

Umstellung der Zeitreihenanalyse

Einleitung

Die monatlichen Daten einer Zeitreihe von einheitlich definierten statistischen Meßgrößen und ihr Verlauf werden durch einen Komplex von Ursachen oder Einflußgrößen bestimmt. Diese Größen prägen die Erscheinungsformen der Zeitreihen. Die Analyse hat die Aufgabe, die den Reihen gemeinsamen **c h a r a k t e r i s t i s c h e n S t r u k t u r e n** festzustellen und in Komponenten aufzugliedern, um Erkenntnisse über Regelmäßigkeiten zu vermitteln. Die Zeitreihenanalyse orientiert sich am formalen Erscheinungsbild einer Reihe und verzichtet bewußt auf die Berücksichtigung von speziellen ökonomischen Theorien oder Modellen, deren Quantifizierung unsicher ist.

Als systematische Bewegungskomponenten einer Zeitreihe gelten

- die **T r e n d k o m p o n e n t e**, die die langfristige Entwicklungsrichtung anzeigt,
- die **K o n j u n k t u r k o m p o n e n t e**, die meist in mehrjährigen Abständen wiederkehrenden, meist konjunkturell bedingten Schwankungen umschreibt, und
- die **S a i s o n k o m p o n e n t e**, die die jährlich wiederkehrenden, jahreszeitlich bedingten Schwankungen angibt.

Es verbleiben die irregulären oder **R e s t s c h w a n k u n g e n**, für deren Erscheinungsbild keine regelmäßigen Einflüsse erkennbar sind. Ihnen zugeordnet sind gelegentlich zu beobachtende **E x t r e m w e r t e** oder Ausreißer, die den Vorgang der Reihenerlegung in die systematischen Komponenten beeinträchtigen oder verzerren können und deswegen zuvor ausgesondert werden.

Die Trend- und die Konjunkturkomponente werden üblicherweise zur sogenannten **g l a t t e n K o m p o n e n t e** zusammengefaßt, weil ihre begriffliche Trennung nicht völlig unumstritten ist¹⁾ und die gemeinsame Darstellung dem allgemeinen Informationsbedürfnis genügt.

Eine besondere Rolle spielen die Einflüsse, die von den Kalenderunregelmäßigkeiten hervorgerufen werden²⁾. Da nämlich die Struktur dieser Unregelmäßigkeiten bekannt ist, kann eine **K a l e n d e r k o m p o n e n t e**, die diese Einflüsse quantitativ beschreibt, als systematische Komponente aufgefaßt werden. Sie hat allerdings keine vergleichbar wichtige eigenständige Bedeutung wie die glatte oder die Saisonkomponente, sondern dient in erster Linie dazu, die Trennung der Komponenten zu verbessern.

Die Zerlegung von Zeitreihen in ihre Komponenten dient dem Zweck, Informationen über die Entwicklungslinien der betrachteten statistischen Reihe in der Vergangenheit, vor allem aber über den Zustand dieser Entwicklung am aktuellen Ende der Zeitreihe zu gewinnen. Dies geschieht in erster Linie dadurch, daß von wiederkehrenden (Kalender- und) Saisoneinflüssen bereinigte Daten dargestellt werden, um die längerfristige Bewegung und ihren jeweiligen Zustand klarer erkennbar zu machen. Ein noch stärker vereinfachtes Bild bietet der Verlauf der bloßen glatten Komponente, aus der jegliche monatspezifischen Besonderheiten eliminiert sind.

Seit mehr als zehn Jahren führt das Statistische Bundesamt monatlich entsprechende Zeitreihenanalysen durch und veröffentlicht die Ergebnisse³⁾. Dabei wird das Berliner Verfahren verwendet mit der Begründung, daß die mathematischen Ansätze genügend allgemein und die Arbeitshypothesen eindeutig und widerspruchsfrei spezifizierbar sein sollten. Außerdem sollte das Analyseverfahren brauchbare Ergebnisse erwarten lassen.

Die Grundlinien des **B e r l i n e r V e r f a h r e n s** bestehen in dem Modell einer additiven Verknüpfung aller Komponenten der Zeitreihe, in der sukzessiven, aber simultan konzipierten Schätzung der glatten und der Saisonkomponente mit Hilfe von gleitenden, zeitinvarianten Filtern und in deren Wirkungskontrolle in der Frequenzdimension⁴⁾.

Die charakteristischen Verläufe der glatten Komponente werden – abgesehen von etwa vereinzelt auftretenden Strukturbrücken – durch ein Polynom mit niedrigem Grad approximiert; in der **V e r s i o n 2** des Berliner Verfahrens (BV2) war ein Polynom dritten Grades gewählt worden, verbunden mit einer regelmäßigen Dreijahreswelle, und zwar für gleitende Approximationsintervalle oder „Stützbereiche“ von stets 31 Monaten⁴⁾. Als

¹⁾ Siehe z.B. Wagner, A.: „Volkswirtschaftliche Beispiele zur Bedeutung der Statistischen Adäquation: Wachstumszyklen und Realkapital“, Angewandte Statistik und Wirtschaftsforschung heute, Vandenhoeck und Ruprecht, 1982, S. 235 ff.

²⁾ Siehe Fürst, G. und Spilker, H.: „Störungen der kurzfristigen Wirtschaftsbeobachtung durch jahreszeitliche und andere wiederkehrende Einflüsse“, Wirtschaft und Statistik 4/1957, S. 199 ff.

³⁾ Siehe „Indikatoren zur Wirtschaftsentwicklung – Zeitreihen mit Saisonbereinigung“, Verlag W. Kohlhammer GmbH.

⁴⁾ Siehe Nourney, M.: „Methode der Zeitreihenanalyse“, Wirtschaft und Statistik 1/1973, S. 11 bis 17.

günstige „Schätzstelle“ innerhalb dieses Bereiches, für die die gewonnene Lösung gültig sein soll, wurde der 19. Monat gewählt, solange der Stützbereich Monat für Monat über die gesamte Reihe gleiten konnte; sobald jedoch der Stützbereich an das Ende einer Reihe gelangt war, wurde er dort belassen, während die Schätzstelle schrittweise vom 19. bis zum 31. Monat versetzt wurde. Zur approximativen Schätzung der Saisonkomponente sind trigonometrische Polynome verwendet worden, die die regelmäßige Jahresschwingung und ihre fünf Oberwellen umschließen; die am Ende der Reihen variierenden Stützbereichslängen überdecken dreieinhalb bis vier Jahre. Diese Anpassungsmodelle wurden mit den Mitteln der Regressionsschätzung behandelt, aus der bestimmte Filter, das sind Folgen von Gewichtungselementen, hervorgehen, mit denen die Beobachtungswerte im Stützbereich gliederweise zu multiplizieren sind. Die Wirkung der Filter läßt sich durch ihre Transformation in die Frequenzdimension in der Anordnung nach Frequenzen bzw. Wellenlängen prüfen und bewerten.

Die Extremwernerkenntnis und -bereinigung beruht auf der gleitenden Berechnung eines bedingten Erwartungswertes für den jeweiligen Folgemonat und dessen Vergleich mit dem tatsächlich beobachteten Wert⁴). Die sogenannte arbeitstägliche Bereinigung stützt sich auf die Restkomponente, aus der die Einflüsse, die auf die monatlich wechselnde Zusammensetzung aus Wochentagen und Feiertagen zurückzuführen sind, ermittelt und eliminiert werden⁴).

Unter Bezugnahme auf die Erkenntnis, daß die zu analysierenden Zeitreihen (z.B. in der Frequenzdimension messbare) unterschiedliche Eigenschaften besitzen und daß auch die verfügbaren Schätzfilter verschiedene Charakteristiken aufweisen, verfolgte die Weiterentwicklung des Berliner Verfahrens (BV) das Ziel, unterschiedliche und reihenspezifisch besonders günstige Filter festzulegen. Zur Vorbereitung dieser seit 1975 eingesetzten *V e r s i o n 3* (BV3)⁵) wurde der Vorrat an zulässigen Approximations- oder Basisfunktionen für die Schätzung der glatten Komponente auf fünf Typen erweitert, indem neben gewöhnlichen Polynomen auch bis zu zwei Mehrjahreswellen (Drei- und Fünfjahreswelle) einbezogen werden konnten und der Polynomgrad auf zwei oder eins zurückführbar war. Die Festlegung auf einen dieser Typen und den passenden Stützbereich vollzog sich je Reihe auf der Grundlage von Vergleichsrechnungen für alle in Betracht kommenden Filter der Saisonbereinigung am Rand unter Verwendung des Kriteriums, daß die Summe der quadratischen Abweichungen zwischen Original- und bereinigten Werten im gesamten Spektralbereich (mit Ausnahme der Saisonfrequenzen) minimiert werden sollte. Die übrigen Filter für die Schätzung an der vorletzten, drittletzten, ..., sechstletzten Stelle wurden – mit beibehaltenem Funktionstyp – durch geeignete Variationen der Stützbereichslängen optimiert; für den im mittleren Zeitreihenbereich gleitenden Filter wurde eine eigene Optimallösung zum selben Funktionstyp gefunden und ein geeigneter Anschluß an den sechstletzten Filter geschaffen. Auf diese Weise sind für die mehr als 300 Zeitreihen der monatlichen Veröffentlichung 47 verschiedene Filterkombinationen entstanden.

Testrechnungen zeigten eine merkliche Verbesserung der Randschätzungen gegenüber BV2. Die praktische Anwendung von BV3 schloß aber auch Schwierigkeiten ein. Die Zuordnung einer der vorhandenen 47 Filterkombinationen zu jeder neu behandelten Zeitreihe erfordert beachtliche Rechenzeiten und eine ergänzende Bearbeitung, weil keine volle Automatisierung eingebaut worden ist. Ähnliches gilt für die Überprüfung einer bereits bestehenden Zuordnung nach Ablauf mehrerer Jahre oder nach einer Umbasierung der Ausgangsreihen. Zum anderen stellte sich das Modell der individuellen Optimierung für die sechs letzten Werte einer Reihe insoweit als unbefriedigend heraus, als die aufeinanderfolgenden Schätzfilter gegenseitig nicht immer genügend abgestimmt waren, so daß die Analyseergebnisse – etwa der Verlauf der glatten Komponente – in einigen Fällen unbegründete Ausschläge aufwiesen. Ein weiterer Nachteil zeigte sich darin, daß die Schätzwerte für die glatte Komponente in Phasen des Auf- oder Abschwungs am Ende der Reihe dazu neigen, derartige Tendenzen zu überzeichnen. Es war deshalb naheliegend und geboten, die Entwicklungsarbeiten fortzuführen, um die erkannten Mängel zu beseitigen. In mehrjähriger Arbeit ist eine *n e u e V e r s i o n* BV4 entstanden, die auch als ein frequenz-determiniertes Filter-Verfahren (FDF 4) bezeichnet werden kann. Sie stützt sich im Kern auf die gleichen Bestandteile wie ihre Vorläufer, ist jedoch – wie nachstehend dargelegt wird – um zwei methodische Elemente ausgebaut worden und wird ab Ende 1983 vom Statistischen Bundesamt eingesetzt.

1 Ein umfassenderes Schätzmodell

Die Zeitreihenanalyse soll die beobachteten Werte oder „Originalwerte“ x_t für zahlreiche aufeinanderfolgende, als äquidistant unterstellte Zeitpunkte oder Zeiträume t in eine glatte Komponente g_t , eine Saisonkomponente s_t und eine irreguläre oder Restkomponente u_t zerlegen, wobei gemäß

$$x_t = g_t + s_t + u_t \quad t = 1, 2, \dots, T$$

ein *a d d i t i v e s* *Z u s a m m e n w i r k e n* dieser Komponenten unterstellt wird. Dies schließt nicht aus, daß Änderungen im Verlauf einer Komponente auch von Änderungen bei anderen Komponenten begleitet und die Teilprozesse korreliert sein können. Ein streng systematischer Zusammenhang im Sinne der Proportionalität – wie bei einem multiplikativen Ansatz – wird jedoch nicht angenommen.

