 und 2.21) und die Temperatur einen zusätzlichen Erklärungsbeitrag zu den Schwankungen im Margarinekonsum leistet, lässt sich Modell 2 durch Einbeziehung der Temperatur verbessern. Prognosen mit Modell 3 erfordern allerdings, dass dazu für den zu prognostizierenden Monat auch die mittlere Temperatur zu prognostizieren ist, was niemals ganz fehlerfrei gelingen kann.

Bevölkerungs- entwicklung

- Modell 4: Saisonale Dummies + Temperatur + Bevölkerung
Die Trendmodelle 1 bis 3 unterstellen, dass die enge Korrelation zwischen Zeit und Bevölkerungsentwicklung bestehen bleibt, d.h. dass der in den letzten Jahren nahezu linear abfallende Verlauf der Bevölkerung sich fortsetzen wird. Ist dies nicht der Fall, so wird Modell 4 die besseren Prognosen liefern. Bis zum Jahr 2002 war die Bevölkerungsentwicklung in Deutschland noch anwachsend. Die Prognosen des Statistischen Bundesamtes lassen aber erwarten, dass die Bevölkerung Deutschlands auch in den kommenden Jahrzehnten weiterhin schrumpfen wird. Da dieser Prozess nicht linear verlaufen wird, ist für langfristige Prognosen Modell 4 vorzuziehen.

Sowohl die Auswahl eines Prognoseverfahrens (wie hier z. B. die Regressionsanalyse) wie auch die Modellspezifikation sind abhängig vom Urteil des Untersuchers. Es wurden diverse empirische Vergleiche von alternativen Prognoseverfahren durchgeführt, um das beste Verfahren zu ermitteln. Hinsichtlich des größten deutschen Vergleichs, des „DGOR-Prognosevergleichs“²⁹, bemerkt Hüttner (1994), S. 351, dass für den „Sieg“ eher die Personen als die Verfahren entscheidend waren.

Für die Erstellung von Prognosen seien hier einige Hinweise gegeben:³⁰

1. Das Prognoseproblem sollte immer in Verbindung mit dem dahinter stehenden Entscheidungsproblem gesehen werden. Ist vom Ergebnis der Prognose keine Entscheidung abhängig, braucht auch keine Prognose erstellt zu werden.
2. Es sollte überlegt werden, ob die Anwendung eines formalen Prognoseverfahrens Vorteile bringt. Dies ist z. B. zu verneinen, wenn es um die kurzfristige Prognose von Aktienkursen geht oder um die Prognose einer volkswirtschaftlichen Entwicklung, die maßgeblich durch politische Entscheidungen beeinflusst wird.
3. Es muss geklärt werden, welche Daten benötigt und welche Datenquellen verwendet werden sollen. Bei Wirkungsprognosen müssen die relevanten Prädiktorvariablen identifiziert werden. Neben theoretischen Überlegungen ist hier oft auch Kreativität gefordert.
4. Die Länge der Zeitreihe sollte dem Prognosehorizont angepasst sein. Langfristige Prognosen erfordern lange Zeitreihen, während für kurzfristige Prognosen kürzere Zeitreihen hinreichend sind.
5. Vor der Durchführung von Analysen sollten die Daten visualisiert werden, damit sich der Untersucher vom Verlauf der Zeitreihe(n) ein „Bild“ machen kann.
6. Für die Spezifikation eines Modells sind gesunder Menschenverstand, Sachkenntnis und theoretische Überlegungen genau so wichtig wie Methodenkenntnisse. Der Untersucher sollte z. B. eine Vorstellung davon haben, welches Prognoseergebnis zu erwarten ist bzw. welche Ergebnisse plausibel sind (z. B. progressives oder degressives Wachstum, Schrumpfung).
7. Bei der Auswahl von Variablen oder der Auswahl eines Modelltyps sollte man sich nicht allein von statistischen Kriterien leiten lassen. Ein guter Fit im Stützbereich ist keine Gewähr für Prognosegüte. Häufig werden Indikatoren nicht auf Basis einer kausalen Beziehung zur Prognosevariablen ausgewählt, sondern auf Basis ihrer Korrelation mit der Prognosevariablen. Diese aber kann zufällig bedingt und somit bedeutungslos für die Zukunft sein.
8. Kausale Zusammenhänge sollten genutzt werden, wenn immer möglich. Zu bedenken ist aber, dass in Strukturmodellen die Prognosen der Wirkungen auch Prognosen der Ursachen erfordern.
9. Simpler Modellen ist gegenüber komplexen Modellen der Vorzug zu geben, solange empirische Befunde nicht dagegen sprechen.