⁵) Siehe Nourney, M.: „Weiterentwicklung des Verfahrens der Zeitreihenanalyse“, *Wirtschaft und Statistik* 2/1975, S. 96 bis 101.

Die mittel- und längerfristigen Einflüsse auf die Werte einer Zeitreihe werden sich im Zeitverlauf im wesentlichen nur langsam und kontinuierlich ändern. Die möglichen Realisationen dieses Entwicklungsprozesses werden deshalb einen „glatten“ Verlauf haben, der höchstens bei einer geringen Zahl von Zeitpunkten von Strukturbrüchen gestört ist. Das sachlich begründete Postulat eines glatten Verlaufs führt auf die mathematisch-formale Forderung, daß sich die Realisationen der *g l a t t e n K o m p o n e n t e* g_t für eine gewisse Anzahl benachbarter Zeitpunkte durch Polynome eines relativ niedrigen Grades mit genügender Genauigkeit approximieren lassen (durch sog. Basisfunktionen). Der Grad des Polynoms soll hierbei im Verhältnis zur Länge des Approximations- oder Stützbereichs, d.h. zur Anzahl der für die Schätzung herangezogenen benachbarten Zeitpunkte, möglichst klein sein. Da jeder Approximationsbereich nicht nur konjunkturelle Maxima und Minima, sondern auch Wendepunkte hinsichtlich der Krümmungsrichtung enthalten kann, wurde bisher als Standardlösung ein Polynom dritten Grades, verbunden mit Schwingungskomponenten im Drei- und/oder Fünfjahresrhythmus, bei einer Länge des Stützbereichs von etwa drei Jahren zugrunde gelegt.

Der neue Ansatz beschränkt sich auf Polynome dritten Grades von der Art

$$\hat{g}_t = a_0 + a_1 t + a_2 t^2 + a_3 t^3,$$

mit denen der Schätzwert \hat{g}_t für die glatte Komponente als Funktion der Zahl t aufgefaßt wird und die Koeffizienten a_i ($i = 0,1,2,3$) zufallsbedingte Größen sind, die auch gewissen zeitlichen Veränderungen unterworfen sein können. Zur Schätzung dieser Koeffizienten wird die Regressionsmethode benutzt, die sich des Prinzips der kleinsten (Abweichungs-) Quadrate zwischen den Originalwerten x_t und den Schätzwerten \hat{g}_t bedient (hier zur Vereinfachung abgesehen von Saisoneinflüssen):

$$\sum_{t \in \Theta} (x_t - \hat{g}_t)^2 = \text{Minimum}$$

Die Summe über die Abweichungsquadrate erstreckt sich über den jeweils vorgegebenen Approximations- oder Stützbereich Θ . Eine logische Schwierigkeit⁶⁾ besteht darin, daß die geschätzten Koeffizienten a_i nach dem üblichen Regressionsmodell für den gesamten Stützbereich gültig und dort feste Größen sind, sie tatsächlich aber nur für einen Zeitpunkt, nämlich die gewählte Schätzstelle, verwendet werden und ihre „Veränderlichkeit“ erst dadurch entsteht, daß zwecks Schätzung an den nächsten Stellen der gesamte Stützbereich um entsprechend viele Zeitintervalle versetzt wird (gleitende Schätzung). Dieser Schwierigkeit wird im neuen Ansatz tendenziell dadurch begegnet, daß in die Minimalbedingung des Regressionsansatzes ein *G e w i c h t u n g s s y - s t e m* w_t eingeführt wird:

$$Q_g = \sum_{t \in \Theta} w_t (x_t - \hat{g}_t)^2 = \text{Minimum} .$$

Die Festlegung des Gewichtungssystems für das Intervall Θ ist grundsätzlich frei. Ist der Prozeß der Restschwankungen u_t autokorreliert, so könnte theoretisch für w_t die Inverse der Kovarianzmatrix dieses Restprozesses gewählt werden, um auf diese Weise zu besten linearen unverzerrten Schätzwerten für g_t zu gelangen⁷⁾. Die quantitativen Eigenschaften des Restprozesses sind jedoch nicht bekannt und ließen sich allenfalls auf der Grundlage einer provisorischen Analyse abschätzen.

Die Gewichte w_t werden hier statt dessen so gesetzt, daß sie an der Schätzstelle τ und in ihrer näheren Umgebung relativ groß, in der Nähe der Enden des Stützbereichs Θ dagegen klein sind. Die einfachste funktionale Form für dieses gewünschte Prinzip ist der gebrochen-lineare Ansatz hinsichtlich t

$$w_t = 1 - \frac{|t - \tau|}{D},$$

worin D die größere der beiden Entfernungen zwischen der Schätzstelle τ und den Enden des Stützbereichs bezeichnet. Diese Gewichtungsfunktion ist so normiert, daß sie ihren höchsten Wert $w_t = 1$ an der Stelle $t = \tau$ annimmt und von dort nach beiden Seiten hin linear abfällt, aber stets positiv bleibt. Handelt es sich z.B. um eine symmetrische Schätzung der betreffenden Komponente, bei der die Stelle τ in der Mitte des Stützbereichs Θ liegt, dann stellt w_t eine symmetrische Dreiecks- oder Dachfunktion dar mit einer Spitze in der Mitte und gegen Null hin abfallenden Seiten.

Eine solche Gewichtung der Abweichungsquadrate ist zulässig. Sie „konzentriert“ die Schätzprozedur ihrer Wirkung nach auf die engere Umgebung der gerade zu schätzenden Stelle τ und unterstreicht den Ausschnitt-

⁶⁾ Siehe Oswald, E.: „Bemerkungen zum Berliner Verfahren“ in: Allgemeines Statistisches Archiv, Heft 2/3, 1972, S. 191 bis 196.

⁷⁾ Siehe Heiler, S.: „Überlegungen zu einem statistischen Modell einer wirtschaftlichen Zeitreihe und einem daraus resultierenden Analyseverfahren“, DIW-Beiträge zur Strukturforschung, Heft 7, 1969, S. 19 bis 43.

Charakter der gleitenden Schätzung, die sich immer nur auf Teilintervalle der gesamten Zeitreihe, nämlich die Stützbereiche, beschränkt. Die Gewichtung findet übrigens ihre Analogie in der üblichen Verwendung von sog. Fenstern beim Schätzen von Spektralintensitäten in der Frequenzdimension⁸⁾.

Das Einbeziehen einer derartigen Gewichtung der Abweichungsquadrate bietet eine Reihe von **Vorteilen**. Die Beobachtungswerte im Stützbereich erhalten einen um so bestimmteren Einfluß auf den Schätzwert unter Bezugnahme z.B. auf das unterlegte Polynom dritten Grades (für die glatte Komponente), je näher sie an der Schätzstelle liegen; entferntere Werte prägen die Koeffizienten dieser Basisfunktionen nur schwach. Dadurch verliert der ursprünglich variierend gewählte Funktionstyp, der durch die am Ende der Zeitreihe vorhandenen Verhältnisse wesentlich bedingt und für die reihenspezifisch unterschiedlichen Lösungen nach BV3 maßgebend war, so stark an Bedeutung, daß es möglich wird, einen einheitlichen, für sämtliche statistischen Monatsreihen universell gültigen Funktionsansatz zu verwenden. Zum anderen kann erwartet werden, daß das Einbeziehen immer neuer Beobachtungswerte in den gleitend eingesetzten Stützbereich und das gleichzeitige Entlassen von Beobachtungswerten am anderen Ende des Stützbereichs auf die Schätzergebnisse nun kontinuierlicher wirksam wird, weil diese Änderungen an den Rändern des Stützbereichs geschehen und dort ein zunächst geringes Bedeutungs-Gewicht erhalten. Für benachbarte Monate entstehen somit besser vergleichbare Ergebnisse.

Diese ohnehin erwünschte Wirkung gewinnt einen besonderen Wert bei der Schätzung der Saisonkomponente, bei der zwingend Lösungen für die Stellen $\tau = 12, 24, 36, \dots$ innerhalb des Stützbereichs notwendig sind. Dies besagt, daß bei der Schätzung in der Nähe des aktuellen Randes einer Zeitreihe irgendeinmal ein Wechsel von $\tau = 24$ auf 36 und später auf 48 notwendig wird, wodurch die linke Seite des Stützbereichs plötzlich um jeweils 11 Beobachtungswerte erweitert wird. Der nachteilige Effekt einer solchen schlagartig verbreiterten Schätzgrundlage, der sich in einer Art „Saisonsprung“ niederschlagen kann, wird infolge der relativ geringeren Gewichtung am Rand des Stützbereichs gedämpft.

Die Gewichtung mit einer Dreiecksfunktion führt ferner zu einer treffenderen und flexibleren Anpassung an veränderliche Kurvenverläufe der glatten Komponente. Die Flexibilität der Saisonschätzung, die wegen der Beschränkung der Länge ihres Stützbereichs auf ungefähr vier Jahre ohnehin vergleichsweise hoch ist, wird durch die Gewichtung gesteigert; auch etwas raschere Änderungen der Saisonfigur werden von den Schätzwerten gut nachgezeichnet. Für die Analysen in der Nähe des Randes der Reihen gilt entsprechend, daß die dort etwa veränderten Verhältnisse hinsichtlich der Amplitude und Struktur der Saisonkomponente frühzeitiger und treffender geschätzt werden können als mit den bisherigen Varianten des Berliner Verfahrens, bei denen lediglich ein „ungewogener Durchschnitt“ der Daten im Stützbereich ermittelt wird. Dessen Ergebnis mag zwar für die tatsächliche Situation in der Mitte des Stützbereichs zutreffend sein, ist es aber weniger für den Rand dieses Bereichs, wenn z.B. eine allmähliche Amplitudenveränderung der Saisonkomponente eintritt.

Zur konkreten Durchführung der eigentlichen Analyse werden **Filter** benutzt, also durch die jeweilige Länge des Stützbereichs begrenzte Folgen von Elementen, mit denen die Ausgangswerte einzeln multipliziert und dann zusammengefaßt werden. Sie entstehen aus einer **Rechenprozedur**, die mit der Festlegung der Basisfunktionen (Polynom dritten Grades für die glatte, trigonometrisches Polynom für die Saisonkomponente) und der Länge des Stützbereichs beginnt und eine varianzminimale Anpassung dieser Approximationsfunktionen an die Beobachtungswerte x_t der Zeitreihe nach der Regressionsmethode vornimmt. Hierbei werden die bekannten, nur von $t = 1, 2, \dots, N$ abhängigen Werte der einzelnen Basisfunktionen⁹⁾ als Matrix F behandelt, die entsprechend der Länge des Stützbereichs N Zeilen besitzt und q Spalten, worin $q (= 15)$ die Gesamtzahl der Glieder der Basisfunktionen bezeichnet. Die ersten Ableitungen der mit w_t gewichteten Minimalbedingung nach den Koeffizienten $a_0, a_1, a_2, a_3, b_1, b_2, \dots, b_{11}$ der Basisfunktionen werden gleich null gesetzt, und diese q Normalgleichungen lassen sich nach den q Koeffizienten auflösen, mit deren Hilfe die konkreten approximierenden Funktionen gefunden werden. Zu jeder Stelle τ im Stützbereich Θ gibt es dann einen (geschätzten) Wert der approximierenden Funktion. Die entsprechenden Matrixrechnungen (Multiplikationen, Inversion) können so weit vorangetrieben werden, daß nicht erst Schätzwerte für die Koeffizienten a_0, a_1, \dots, b_{11} entstehen, sondern unmittelbar eine Matrix G mit N Zeilen und N Spalten, deren einzelne Zeilen elementweise mit dem Spaltenvektor der N Beobachtungswerte im Stützbereich zu multiplizieren sind. Jede Zeile der Matrix G kann folglich als Filter gedeutet werden, und die Wahl der Zeile aus G bestimmt die Stelle τ im Stützbereich, für die die betreffende Schätzung gültig ist. Alle Filter sind zeitinvariant, weil sie von den Basisfunktionen, der Länge des Stützbereichs, der Gewichtung der Abweichungsquadrate und der Schätzstelle τ abhängen, nicht aber von den zeitbedingten Beobachtungswerten.