²⁹Vgl. dazu Schwarze/Weckerle (1982).

³⁰Vgl. dazu auch Armstrong, J. (Hrsg.) (2002), S. 680 ff.

10. Neben Punktprognosen sollten Intervallprognosen erstellt werden, damit das Risiko der mit der Prognose verbundenen Entscheidung abgeschätzt werden kann.
11. Die Schätzung des Modells sollte aktualisiert werden, sobald neue Daten verfügbar sind.
12. Die Verwendung alternativer Datenquellen sowie alternativer Prognoseverfahren und deren Kombination hat sich als nützlich erwiesen.³¹

2.6 Mathematischer Anhang

Nachfolgend sei unter Verwendung der Matrixschreibweise die Berechnung des Multivariaten Prognosefehlers angegeben.

Prognose für Periode t' :

Prognosefehler

$$\hat{y}_{t'} = b_0 + \sum_{j=1}^J b_j x_{t'j}$$

Prognosefehler:

$$s_p(t') = s \sqrt{1 + \frac{1}{T} + (\mathbf{X}_{t'} - \bar{\mathbf{X}})'(\dot{\mathbf{X}}'\bar{\mathbf{X}})^{-1}(\mathbf{X}_{t'} - \bar{\mathbf{X}})}$$

mit

$\mathbf{X}_{t'}$ = J-Vektor der Werte der Prädiktorvariablen für Periode t'

$\bar{\mathbf{X}}$ = J-Vektor der Mittelwerte der Prädiktorvariablen

s = Standardfehler der Regression

Prognoseintervall für Vertrauenswahrscheinlichkeit $1 - \alpha$:

Prognoseintervall

$$\hat{y}_{t'} - t_{\alpha/2} s_p \leq y_{t'} \leq \hat{y}_{t'} + t_{\alpha/2} s_p$$

mit

$t_{\alpha/2}$ = Quantil der Verteilung (Student-Verteilung) für die Vertrauenswahrscheinlichkeit $1 - \alpha$ bei zweiseitigem Test und $T - J$ Freiheitsgraden.

Wurden nichtlineare Transformationen f_j ($j = 1, \dots, J$) der Prädiktoren bei der Regressionsrechnung berücksichtigt, so sind die entsprechend transformierten Prädiktoren in die Prognosefunktion einzusetzen. Wurde auch die abhängige Variable transformiert, so ergibt sich allgemein der Prognosewert durch:

Transformation

$$z = b_0 + \sum_{j=1}^J b_j f(x_{t'j})$$

$$\hat{y}_{t'} = f_0^{-1}(z)$$

Das Prognoseintervall ergibt sich dann durch

$$f_0^{-1}(\hat{y}_{t'} - t_{\alpha/2} s_p) \leq y_{t'} \leq f_0^{-1}(\hat{y}_{t'} + t_{\alpha/2} s_p)$$

Das Prognoseintervall ist bei nichtlinearer Transformation der abhängigen Variablen asymmetrisch.

³¹Vgl. dazu Armstrong, J. (Hrsg.) (2002), S. 417 ff., Bates/Granger (1969), Hüttner (1994).

Literaturhinweise

A. Basisliteratur zur Zeitreihenanalyse

- Armstrong, J. (Hrsg.) (2002)**, Principles of Forecasting, Boston (Mass.) u.a.
- Hamilton, J. (1994)**, Time Series Analysis, Princeton.
- Hanke, J./Reitsch, A. (1998)**, Business Forecasting, 6. Auflage, Upper Saddle River (N.J.) u.a.
- Makridakis, S./Wheelwright, S./ Hyndman, R. (1998)**, Forecasting: Methods and Applications, 3. Auflage, New York u.a.
- Mertens, P./Rässler, S. (Hrsg.) (2004)**, Prognoserechnung, 6. Auflage, Heidelberg.
- Rinne, H./Specht, K. (2002)**, Zeitreihen; Statistische Modellierung, Schätzung und Prognose, München.