Es wäre nun zwar möglich, die für die Zeitreihenanalyse besonders gefragte Saisonkomponente unmittelbar zu schätzen; ihre Ergebnisse sind jedoch zuverlässiger, wenn aus den Originalwerten zuvor die glatte Komponente eliminiert wird. Der **Ablauf des Analyseverfahrens** beginnt¹⁰⁾ deshalb mit der Schätzung der

⁸⁾ Das dortige Analogon ist das Bartlett-Fenster. Siehe z.B. Dub, W.: „Zur Klassifikation der Fenster in der Spektralanalyse“ in der Reihe „Quantitative Methoden der Unternehmensplanung“, Heft 11, 1980.

⁹⁾ Das Polynom dritten Grades zur glatten Komponente enthält 4 Glieder mit den Koeffizienten a_0, a_1, a_2, a_3 , das trigonometrische Polynom zur Saisonkomponente 11 Glieder von sin- und cos-Funktionen der Jahreswelle und ihrer 5 Oberwellen mit den Koeffizienten b_1, b_2, \dots, b_{11} (ohne das sin-Glied der letzten Oberwelle).

¹⁰⁾ Abgesehen von einer gegebenenfalls vorweg durchgeführten Extremwert- und Kalenderbereinigung (siehe Abschnitte 6 und 5).

glatten Komponente und ihrer Herauslösung aus den Originaldaten und wird mit der Saisonschätzung fortgesetzt. Die sog. Teilschätzungen der beiden Komponenten berücksichtigen hierbei die (etwa gegebene) Existenz der jeweils anderen Komponente auf eine einfache Weise: In den Regressionsansatz und die Bedingung für die Abweichungsquadrate Q werden stets die Basisfunktionen der beiden Komponenten einbezogen; die eigentliche Schätzung beschränkt sich dann allerdings auf nur je eine Komponente, indem nur deren Koeffizienten der Basisfunktionen beteiligt, die übrigen Koeffizienten aber übergangen werden. Damit läßt sich erreichen und gewährleisten, daß beide Komponenten sauber voneinander getrennt werden und bei der zuerst vollzogenen Schätzung der glatten Komponente nicht fälschlich etwa bereits Teile der Saisonkomponente mit einfließen.

Bevor auf die konkrete Festlegung der universellen Filter für die Ermittlung von glatter und Saisonkomponente ausführlicher eingegangen wird, muss noch eine zweite Modellerweiterung gegenüber BV2 und BV3 erwähnt und begründet werden. Die Gewichtung der Abweichungsquadrate im Regressionsansatz reicht nicht aus, die Ergebnisse der glatten Komponente in unmittelbarer Nähe des Reihenendes befriedigend zu stabilisieren. Nach wie vor haben diese Ergebnisse die Tendenz, bei Auf- oder Abschwüngen der konjunkturellen Entwicklung übertreibend nach oben oder unten auszuspringen und damit die tatsächlichen Bewegungen zu überzeichnen. Wegen der besonderen Bedingungen der zwangsläufig unsymmetrischen Schätzansätze in der Nähe des Randes muß die Gruppe der Basisfunktionen zwecks Stabilisierung der Ergebnisse dort erweitert werden um eine lineare Funktion, d.h. ein Polynom ersten Grades mit zwei Koeffizienten, das selbst keine Schwingungen besitzt. Der technischen Einfachheit halber werden die resultierenden Filter zu beiden Funktionsgruppen gliedweise zusammengefaßt, wobei im so **k o m b i n i e r t e n F i l t e r** der Anteil aus der linearen Funktionsgruppe allmählich wächst, während die Schätzstelle τ an den Rand des Stützbereichs – und damit an das Ende der Zeitreihe – rückt. Die Stabilität und Verlässlichkeit der provisorischen Schätzungen am Rand, die später nach Hinzutreten weiterer Beobachtungswerte erneuert und hierdurch gegebenenfalls revidiert werden, lassen sich auf diesem Weg merklich verbessern.

Die Erweiterung um lineare Funktionen im Bereich der glatten Komponente sowie um das Arbeitsmittel einer geeigneten Kombination von Filtern wird auch für die Ermittlung der Saisonkomponente nutzbar gemacht.

2 Die Filter zur Komponentenschätzung

Geeignete Filter sind das Mittel zur Schätzung der glatten und der Saisonkomponente. Die Filter entstehen auf der Grundlage von Approximationsansätzen unter Verwendung der Methode der kleinsten Quadrate (siehe Abschnitt 1). Sie sind abhängig von und eindeutig bestimmt durch

- die Länge N des Stützbereichs Θ , die auch die Länge des Filters bestimmt;
- die Schätzstelle τ innerhalb des Stützbereichs Θ , die sich aus der Zeilennummer der „Schätzmatrix“ G ergibt;
- den Polynomgrad P_g der Basisfunktionen für die glatte Komponente, der 3 oder 1 beträgt (der Polynomgrad P_s der trigonometrischen Polynome für die Saisonkomponente ist stets 6);
- die Position τ_w innerhalb des Stützbereichs Θ , an der die Gewichtungsfunktion w_t der Abweichungsquadrate ihren Maximalwert 1 besitzt (τ_w und τ müssen nicht unbedingt übereinstimmen);
- das Anteilsgewicht p (≤ 1) des betreffenden individuellen Filters im kombinierten Filter und schließlich
- die Kennung, für welche der beiden Komponenten (g, s) der Filter gültig ist.

Zu jeder beliebigen vorgegebenen Kombination dieser Parameter läßt sich ein Filter berechnen, mit dessen Elementen die Ausgangswerte im Stützbereich zu multiplizieren sind.

Die Zeitreihenanalyse wird im wesentlichen als gleitende Schätzung durchgeführt. Der nur einen relativ kurzen Ausschnitt der Zeitreihe überdeckende Stützbereich Θ wird, am Anfang der Reihe beginnend, Monat für Monat zu aktuelleren Daten hin versetzt, und seine irgendwo im mittleren Teil von Θ liegende Schätzstelle τ bleibt dabei, relativ gesehen, stets dieselbe, so daß Ergebnisse für zahlreiche aufeinanderfolgende Monate entstehen (als gewogene „gleitende Durchschnitte“). Da hierbei die Länge des Stützbereichs und die Stelle τ konstant sind, handelt es sich um ein und denselben Filter, der **M i t t e l f i l t e r** genannt wird, weil er für den gesamten mittleren Bereich einer Zeitreihe eingesetzt wird.

Sobald der Stützbereich das Ende einer Zeitreihe erreicht hat, kann er nicht noch weiter versetzt werden. Die Schätzstelle τ liegt in seinem mittleren Teil, so daß die Ermittlung der Ergebnisse dort zunächst abbricht. Es bleiben somit am Rand der Reihe noch mehrere Monate offen, für die eine Schätzung mit dem Mittelfilter nicht möglich ist. Dieser Filter muß deshalb ersetzt werden durch mehrere andere sogenannte **R a n d f i l t e r**, bei denen die Position von τ so verändert ist, daß ihr Abstand vom „aktuellen“ Rand des Stützbereichs Schritt für Schritt kürzer wird.

Für die Zeitreihenanalyse wird also je eine **F i l t e r f o l g e** zur Schätzung der glatten und der Saisonkomponente gebraucht, die sich aus dem Mittelfilter und mehreren, jeweils anschließenden Randfiltern zusammensetzt.

Die konkrete Festlegung der Analysefilter richtet ihr Hauptaugenmerk auf die Ähnlichkeit der aufeinanderfolgenden Filter, damit die Analyse im Mittelbereich einer Reihe eine möglichst gleichwertige Fortsetzung bis zum Rand findet. Als Kriterium zur Beurteilung der Ähnlichkeit wird die Frequenzantwortfunktion eines jeden Filters, das ist die Transformierte des Filters in die Frequenzdimension, herangezogen (siehe Abschnitt 3). Das eigentliche Problem ergibt sich aus der Tatsache, daß Mittelfilter zwar generell gute bis optimale Eigenschaften besitzen, die Eigenschaften von Randfiltern jedoch prinzipiell um so ungünstiger werden, je näher die Schätzstelle τ an den Rand gerückt wird.

Als Stützbereichslänge N_g für den Mittelfilter der *g l a t t e n K o m p o n e n t e* wird die Spanne von 27 Monaten gewählt, weil diese kurz genug ist, um auch noch konjunkturelle Bewegungen, die sich z.B. zwei- bis dreijährlich wiederholen, in nahezu vollem Ausmaß ermitteln zu können. Andererseits ist diese Spanne genügend lang, um nicht mehr von individuellen Unregelmäßigkeiten gestört zu werden. Zur Vermeidung jeglicher Phasenverschiebungen, d.h. zeitlicher Vor- oder Nachläufe der Ergebnisse, wird der Mittelfilter (im Gegensatz zu BV2 und BV3) *s y m m e t r i s c h* angesetzt, so daß seine Schätzstelle τ in der Mitte des Stützbereichs, also beim 14. Monat, zu liegen kommt. Stellt man alle Parameter dieses Filters zur glatten Komponente in der Kurzform $f_g(N_g, \tau; \tau_w, P_g; p)$ zusammen, so lautet sie für den Mittelfilter $f_g(27, 14; 14, 3; 1)$. Dieser Filter ist auf den gesamten Mittelbereich einer Zeitreihe anwendbar bis auf Lösungen für die ersten und die letzten 13 Monate der Reihe. Für die nächste Festlegung wird ein geeigneter (Rand-)Filter gesucht, dessen Schätzstelle τ nur noch genau 12 Monate vom „aktuellen“ Ende des Stützbereichs entfernt ist (für den „dreizehnt-letzten“ Monat), für den also $N_g - \tau = 12$ ist. Hierfür kommen verschiedene mögliche Filter in Betracht, die sich im wesentlichen nur noch durch ihre Länge N_g unterscheiden. Aufgrund eines quantitativen Vergleichs der Frequenzantwortfunktionen dieser möglichen Filter mit der des bereits festgelegten Mittelfilters vornehmlich hinsichtlich des Realteils im entscheidenden Niederfrequenzbereich ist entschieden worden, daß der erste der Randfilter um einen Monat länger als der Mittelfilter sein soll, in Kurzform also durch $f_g(28, 16; 16, 3; 1)$ zu beschreiben ist.

In entsprechender Weise wird bei der Festlegung der folgenden Randfilter verfahren. Der Abstand $N_g - \tau$ zum Rand des Stützbereichs und – da dieser Bereich nun immer bis zum Ende der Zeitreihe reicht – auch zum aktuellen Rand der Reihe wird ständig um einen Monat verringert und der Vergleich der Spektraleigenschaften der neuerlich in Betracht kommenden Filter auf ihren Vorläufer bezogen. So setzt sich für die Schätzung der glatten Komponente die Filterfolge fort gemäß: $f_g(29, 18; 18, 3; 1)$, $f_g(30, 20; 20, 3; 1)$, $f_g(30, 21; 21, 3; 1)$, $f_g(29, 21; 21, 3; 1)$, $f_g(28, 21; 21, 3; 1)$, $f_g(26, 20; 20, 3; 1)$. Hierbei werden die spektralanalytisch gemessenen Eigenschaften der Filter allmählich ungünstiger, vor allem auch hinsichtlich des Imaginärteils der Frequenzantwortfunktion, der für Phasenverschiebungen verantwortlich ist und im Idealfall null sein sollte.

Eine weitere Fortsetzung der Filterfolge nach dem begonnenen Modell schien nicht mehr vertretbar, weil insbesondere in der Spanne zwischen Jahres- und Halbjahreswelle der Realteil der Frequenzantwortfunktion zu stark wächst. In dieser Spanne haben Filter auf der Basis von linearen Funktionen wesentlich bessere Konditionen. Da die Einbeziehung dieser Gruppe von Filtern auch logisch vernünftig ist (siehe Abschnitt 1), sind schließlich die folgenden *K o m b i n a t i o n e n* entwickelt worden:

$$f_g(25, 20; 20, 3; 11/12) + f_g(25, 20; 25, 1; 1/12)$$

$$f_g(25, 21; 21, 3; 10/12) + f_g(24, 20; 24, 1; 2/12)$$

$$f_g(25, 22; 22, 3; 9/12) + f_g(23, 20; 23, 1; 3/12)$$

$$f_g(25, 23; 23, 3; 8/12) + f_g(22, 20; 22, 1; 4/12)$$

$$f_g(26, 25; 25, 3; 7/12) + f_g(21, 20; 21, 1; 5/12)$$

$$f_g(27, 27; 27, 3; 6/12) + f_g(20, 20; 20, 1; 6/12)$$

Der Gewichtsanteil der neuen Gruppe von Filtern aus linearen Funktionen (rechter Summand) nimmt monoton von zunächst 1/12 bis auf 1/2 zu. Die Filter dieser Gruppe werden ständig kürzer, die Stelle τ_w der maximalen Gewichtung der Abweichungsquadrate befindet sich aber stets am Ende des Stützbereichs, weicht also meist von der Schätzstelle τ ab.

Zur Schätzung der glatten (und Saison-)Komponente am vorderen Rand der Zeitreihe werden die gleichen Lösungen benutzt, jedoch in spiegelbildlich umgekehrter Reihenfolge der einzelnen Filter und ihrer einzelnen Elemente.

Für die Schätzung der *S a i s o n k o m p o n e n t e* wird ebenfalls ein symmetrischer Mittelfilter bestimmt, dessen Länge N_s wegen der Schätzstelle $\tau = 24$ (Vielfaches von 12 erforderlich!) 47 Monate umfaßt. Aus Gründen der besseren Kondition im Niederfrequenzbereich ist bereits dieser Filter geringfügig mit einem aus der Gruppe linearer Funktionen überlagert. Die Kurzformen des Mittelfilters und seiner Nachbarn lauten

$$f_s(47, 24; 24, 3; 6/7) + f_s(47, 24; 24, 1; 1/7)$$

$$f_s(46, 24; 24, 3; 6/7) + f_s(46, 24; 24, 1; 1/7)$$

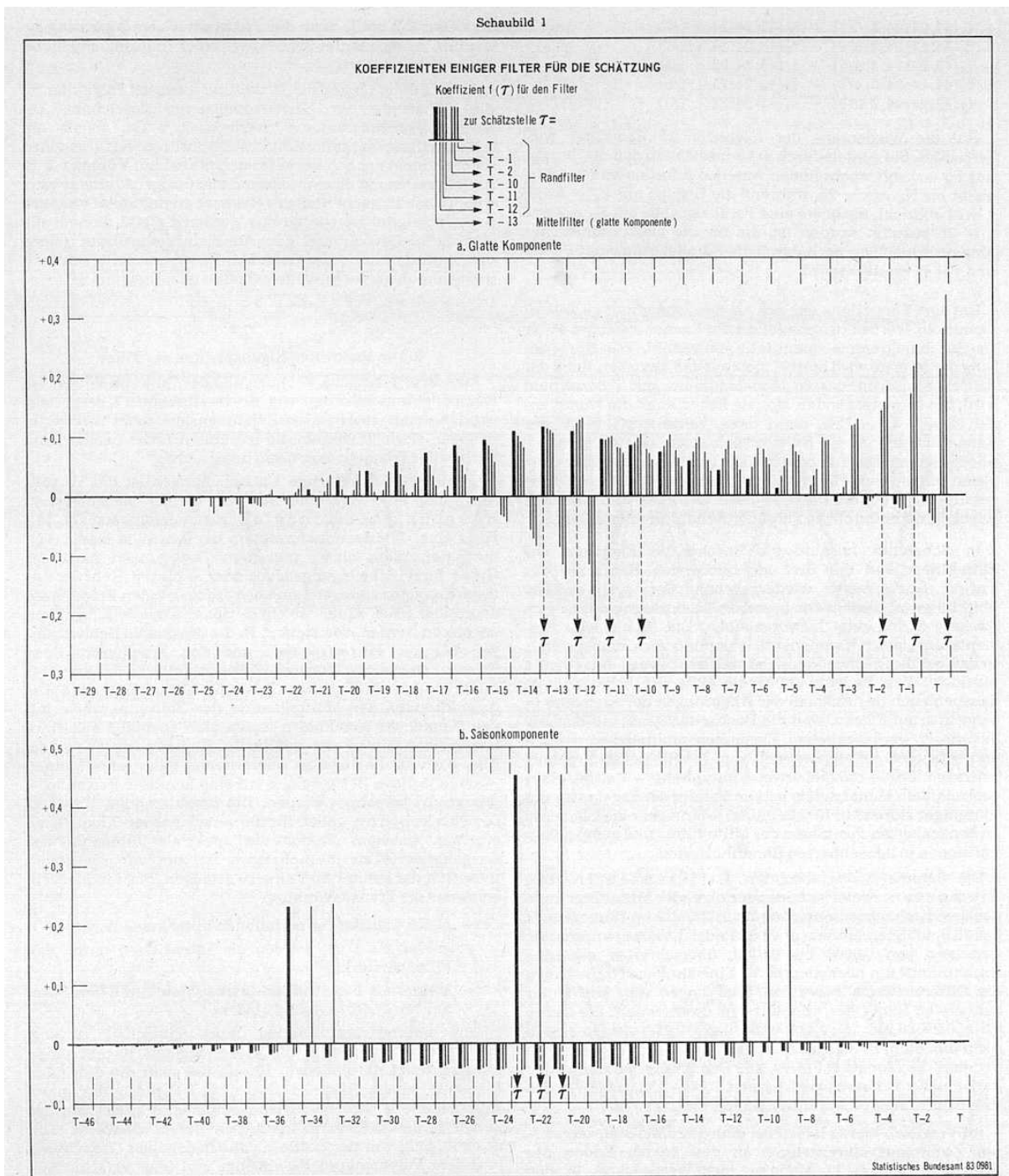
$$f_s(45, 24; 24, 3; 4/5) + f_s(45, 24; 23, 1; 1/5)$$

$$f_s(44, 24; 24, 3; 4/5) + f_s(44, 24; 23, 1; 1/5)$$

$$f_s(43, 24; 24, 3; 2/3) + f_s(43, 24; 22, 1; 1/3)$$

Auf die Wiedergabe der folgenden 19 Randfilter wird verzichtet. Sie sind dadurch gekennzeichnet, daß die Gruppe mit $P_g = 1$ mit wechselnden Anteilen p fortlaufend beteiligt bleibt bis $N_s = \tau = 24$, während die Gruppe mit $P_g = 3$ und $\tau = 24$ abbricht, nachdem eine Parallelgruppe mit $P_g = 3$ und $\tau = 36$ beteiligt worden ist, die für die vier randnächsten Monate schließlich noch durch die Parallelgruppe mit $P_g = 3$ und $\tau = 48$ verstärkt wird.

Um eine Vorstellung von der Art der Schätzfilter zu BV4 zu vermitteln, werden in Schaubild 1 die Elemente einiger Filter als Strichdiagramme quantitativ dargestellt. Die Numerierung der Monate wird hierbei stets auf den aktuellen Rand der Zeitreihe bezogen, dessen Monatsnummer mit T bezeichnet wird; die vorausgehenden Monate haben dann die Nummern T-1, T-2, T-3, ... Dies dient dazu, benachbarte Filter der Randschätzung, deren



Stützbereich stets bis zum Monat T reicht, vergleichend nebeneinanderzustellen. In diesen Vergleich wird auch der Mittelfilter einbezogen, der zwar über die gesamte Reihe gleitet, aber zuletzt auch bis zum Monat T reicht, bevor er durch die Folge der Randfilter ersetzt wird.

In Schaubild 1a sind als Striche die Elemente des Mittelfilters und der drei anschließenden Randfilter zur glatten Komponente wiedergegeben; der symmetrische Mittelfilter ist hierbei zur besseren Orientierung durch eine stärkere Strichbreite hervorgehoben. Die Summe der Elemente zur glatten Komponente ist automatisch stets gleich 1, so daß bei ihrer Anwendung das mittlere Niveau einer Reihe resultiert. Von Bedeutung für die Güte der Schätzung ist daneben auch das Ausmaß der Änderungen der Elemente in benachbarten Filtern, weil die Beobachtungswerte sukzessiv mit diesen veränderlichen Elementen multipliziert werden und entsprechend abgewandelt zum jeweils neuen Ergebnis beitragen. Der - möglicherweise untypische - Beitrag eines beobachteten Monatswerts tritt im Verlauf der Ergebnisse um so weniger störend in Erscheinung, je ähnlicher die Elemente an benachbarten Positionen des Mittelfilters und auf gleichen Positionen in benachbarten Randfiltern sind.

Die Summe Δ der absoluten Differenzen in den Elementen zweier aufeinanderfolgender Mittelfilter zu je gleichen Reihenpositionen beträgt 0,295, also im Durchschnitt der 28 beteiligten Monate etwa 0,01 oder 1 %; sie variieren im einzelnen von 0,0006 bis 0,0159, überschreiten also den Durchschnitt um höchstens 60 %. Eine ähnliche Größe haben die Differenzwerte Δ bei den Übergängen vom Mittelfilter zum ersten Randfilter, von dort zum zweiten u.s.f. Sie lauten 0,319; 0,290; 0,246; ...; 0,414; 0,458; 0,426; 0,432, steigen also in nicht monotoner Folge bis höchstens 0,458 – im Durchschnitt der dort 25 Monate 0,018 – an. Der größte vorkommende individuelle Differenzwert beträgt 0,0804 (für den Wechsel vom Schätzmonat $\tau = T-3$ zum Monat $\tau = T-2$).

Im Vergleich hierzu treten bei dem gewöhnlichen zentrierten Zwölfmonatsdurchschnitt an den beiden Enden des Stützbereichs von 13 Monaten vier Wertwechsel in den Elementen auf, die jeweils $1/24 = 0,042$ betragen und somit viermal so groß sind wie für die Mittelschätzung nach BV4 im Durchschnitt der Monate; während sich die unumgänglichen Elementdifferenzen beim Zwölfmonatsdurchschnitt auf vier Monate konzentrieren, sind sie beim Mittelfilter von BV4 auf alle Monate des Stützbereichs verteilt.

Die relativ ungünstigeren Übergänge bei den Elementen der Filter in der unmittelbaren Nähe des Randes sind im jeweils rechten Teil der Positionen von Schaubild 1a demonstriert. Der auffallend deutliche Rückgang der Elemente für die Monate mit den Nummern T-14 bis T-12 stellt offensichtlich ein notwendiges Gegengewicht dar zum starken Anstieg der Elemente (im Jahresabstand) bei den Monaten T-2 bis T, denn die „Neutralität“ der Schätzung der glatten Komponente gegenüber einer Saisonkomponente bleibt stets gewahrt.

In Schaubild 1b sind der Mittelfilter und zwei Folgefilter für die Schätzung der Saisonkomponente gezeichnet. Die Elemente, deren Summe stets gleich 0 ist, zeigen ein andersartiges, markantes Bild. Hohe positive Werte bestehen für den Monat der Schätzstelle τ und die beiden Monate $\tau \mp 12$ im Jahresabstand davon, während alle übrigen Werte negativ sind. In der Tatsache, daß das Element an der Stelle τ nahezu doppelt so groß ist wie an den Stellen $\tau \mp 12$, kommt die abgestufte Gewichtung der Abweichungsquadrate (siehe Abschnitt 1) zum Ausdruck: Die Werte im Jahresabstand tragen nur in abgeschwächtem Maß zu dem Ergebnis bei.