B. Zitierte Literatur

- Armstrong, J. (Hrsg.) (2002)**, Principles of Forecasting, Boston (Mass.) u.a.
- Backhaus, K./Simon, W. (1981)**, Indikatorprognosen im Investitionsgüter-Marketing, in: *Die Betriebswirtschaft*, Vol. 41, S. 419–422.
- Bates, J./Granger, C. (1969)**, The Combination of Forecasts, in: *Operational Research Quarterly*, Vol. 20, S. 451–468.
- Greene, W.**, Econometric Analysis, 5. Auflage, Upper Saddle River (N.J.) u.a.
- Hamilton, J. (1994)**, Time Series Analysis, Princeton.
- Hammann, P./Erichson, B. (2000)**, Marktforschung, 4. Auflage, Stuttgart.
- Hanke, J./Reitsch, A. (1998)**, Business Forecasting, 6. Auflage, Upper Saddle River (N.J.) u.a.
- Hanssens, D./Parsons, L./Schultz, R. (2001)**, Market Response Models. Econometric and Time Series Analysis, 2. Auflage, Boston (Mass.) u.a.
- Hüttner, M. (1994)**, Vergleich und Auswahl von Prognoseverfahren für betriebswirtschaftliche Zwecke, in: Mertens, P. (Hrsg.), Prognoserechnung, 5. Auflage, Heidelberg, S. 349–363.
- Kmenta, J. (1997)**, Elements of Econometrics, 2. Auflage, New York.
- Lambin, L. (1969)**, Measuring the Profitability of Advertising: An Empirical Study, in: *Journal of Industrial Economics*, Vol. 2, S. 86–103.
- Mahajan, V./Muller, E./Bass, F. (1990)**, New Product Diffusion in Marketing: A Review and Directions for Research, in: *Journal of Marketing*, Vol. 54, Nr. 1, S. 1–26.

- Makridakis, S./Wheelwright, S./Hyndman, R. (1998)**, Forecasting: Methods and Applications, 3. Auflage, New York u.a.
- Meffert, H./Steffenhagen, H. (1977)**, Marketing-Prognosemodelle, Stuttgart.
- Mertens, P. (Hrsg.) (1994)**, Prognoserechnung, 5. Auflage, Heidelberg.
- Mertens, P./Rässler, S. (Hrsg.) (2004)**, Prognoserechnung, 6. Auflage, Heidelberg.
- Niederhübner, N. (1994)**, Indikatorprognosen, in: Mertens, P. (Hrsg.), Prognoserechnung, 5. Auflage, Heidelberg, S. 205–212.
- Palda, K. (1984)**, The Measurement of Cumulative Advertising Effects, Englewood Cliffs (N.J.).
- Ramanathan, R. (1998)**, Introductory Econometrics with Applications, 4. Auflage, Fort Worth.
- Rinne, H./Specht, K. (2002)**, Zeitreihen; Statistische Modellierung, Schätzung und Prognose, München.
- Schneeberger, H. (1994)**, Punkt-, Intervallprognose und Test auf Strukturbruch, in: Mertens, P. (Hrsg.), Prognoserechnung, 5. Auflage, Heidelberg.
- Schneeweiß, H. (1990)**, Ökonometrie, 4. Auflage, Heidelberg.
- Schwarze, J./Weckerle, J. (1982)**, Prognoseverfahren im Vergleich, Braunschweig.
- SPSS Inc. (2007)**, SPSS Base 16.0 - Benutzerhandbuch, Chicago.
- Studenmund, A. (2001)**, Using Econometrics: A Practical Guide, 4. Auflage, Boston (Mass.).
- Weiber, R. (1992)**, Diffusion von Telekommunikation, Wiesbaden.
- Wooldridge, J. (2005)**, Introductory Econometrics: A modern Approach, 3. Auflage, Cincinnati (Ohio) u.a.