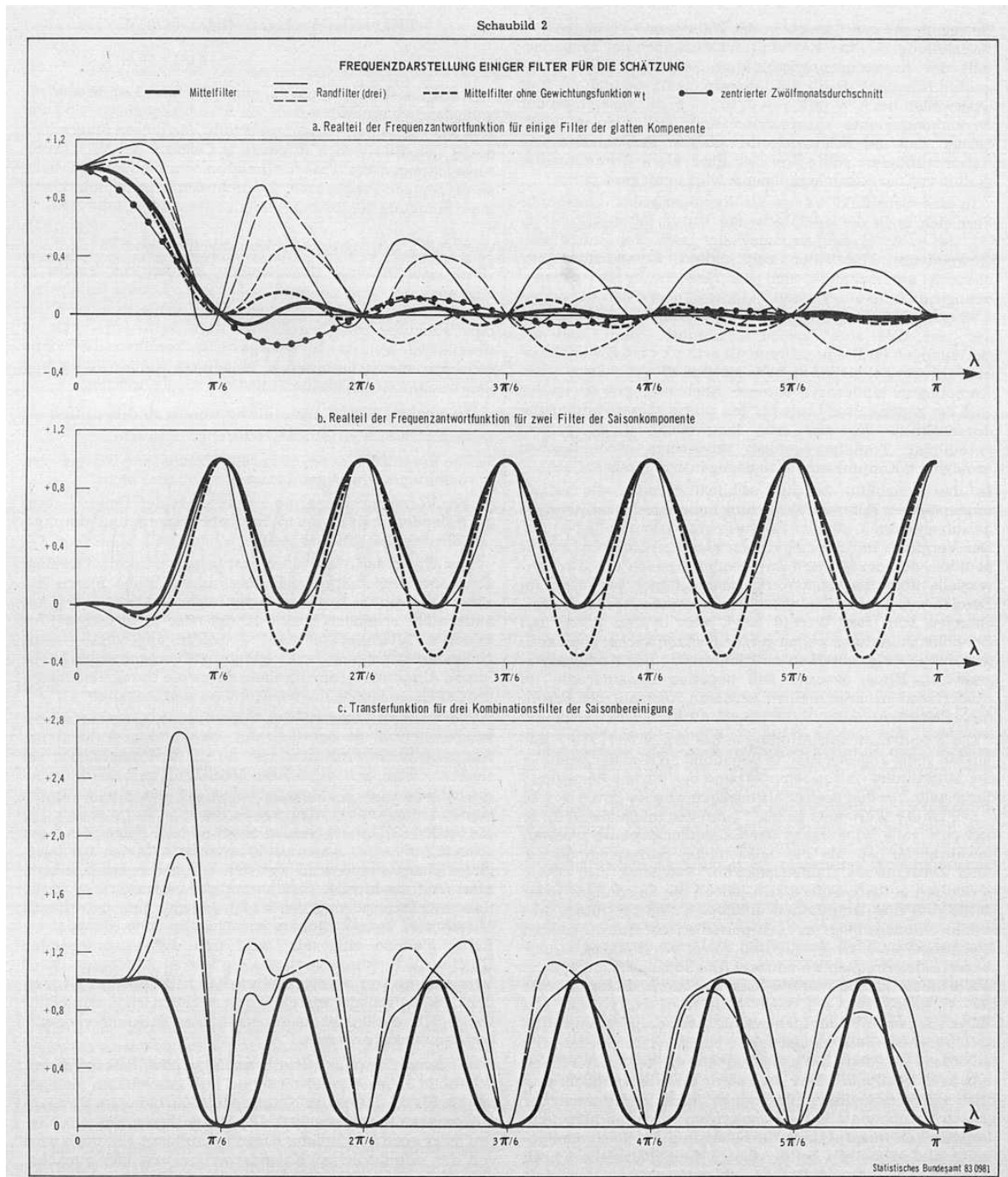
3 Die spektralen Eigenschaften der Filter

Eine Beurteilung der Wirkung und der Güte von Filtern zur Komponentenschätzung, wie sie in Abschnitt 2 dargestellt worden sind, ist in der Zeitdimension nicht vollwertig möglich. Deshalb werden die gewählten Filter nachstehend auch in der Frequenzdimension untersucht.

Grundlage für derartige Untersuchungen ist die Vorstellung, daß eine Zeitreihe aus zahlreichen verschiedenen Bewegungskomponenten zusammengesetzt ist. Mit Hilfe einer Fouriertransformation der Zeitreihe lassen sich die Intensitäten dieser einzelnen Bewegungen nach der Größe ihrer Schwingungsdauer oder – deren Kehrwert – ihrer Frequenz zerlegen und nach aufsteigenden Frequenzen anordnen. Aus einer solchen Spektraldarstellung kann abgelesen werden, wie stark z.B. die saisonalen Schwingungen bei der Jahresfrequenz und den Frequenzen ihrer Oberwellen oder die längeren Wellen im unteren Frequenzbereich in der vorliegenden Zeitreihe ausgeprägt sind. Die Auswirkungen der Multiplikation der Zeitreihenwerte mit den Elementen von Filtern lassen sich ebenfalls aus einer Spektraldarstellung der gefilterten Reihenwerte erkennen. Aber auch für die linearen, zeitvarianten Filter selbst können – ohne jegliche Bezugnahme auf eine konkrete Zeitreihe – „Spektrale“ berechnet werden. Die resultierenden Transferfunktionen geben für die verschiedenen Frequenzen an, mit welchem Faktor die spektrale Intensität der Ausgangswerte zu multiplizieren ist, um zur spektralen Intensität der gefilterten Reihe zu gelangen. Für Frequenzen, bei denen die Transferfunktion

- gleich 1 ist, hat der betreffende Filter keine Wirkung;
- größer als 1 ist, werden die Intensitäten durch den Filter verstärkt;
- kleiner als 1 ist, werden Intensitäten durch den Filter teilweise oder ganz absorbiert.

Schaubild 2



Eine Fouriertransformation führt (zunächst) zu sog. komplexen Ergebnissen, die einen Realteil R und einen Imaginärteil I umschließen. Bezeichnet man die einzelnen Elemente eines Filters $f(N, \tau; \tau_w, P_g)$ für die glatte Komponente g oder die Saisonkomponente s , der von der Länge N des Stützbereichs Θ , von der Schätzstelle τ im Stützbereich, von der Stelle τ_w der maximalen Gewichtung und dem Polynomgrad P_g abhängt und über verschiedene Teilfilter mit $p_i < 1$, $\sum_i p_i = 1$ zusammengesetzt sein kann, abkürzend mit $f_{t(\tau)}$, weil diese Elemente sich für die verschiedenen Positionen t im Stützbereich Θ unterscheiden (hier $t = 1, 2, 3, \dots, N$) und von der relativen Position τ der Schätzstelle abhängig sind, dann lauten die Transformationsformeln für die sogenannte Frequenzantwortfunktion

$$f_R(\lambda) = \sum_{t \in \Theta} f_{t(\tau)} \cdot \cos(t - \tau)\lambda \quad (\text{Realteil})$$

$$f_I(\lambda) = \sum_{t \in \Theta} f_{t(\tau)} \cdot \sin(t - \tau)\lambda \quad (\text{Imaginärteil})$$

Die Lösungen sind Funktionen der Frequenzvariablen λ und können für jeden beliebigen Wert dieser Variablen im Intervall $0 \leq \lambda \leq \pi$ berechnet werden. Der Imaginärteil ist für Phasenverschiebungen, d.h. einen Vor- oder Nachlauf der gefilterten Ergebnisse verantwortlich. Es ist leicht einzusehen, daß symmetrische Filter keine Phasenverschiebung bewirken, denn die Summe der Produkte aus den Elementen $f_i(\tau)$, die symmetrisch bezüglich der Schätzstelle τ sind, und der um τ schiefssymmetrischen Sinus-Funktion ist für alle λ -Werte gleich null.

In Schaubild 2a ist der Verlauf des Realteils des Mittelfilters zur **g l a t t e n K o m p o n e n t e** als fette Linie wiedergegeben. Diese Ergebnisse wie alle übrigen sind für die 61 äquidistanten λ -Werte $\lambda_i = i\pi/60$, $i = 0, 1, 2, \dots, 60$ berechnet worden. Die hier nicht dargestellten Lösungen für die dem Mittelfilter benachbarten Randfilter sehen sehr ähnlich aus; diese Tatsache lag gerade dem Konstruktionsprinzip der Folgefilter in der Nähe des Randes zugrunde (vgl. Abschnitt 2), um Filterfolgen mit ähnlichen (Spektral-)Eigenschaften zu gewinnen.

Im Niederfrequenzbereich der Mehrjahreswellen verharrt die Kurve lange auf dem Wert 1 (\rightarrow volle Wiedergabe der Bewegungen), bevor sie abfällt und für die Jahresfrequenz an der Stelle $\lambda = \pi/6$ den Wert 0 erreicht (\rightarrow volle Löschung der Bewegungen) und danach in der Nähe von 0 verbleibt. Die Feststellung, daß der Realteil (und stets auch der Imaginärteil) der Frequenzantwortfunktion bei allen Filtern zur glatten Komponente für die Jahresfrequenz $\pi/6$ und alle fünf Oberwellen bei $\lambda = j\pi/6$, $j = 2, 3, \dots, 6$, die zusammen der Saisonkomponente zuzurechnen sind, strikt gleich 0 ist, besagt, daß die Schätzung der glatten Komponente von Saisoneinflüssen völlig frei ist. Eine klare Trennung der glatten von der Saisonkomponente wird somit garantiert.

In das Schaubild 2a ist als kurz gerissene Linie zum Vergleich auch die entsprechende Kurve für den Filter $f_g(27, 14; -, 3; 1)$ eingezeichnet, der sich vom tatsächlich verwendeten Mittelfilter zur glatten Komponente nur dadurch unterscheidet, daß die Gewichtung der Abweichungsquadrate unterlassen ist (entsprechend den Versionen BV2 und BV3). Das frühere Sinken dieser Kurve gegenüber der von BV4 vom Niveau 1 hinunter zeigt, daß die Bewegungen im Frequenzintervall $\pi/12 < \lambda < \pi/6$ (Zweijahres- bis Jahreswelle) bisher nicht so stark in die geschätzte glatte Komponente einbezogen wurden. Noch ausgeprägter ist das Bild der gepunkteten Linie für den zentrierten Zwölfmonatsdurchschnitt, der nur sehr langwellige, d. h. nahezu geradlinige Trendbewegungen unverzerrt wiederzugeben ermöglicht, Konjunkturwellen dagegen unzulässig abflacht.

In das Schaubild 2a sind schließlich noch die relativ ungünstigsten Filter der Schätzung in unmittelbarer Randnähe aufgenommen, nämlich für die Schätzstellen $\tau = T, T-1, T-2$. Der Vergleich mit dem gleichsam idealen Mittelfilter macht deutlich, daß der Realteil im Frequenzbereich der Zweijahreswelle über den Best-Wert 1 ansteigt und vor allem im Bereich $\pi/6 < \lambda < 2\pi/6$ (Jahres- bis Halbjahreswelle) Abweichungen vom Best-Wert 0 zeigt, was in den genannten Bereichen allerdings wegen gerade dort schwacher Intensitäten einiger Originalreihen möglicherweise nicht schwerwiegend ist. Hinzu kommt, daß negative Imaginärteile im Niederfrequenzbereich einen gewissen Nachlauf der Ergebnisse bewirken.

Die Frequenzantwortfunktionen der **S a i s o n f i l t e r** zeigen ein völlig anderes Bild. In Schaubild 2b sind die Realteile des Mittelfilters (fett gezeichnet) und des letzten Randfilters dargestellt. An den sechs Saisonfrequenzen $\lambda = j\pi/6$, $j = 1, 2, \dots, 6$, ist der Wert stets genau 1 (und der Imaginärteil 0), so daß eine volle Wiedergabe der Saisonkomponente insoweit gewährleistet ist, als die tatsächliche Saisonkomponente einer Zeitreihe als „Linienspektrum“ erscheint, also streng periodisch auftritt. Sofern sich jedoch die Saisonfigur einer Reihe und ihre Amplituden allmählich ändern, findet dies seinen Niederschlag in fast-periodischen Schwingungen leicht abgewandelter Frequenzen, so daß ein „Bandspektrum“ an den Saisonfrequenzen entsteht. Die Saisonschätzung nach BV4 schließt ziemlich breite Saisonfrequenzbänder ein, wie der Vergleich mit der kurz gerissenen Kurve für den Mittelfilter von BV3 deutlich macht. Dies besagt, daß mit BV4 auch raschere Saisonänderungen zügig erfaßt werden und außerdem in den Zwischenfrequenzen keine störenden negativen Realteile mehr auftreten. Im Niederfrequenzbereich sind die Realteile minimal, so daß in den Bereich der glatten Komponente kaum eingegriffen wird. Die hier nicht dargestellten Imaginärteile zur Schätzung der Saisonkomponente sind sehr dicht bei 0, solange die Schätzstelle τ noch mehr als 12 Monate vom Reiheneende entfernt ist.

Entscheidend für die Beurteilung der Saisonbereinigung ist die kombinierte Wirkung aller Schätzvorgänge: Berechnung der glatten Komponente, Elimination dieser Ergebnisse aus den Originalwerten, Berechnung der Saisonkomponente aus den trend-konjunktur-bereinigten Originalwerten und Subtraktion dieser Saisondaten von den Originalwerten. Die entsprechende Verarbeitung der Real- und Imaginärteile der glatten und Saisonkomponente schließt mehrere Rechengänge ein. In Schaubild 2c werden die Ergebnisse für die Mittelschätzung und zwei Randschätzungen in der Form von Transferfunktionen $TF(\lambda)$ dargestellt, die als Quadratsummen (konjugiert komplexe Produkte) der betreffenden Frequenzantwortfunktionen definiert sind:

$$\begin{aligned} TF(\lambda) &= (f_R(\lambda) + if_I(\lambda)) \cdot (f_R(\lambda) - if_I(\lambda)) \\ &= f_R^2(\lambda) + f_I^2(\lambda) \end{aligned}$$

Die Mindestanforderungen, nämlich Werte 1 im Niederfrequenzbereich und Werte 0 an den Saisonfrequenzen, sind von BV4 wie von seinen Vorläufern erfüllt. Von diesen unterscheidet¹¹⁾ sich BV4 jedoch in bezug auf wesentlich schwächere Abweichungen der Transferfunktion von den Idealwerten 1 an denjenigen Frequenzen, die nicht den Saisonbändern und nicht eindeutig der glatten Komponente zuzurechnen sind.

4 Die Analyse der Monatsreihen mit BV4

Alle statistischen Monatsreihen können mit einem je Komponente einheitlichen Satz von Filtern analysiert werden. Eine Voruntersuchung hinsichtlich passender Filter oder günstiger Optionen für die Wahl freier Parameter ist überflüssig, weil das hier dargestellte Verfahren BV4 ohne freie Parameter unmittelbar eingesetzt werden kann und damit subjektive Beeinflussungen vermieden werden.

Die einzigen Steuerungen, die notwendig sind, betreffen

- die Art der Daten (Monats- oder Quartalswerte),
- die Vorentscheidung, ob gegebenenfalls eine Extremwertbereinigung durchgeführt werden soll oder nicht,
- die Vorentscheidung, ob – nach Art der Daten – eine Kalendereinigung in Betracht kommt und durchgeführt werden soll oder nicht.

Der Ablauf der Analyse beginnt (gegebenenfalls) mit einer Extremwelterkennung und -bereinigung (siehe hierzu Abschnitt 6), weil die Ergebnisse der folgenden Schätzvorgänge andernfalls erheblich beeinträchtigt oder verfälscht werden könnten. Anschließend wird – sofern angezeigt – eine Kalendereinigung der Originalwerte vorgenommen (siehe hierzu Abschnitt 5), um die allein durch die Unregelmäßigkeiten des Kalenders bedingten Einflüsse auszuschalten.

Die weiteren Schritte des Saisonbereinigungsverfahrens vollziehen sich in der üblichen Reihenfolge. Anhand der Ausgangswerte wird zunächst die glatte Komponente geschätzt, wobei der zugehörige Mittelfilter mit der Stützbereichslänge von 27 Monaten über die gesamte Zeitreihe gleitend angewendet wird; zur Schätzung für die ersten und die letzten 13 Monate werden abgewandelte Folgefilter (oder „Randfilter“, siehe Abschnitt 2) eingesetzt. Die so ermittelte glatte Komponente wird von den Ausgangswerten subtrahiert. Auf die hieraus resultierenden Monatswerte wird die Saisonschätzung angewendet¹²⁾, indem der betreffende Mittelfilter, der 47 Monate umschließt, über die gesamte Reihe gleitend eingesetzt wird und die entsprechenden Ergebnisse für die ersten und die letzten 23 Monate durch Verwendung von abgewandelten Randfiltern ermittelt werden. Die saisonbereinigten Werte ergeben sich schließlich durch Subtraktion der zuvor berechneten Saisonkomponente von den Ausgangswerten.

Als Beispiel für die Zeitreihenanalyse nach BV4 werden in Schaubild 3 die Originalwerte und ihre geschätzten Komponenten für die Reihe des Nettoproduktionsindex im Straßenfahrzeugbau wiedergegeben. Die Daten dieser Reihe schließen zwar auch beachtliche Kalendereinflüsse ein, doch wird von der erforderlichen Kalendereinigung hier zunächst abgesehen, um ihr Verfahren und ihre Auswirkung im nächsten Abschnitt gesondert darzustellen.

Die Originalwerte der Musterreihe zeigen ausgeprägte Bewegungen, die in starkem Maße auf regelmäßig wiederkehrende saisonale Schwingungen zurückzuführen sind. Die ermittelte Saisonkomponente läßt zwei Höhepunkte im Frühjahr und Herbst sowie zwei Tiefpunkte im Sommer und Winter erkennen; das markante Minimum im Sommer ist durch Betriebsferien zu erklären. Der mehrjährige Verlauf dieser Komponente ist zwar allmählichen Änderungen unterworfen, aber ihr Grundmuster bleibt ständig das gleiche.

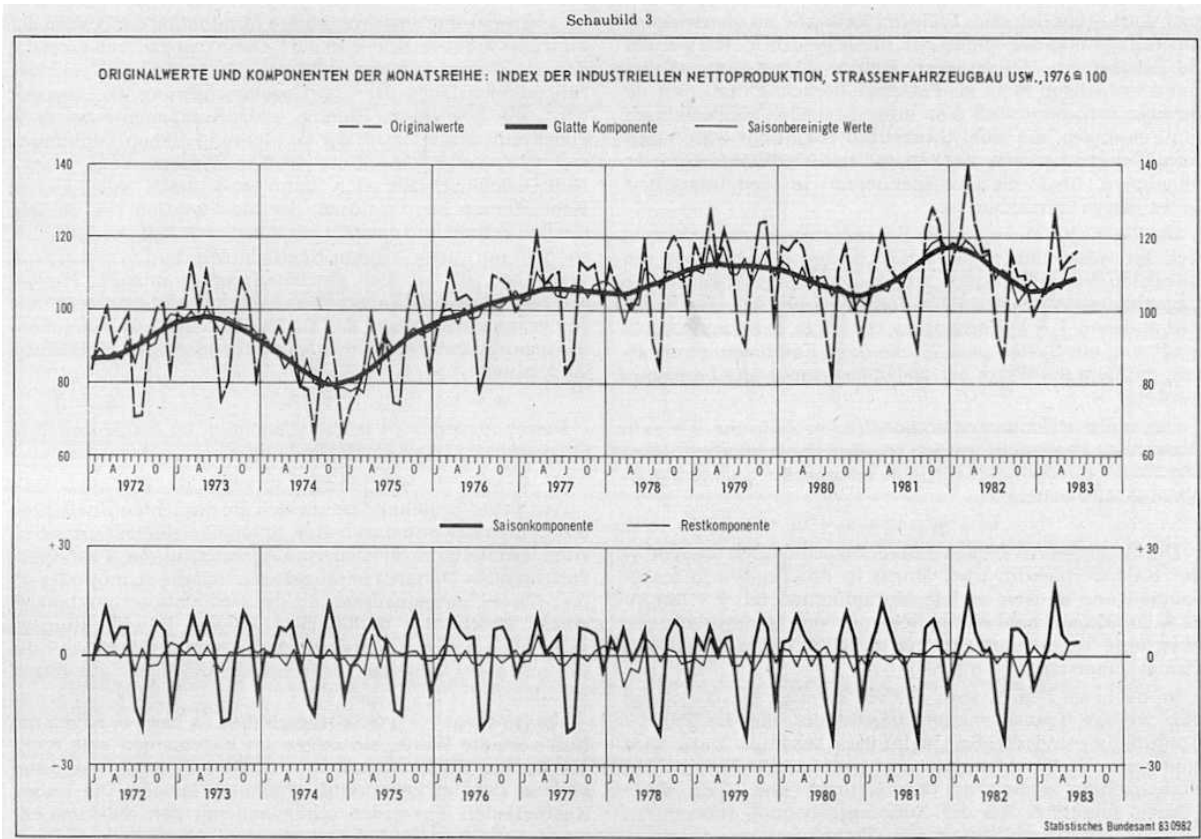
Auch die (stark gezeichnete) glatte Komponente hat einen merklichen Anteil an den Bewegungen der Originalwerte, mit konjunkturellen Hochpunkten im Mai 1973, im April/Mai 1979 sowie im Januar 1982; im Frühsommer 1977 zeichnet sich außerdem ein kleines Zwischenhoch ab, gefolgt von einer minimalen Abschwächung. Der Tiefpunkt nach der ersten Ölkrise war im November 1974 erreicht; weitere Minima sind im November/Dezember 1980 und im Dezember 1982/Januar 1983 zu beobachten.

Die saisonbereinigten Werte (im oberen Teil des Schaubilds) schwingen mit ziemlich geringen Ausschlägen um die Werte der glatten Komponente, geben also den längerfristigen Verlauf der Entwicklung recht gut wieder. Die Differenz zwischen diesen beiden Wertereihen ist die Restkomponente, die im unteren Teil des Schaubilds auch gesondert dargestellt ist.

¹¹⁾ Siehe Nullau, B. u. a.: „Das Berliner Verfahren; ein Beitrag zur Zeitreihenanalyse“, DIW-Beiträge zur Konjunkturforschung, Heft 7, 1969, S. 77/78; sowie Nourney, M.: „Weiterentwicklung des Verfahrens der Zeitreihenanalyse“, Wirtschaft und Statistik, 2/75, S. 98.

¹²⁾ Die Saisonkomponente ließe sich auch unmittelbar aus den Ausgangswerten schätzen, doch wären dann die Ergebnisse etwas weniger zuverlässig.

Schaubild 3



5 Die Berücksichtigung von Einflüssen des Kalenders

Die Verfahren der Analyse von Zeitreihen unterstellen gewöhnlich, daß die Zeitabschnitte bei den Beobachtungswerten (Monate, Quartale) äquidistant sind. Diese meist unausgesprochene Hypothese ist jedoch allenfalls für Jahresdaten näherungsweise erfüllt. Sie gilt mit Sicherheit nicht für Monatsdaten, da die Länge der Monate nach der Zahl ihrer Kalendertage um bis zu 10 % variiert. Dementsprechend müssen Unterschiede in den Monatsergebnissen erwartet werden, die lediglich auf die unterschiedliche Länge des Monats zurückzuführen sind¹³⁾. Beeinflussungen von Monatsergebnissen der Wirtschaftsstatistik gehen darauf zurück, daß die Zahl der monatlichen Arbeitstage variiert, und zwar bedingt durch die genannte unterschiedliche Zahl von Kalendertagen, aber auch durch „eingestreute“ Feiertage und durch den wechselnden Wochenrhythmus, der z. B. dazu führen kann, daß ein Monat mit 30 Tagen und fünf Wochenenden ebenso viele Arbeitstage hat wie ein Monat mit 28 Tagen. Und schließlich sind die einzelnen Arbeitstage im allgemeinen nicht gleichwertig, sondern sollten nach den Wochentagen unterschieden werden, vor allem wenn die wirtschaftlichen Aktivitäten ein markantes Wochentagsmuster besitzen.

Diese Feststellungen legen eine Bereinigung von Monatsdaten um die Einflüsse aus der Kalenderstruktur nahe, damit die Komponente der nicht erklärbaren Reste reduziert und die beiden systematischen Komponenten der längerfristigen Trend-Konjunktur und der jahreszeitlich bedingten Saison um so klarer herausgearbeitet werden können. Einschränkung ist jedoch zu bemerken, daß nicht alle Monatsreihen auch Kalendereinflüsse umschließen (Gegenbeispiele: Stichtagswerte über den Personalbestand, Preise). Für welche statistischen Monatsreihen eine Kalenderbereinigung faktisch in Betracht kommt, wird aus fachlicher Sicht entschieden.

Bevor das Berechnungsmodell zur „Kalenderkomponente“ im einzelnen dargelegt wird, sollen die Eigenheiten des Kalenders skizziert werden. Für jeden Monat werden in erster Linie die Häufigkeiten der sieben Wochentage herangezogen. Eine besondere Berücksichtigung verlangen die Feiertage, die den Charakter des Wochentags, an dessen Stelle sie treten, in bestimmter Weise verändern. Sie werden als achte Variable eingeführt und gelten als dominant gegenüber dem betreffenden Wochentag, wenn dieser nicht ein Sonntag ist. Die festen Feiertage umschließen den 1. Januar, 1. Mai, 17. Juni, 25. und 26. Dezember sowie den 24. und 31. Dezember je zur Hälfte. Als bewegliche Feiertage werden berücksichtigt der Karfreitag, Oster- und Pfingstmontag, der Himmelfahrtstag sowie der Buß- und Betttag. Auch Fronleichnam wird als voller Feiertag gewertet, obwohl dieser Tag nicht bundeseinheitlich gesetzlicher Feiertag ist, aber doch – ähnlich dem Buß- und Betttag – im überwiegenden Teil des Bundesgebietes gilt. Im Gegensatz hierzu werden die Festtage der Heiligen Drei Könige, Mariä Himmelfahrt und Allerheiligen nicht als Feiertage berücksichtigt, weil sie geringer verbreitet sind. Von einer

¹³⁾ Siehe Fußnote 2.

partiellen Einbeziehung von Feiertagen, die wohl theoretisch erwünscht wäre, wird bewußt abgesehen, um sachlich und zeitlich differenzierende Annahmen über die Gegebenheiten in den einzelnen Zeitreihen zu vermeiden.

Die Durchführung der Kalendereinigung vollzieht sich im wesentlichen in zwei Schritten. Zuerst wird ein Ausgleich der Monatslängen vorgenommen und danach durch Gegenüberstellung der Monatswerte und der je Monat vorhandenen 7 + 1 Wochentage mit Hilfe der Regressionsschätzung ein System von Wochentags-Koeffizienten ermittelt, mit dem die Werte der Kalenderkomponente berechnet werden.

Das additive Komponentenmodell einer Zeitreihe, wie es in Abschnitt 1 dargestellt worden ist, wird in einem allgemeineren Modell um die arbeitstägliche oder besser Kalenderkomponente a erweitert:

$$x_t = g_t + s_t + a_t + u_t$$

Die Monatswerte werden mittels Division durch die Zahl y_t der Kalendertage im t -ten Monat in Pro-Tag-Werte umgewandelt und danach mittels Multiplikation mit $\bar{y} = 365,25/12 = 30,4375$ auf eine einheitliche mittlere Monatslänge von etwa 30,44 Tagen gebracht. Die transformierten Daten sind formal äquidistant geworden¹⁴).

Im nächsten Schritt wird auf beide Seiten der Gleichung eine weitere Transformation angewendet, die die Trend-/Konjunktur- und die Saisoneinflüsse beseitigt. Dazu wird eine gleitende „Gesamtschätzung“ benutzt, die diese beiden Komponenten gemeinsam herausfiltert, und dann diese Lösung monatlich von der Ausgangsgleichung subtrahiert. Der als Matrix aufzufassende Transformations-Operator einer solchen Elimination von glatter und Saisonkomponente werde mit L bezeichnet, so daß sich die Beziehung

$$Lx_t = La_t + u_t$$

ergibt, da die Komponenten g und s eliminiert sind und die Restwerte u von der Elimination definitionsgemäß nicht betroffen werden.

Diese Gleichung unterscheidet sich von allen bisher praktizierten Analyseverfahren, die von der Hypothese einer Schätzbarkeit der Kalendereinflüsse allein aus der Restkomponente ausgehen, dadurch, daß auf der linken Seite der Gleichung zwar ebenfalls Restwerte Lx erscheinen, auf der rechten Seite aber nicht die Kalenderkomponente a selbst, sondern die gleichermaßen transformierte Komponente La steht. Das bedeutet, daß die gesuchten Lösungen für a gleichsam auf der Ebene der originalen x -Werte und nicht auf der Ebene der Restwerte Lx zu erwarten sind. Die Lösungswerte für a werden damit prinzipiell größer sein als bisher, weil auch solche der Kalenderkomponente zuzurechnenden Teile, die bisher von den Schätzwerten für die glatte und die Saisonkomponente zuvor absorbiert worden sind, nun zu Recht in die Komponente a einbezogen werden.

Für die Kalenderkomponente a selbst wird – wie bisher – ein strukturiertes Modell

$$a_t = y_{ti}^* b_{i(t)}, \quad t = 1, 2, \dots, T; \quad i = 1, 2, \dots, 8$$

unterstellt, das eine lineare Abhängigkeit von den Häufigkeiten y_{ti}^* der 7 + 1 Wochentage im jeweiligen Monat t bezeichnet. Dabei können die Einflüsse der 8 einzelnen Wochentage auf die beobachteten Ergebnisse der Zeitreihe und damit auf ihre Kalenderkomponente durchaus verschieden sein; derartige Unterschiede prägen die Größen der Wochentags-Koeffizienten $b_{i(t)}$, die geschätzt werden müssen und im Lauf der Zeit auch allmählichen Veränderungen unterliegen können.

Zuvor muss noch die Nebenbedingung berücksichtigt werden, daß die Kalenderkomponente im Erwartungswert null sein soll, daß also ihre spätere Elimination das Niveau der Zeitreihe nicht verschieben darf. Dies wird dadurch erreicht, daß in gleitenden Zeitabschnitten, deren Länge N mit der Stützbereichslänge der o. g. Gesamtschätzung gleichgesetzt wird, die gewichtete Summe aller Komponentenwerte a_t gleich null gesetzt wird; zur Gewichtung dienen Funktionen w_t , wie unter Abschnitt 1 beschrieben. Jede solche Null-Gleichung läßt sich dann z. B. nach dem letzten Koeffizienten $b_{8(t)}$ auflösen, der als Funktion der übrigen Größen auftritt und deshalb mit ihnen vereinigt werden kann, so daß nur noch sieben Koeffizienten $b_{j(t)}$, $j = 1, 2, \dots, 7$, verbleiben, die letztlich geschätzt werden müssen. Hierbei erfahren die Häufigkeiten y_{ti}^* der Wochentage entsprechende Änderungen zu y_{tj} , und das Modell, in dem die Nullsummenbedingung (und der Monatslängen-Ausgleich) berücksichtigt sind, lautet

$$a_t = y_{tj} \cdot b_{j(t)}, \quad t = 1, 2, \dots, T; \quad j = 1, 2, \dots, 7$$

Dieses Modell wird in die Gleichung der transformierten Originalwerte Lx eingesetzt und ergibt

¹⁴) Die beschriebene Transformation wird im folgenden aus Vereinfachungsgründen nicht durch ein Sondersymbol kenntlich gemacht.

$$Lx_t = (Ly_j)_t b_j + u_t$$

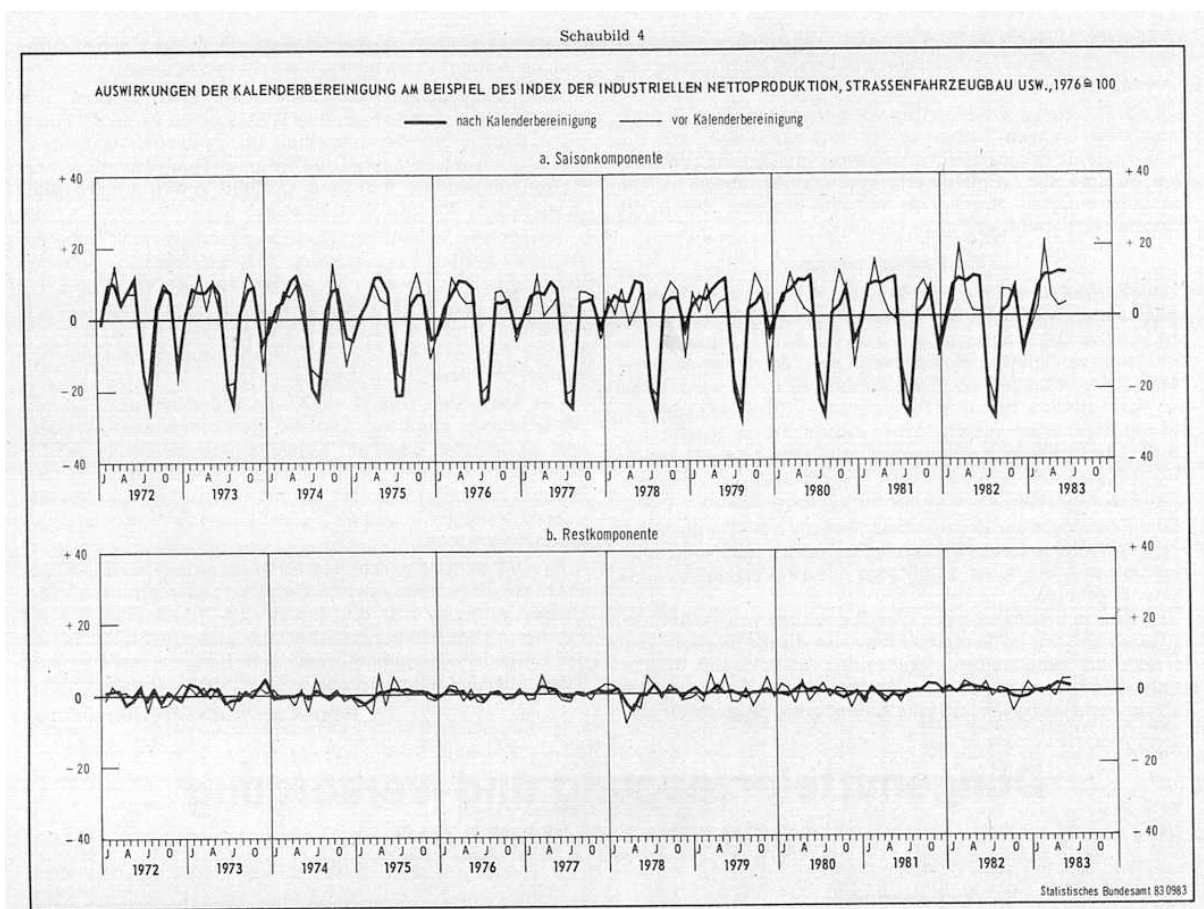
Aus dieser Beziehung lassen sich die gesuchten Koeffizienten b_j monatsweise durch eine „gleitende Regressionsschätzung“ ermitteln, wozu wiederum Ausschnitte von N aufeinanderfolgenden Monaten herangezogen und die zu minimierenden Abweichungsquadrate mit der Gewichtungsfunktion w_t gemäß Abschnitt 1 multipliziert werden. Die Funktion w_t sorgt auch dafür, daß bei der Schätzung in Randnähe die Lösungen der aktuellen Entwicklung möglichst gut folgen können.

Die je Monat $7 + 1$ Schätzergebnisse \hat{b}_j bzw. \hat{b}_{j+1} haben um Null verteilte Werte; sie zeigen im allgemeinen eine recht ausgeprägte Wochentagsstruktur der jeweiligen Zeitreihe und können sich im Zeitverlauf allmählich ändern. Die sieben Koeffizienten \hat{b}_j werden schließlich mit den nullsummentransformierten Häufigkeitswerten y_{tj} (nicht jedoch Ly_{tj} !) multipliziert und ergeben zusammengefaßt eine vorläufige Kalenderkomponente.

Diese Ergebnisse, die auf der Grundlage einer künstlich herbeigeführten Äquidistanz der Monatswerte berechnet worden sind, erhalten schließlich je Monat ihren notwendigen Rückbezug auf die ursprüngliche Monatsgliederung durch Hinzufügen der Differenz zwischen den Originaldaten in ihrer ursprünglichen und ihrer auf gleiche Monatslänge gebrachten Form.

In diesem Zusammenhang ist besonders zu beachten, daß die **Kalenderbereinigung** künftig als **erster Schritt** einer Zeitreihenanalyse eingesetzt wird¹⁵⁾, bevor die glatte und die Saisonkomponente geschätzt werden. Die beiden letztgenannten Schätzungen stützen sich demzufolge auf Ausgangsdaten, die nunmehr äquidistant sind und somit die allgemeine Voraussetzung für die Analyse erfüllen. Die inhaltliche Deutung der Analyseergebnisse weicht folglich auch von dem gewohnten Schema ab. Dies gilt insbesondere für die im Anschluß an die Kalenderbereinigung ermittelte Saisonkomponente, die keine Monatslängen-Unterschiede mehr widerspiegelt, sondern nur noch die jahreszeitlich spezifischen Einflüsse wiedergibt.

In Schaubild 4 sind die Auswirkungen der vorweg durchgeführten Kalenderbereinigung auf die Saison- und die Restkomponente für die bereits in Schaubild 3 untersuchte Indexreihe der Produktion im Straßenfahrzeugbau graphisch dargestellt. Neben den fett ausgezogenen Linien beider Komponenten nach Kalenderbereinigung sind als dünne Linien die entsprechenden Ergebnisse von Schaubild 3 wiederholt, um ihren Vergleich zu erleichtern.



¹⁵⁾ Falls erforderlich, wird allerdings vorweg eine Extremwertbereinigung durchgeführt (siehe Abschnitt 6).

Die Amplituden der neuen Saisonkomponente sind keineswegs geringer als die der bisherigen, so daß nicht etwa behauptet werden kann, die Kalenderbereinigung beeinträchtige die Saisonkomponente, indem sie ihr Elemente entziehe, die nicht den Kalendereinflüssen zuzurechnen sind. Der jahreszeitliche Verlauf der neuen Saisonwerte, in dem nach wie vor zwei Hoch- und zwei Tiefpunkte auftreten, ist viel gleichmäßiger und damit auch überzeugender geworden; so passen sich z. B. die Februar-Werte nun besser in die Anstiegsphase ein. Die Werte der neuen Restkomponente sind zwar erwartungsgemäß im quadratischen Mittel kleiner geworden, aber doch nicht so niedrig, daß man an einer zutreffenden Komponentenzersetzung zweifeln müßte.

6 Prüfung auf Extremwerte und deren Bereinigung

Die Komponentenschätzungen der Zeitreihenanalysen sind reagibel und anfällig gegenüber Extremwerten, also sporadisch auftretenden Beobachtungswerten, die von dem natürlichen Gefüge des Werteverlaufs ungewöhnlich stark abweichen. Dies gilt in besonderem Maß auch für das Berliner Verfahren. Um Beeinträchtigungen oder Verzerrungen bei der Schätzung der Komponenten, vor allem auch der Saisonkomponente, zu vermeiden, sollten echte Extremwerte vor Beginn aller weiteren Analysen aus den Zeitreihen eliminiert werden.

Die Extremwertbereinigung umschließt zwei Aufgaben. Zunächst müssen sämtliche monatlichen Beobachtungswerte daraufhin geprüft und jeweils entschieden werden, welche von ihnen tatsächlich durch extreme Ausschläge gestört sind. Danach muß festgestellt werden, um welche (Extremwert-)Beträge die Beobachtungswerte solcher Monate geeignet zu korrigieren sind.

Das seit 1972 angewendete Verfahren der Extremwertbereinigung¹⁶⁾ hat sich relativ gut bewährt. Es gründet sich auf die Annahme schwach stationärer, normal-verteilter stochastischer Prozesse und ermittelt gleitend auf der Basis von jeweils 24 vorausgehenden Monatswerten (die frei von Extremwerten sind) für den folgenden Monat einen hierdurch bedingten Erwartungswert, der sich inhaltlich aus den systematischen Komponenten der Reihe (ohne Restkomponente) zusammensetzt. Um diesen Erwartungswert wird ein Toleranzintervall gebildet, das dem Dreifachen der Standardabweichung der Restkomponente entspricht, und geprüft, ob der tatsächliche Beobachtungswert innerhalb dieses Intervalls liegt. Falls das nicht zutrifft, stellt der betreffende Originalwert einen Extremwert dar, der durch Rückführung des Originalwertes auf den Erwartungswert zuzüglich der (einfachen) Standardabweichung bereinigt wird.

Die Neugestaltung der arbeitstäglichen oder Kalenderbereinigung gemäß Abschnitt 5 bietet nun aber die Möglichkeit, mit dem betreffenden Schätzprozeß auch die Erkennung und Bereinigung von Extremwerten zu verbinden. Zu den sieben erklärenden Variablen der nullsummen-transformierten Wochentags-Häufigkeiten (y_{ij}) in der Regressionsbeziehung wird eine achte Variable als Vektor hinzugenommen, die für alle N Monate des Stützbereichs den Wert 0 hat, mit Ausnahme des einen Monats zur Schätzstelle τ , wo gleichsam als Detektor für einen Extremwert der Wert 1 steht. Nach Durchführung der gleitenden Regressionsrechnung gibt der für jeden Monat geschätzte Regressionskoeffizient zur achten Variablen unmittelbar zu erkennen, welcher Teil des Originalwertes weder der glatten, noch der Saison-, noch der Kalenderkomponente zuzurechnen, sondern strukturbereinigt dem Rest zugeordnet ist. Die genannten Koeffizienten können durch Bezugnahme auf ihre Standardabweichung darauf geprüft werden, ob sie – unabhängig von ihrem Vorzeichen – ungewöhnlich groß sind und deshalb einen Extremwert markieren. Der erforderliche Bereinigungswert läßt sich unmittelbar dem betreffenden Koeffizienten gleichsetzen. Vor der Bereinigung ist lediglich noch festzustellen, ob eins der beiden entsprechenden Prüfmaße im 12-Monats-Abstand nicht etwa – mit umgekehrtem Vorzeichen – noch größer ist und deswegen bei der Bereinigung den Vorrang erhalten muß.

Eine Extremwertbereinigung vor der analytischen Ermittlung aller übrigen Komponenten hat zur Folge, daß die Jahressumme der glatten Komponente von der entsprechenden Summe der Originalwerte systematisch abweicht. Die saisonbereinigten Ergebnisse werden dagegen durch die Extremwertbereinigung nicht verzerrt.

7 Zusammenfassung

Die Zeitreihenanalyse von Monatsdaten nach dem weiterentwickelten Verfahren BV4 ist einfach durchführbar, weil nur je ein einziger Satz universell verwendbarer Filter für die Schätzung der glatten Komponente bzw. der Saisonkomponente jeder statistischen Monatsreihe gebraucht wird. Nach wie vor entfallen iterative Rechengänge. Eine Bezugnahme auf die Eigenarten verschiedener Zeitreihen ist überflüssig, da die methodische Suche nach Optimierung bereits in die Filterkonstruktion eingeschlossen worden ist und deshalb nicht mehr für jede Zeitreihe wiederholt werden muß.

Die Ergebnisse der Zeitreihenanalyse nach BV4 sind wegen der Vermeidung freier Parameter eindeutig und voll übereinstimmend bei allen Benutzern dieses Verfahrens und seiner Ergebnisse.

Die Zusammenfassung der BV4-Ergebnisse von Teilreihen zu Gesamtreihen ist komponentenweise direkt möglich und korrekt ohne eine weitere gegenseitige Abstimmung, da das Analysemodell additiv und eindeutig ist.

¹⁶⁾ Siehe Fußnote 4.

Falls jedoch Extremwerte auftreten und eine Kalenderbereinigung durchgeführt wird, ist die Aggregationsfähigkeit der Komponenten von Teilreihen auf geeignete Weise herbeizuführen.

Die Analyseergebnisse nach BV4 sind flexibel und zuverlässig; sie sind gegenüber BV3 deutlich verbessert auch am aktuellen Ende der Zeitreihen. Durch den Schätzansatz ist methodisch gewährleistet, daß die glatte Komponente von der Saisonkomponente klar und treffend getrennt wird ohne gegenseitige störende Beeinflussung.

Revisionen vorläufiger Analyseergebnisse am Rand der Reihen werden unverändert für erforderlich gehalten angesichts der besseren Eigenschaften der Mittelfilter, die erst nach Hinzutreten mehrerer neuer Beobachtungswerte einsetzbar sind. Das Ausmaß derartiger Revisionen ist jedoch infolge der verbesserten vorläufigen Randschätzungen vermindert worden.

Das Verfahren BV4 versucht die Kalender- und Extremwerteeinflüsse nicht nur von der Restkomponente, sondern von sämtlichen anderen Komponenten zu isolieren. Die Kalenderbereinigung der Originalwerte wird künftig – dem Wunsch mehrerer Benutzer folgend – v o r den Schätzprozeduren für die glatte und die Saisonkomponente durchgeführt und nachgewiesen.

Es wird nicht ausgeschlossen, daß das nunmehr eingeführte Verfahren BV4 noch gewisse Verbesserungen erfahren kann. Hierzu gehören u. a. die analytische Fundierung für die Festlegung der benötigten Filterfolge und eine Stabilisierung der Kalenderkomponente. Nach dem Konzept von BV4 wird ferner ein Analyseprogramm für Quartalsreihen entwickelt.

Dipl.-Mathematiker Martin Nourney

Auszug aus Wirtschaft und Statistik

Herausgeber

Statistisches Bundesamt, Wiesbaden

www.destatis.de

Schriftleitung

Roderich Egeler, Präsident des Statistischen Bundesamtes

Brigitte Reimann (verantwortlich für den Inhalt)

Telefon: + 49 (0) 6 11 / 75 20 86

Ihr Kontakt zu uns

www.destatis.de/kontakt

Statistischer Informationsservice

Telefon: + 49 (0) 6 11 / 75 24 05

Telefax: + 49 (0) 6 11 / 75 33 30

Abkürzungen

| | | |
|----------|---|---|
| WiSta | = | Wirtschaft und Statistik |
| MD | = | Monatsdurchschnitt |
| VjD | = | Vierteljahresdurchschnitt |
| HjD | = | Halbjahresdurchschnitt |
| JD | = | Jahresdurchschnitt |
| D | = | Durchschnitt (bei nicht addierfähigen Größen) |
| Vj | = | Vierteljahr |
| Hj | = | Halbjahr |
| a. n. g. | = | anderweitig nicht genannt |
| o. a. S. | = | ohne ausgeprägten Schwerpunkt |
| St | = | Stück |
| Mill. | = | Million |
| Mrd. | = | Milliarde |

Zeichenerklärung

| | | |
|----------|---|--|
| p | = | vorläufige Zahl |
| r | = | berichtigte Zahl |
| s | = | geschätzte Zahl |
| – | = | nichts vorhanden |
| 0 | = | weniger als die Hälfte von 1 in der letzten besetzten Stelle, jedoch mehr als nichts |
| . | = | Zahlenwert unbekannt oder geheim zu halten |
| ... | = | Angabe fällt später an |
| X | = | Tabellenfach gesperrt, weil Aussage nicht sinnvoll |
| I oder — | = | grundsätzliche Änderung innerhalb einer Reihe, die den zeitlichen Vergleich beeinträchtigt |
| / | = | keine Angaben, da Zahlenwert nicht sicher genug |
| () | = | Aussagewert eingeschränkt, da der Zahlenwert statistisch relativ unsicher ist |

Abweichungen in den Summen ergeben sich durch Runden der Zahlen.