Zur Ausschaltung von Saisonschwankungen wirtschaftlicher Vorgänge

Von Dr. *Ch. Willigens*, Bern

In dieser Abhandlung soll eine Methode dargelegt werden, welche ich bei einer Untersuchung über Saisonschwankungen für das eidgenössische statistische Amt versucht habe. Es war ein Rechnungsverfahren nötig, welches es gestattete, die Ausschaltung abzustufen, ohne Erscheinungen, die nachweisbar keinen saisonmässigen Charakter hatten, zu verwischen. Das gegebene Verfahren war also eine Entwicklung, ähnlich der Fourierschen, in den Grenzen der Genauigkeit der vorliegenden Angaben.

Es sind je nach der Art der Erhebung zwei Arten graphischer Darstellungen zu betrachten:

1. Die gegebenen Zahlen sind Monatsbeträge, z. B. die Exportmengen einer Ware. Die betreffende Zahl ist das Mass für die Höhe eines Rechtecks, dessen Grundlinie die Monatslänge ist. Die exportierten Mengen werden durch die Flächeninhalte der Rechtecke dargestellt.

2. Die Zahlen sind das Ergebnis einer Stichtagerhebung, z. B. die Zahl der Arbeitslosen an einem bestimmten Monatstage. Die Zahlen sind die Masse der entsprechenden Ordinate; die erhaltenen Punkte werden durch eine gebrochene Linie aus Geradenstücken verbunden. Es wird angenommen, dass der Verlauf zwischen zwei Punkten angenähert durch die Gerade dargestellt wird. Der Flächeninhalt zwischen zwei Ordinaten entspricht der Summe der Arbeitslosigkeitsdauern der erfassten Personen während des entsprechenden Zeitraumes.

Im ersten Falle soll die Kurve durch die Ecken, links oben der Rechtecke gelegt werden, im zweiten Falle durch die Ecken der gebrochenen Linie.

Wir legen nun durch diese Punkte eine Kurve, deren Gleichung die Form besitzen soll.

$$1)\quad y = f(x) + (b_1 \cos x + a_1 \sin x) + (b_2 \cos 2x + a_2 \sin 2x) + (b_3 \cos 3x + a_3 \sin 3x) + (b_4 \cos 4x + a_4 \sin 4x) + (b_5 \cos 5x + a_5 \sin 5x) + b_6 \cos 6x.$$

Die Kurve $y = f(x)$ soll als Vergleichsbasis dienen und den Zustand ohne jährliche Schwankungen darstellen und spielt somit die Rolle, die jedenfalls dem Trend von Person zugedacht ist. Unter Saisonschwankungen sind wohl periodische Oszillationen zu verstehen, für welche die Jahresdauer eine Periode ist. In der Formel 1) sollen einstweilen die Ausdrücke in Klammern für ein bestimmtes Jahr gültig sein, der allgemeine Fall soll in der Folge noch besprochen werden.

Integrieren wir den Ausdruck 1) über ein Jahresintervall, so verschwinden, wegen der Periodizität, die trigonometrischen Ausdrücke, und wir erhalten die Bedingung

$$2)\qquad \int_{(J)} y\,dx = \int_{(J)} f(x)\,dx,$$

d. h. der Flächeninhalt des Trend soll für jedes Jahr gleich sein dem Flächeninhalt der ursprünglichen Kurve.

Diese Bedingung genügt nun keineswegs, um den Trend zu bestimmen. Ist eine Kurve bekannt, welche diese Relation erfüllt, so kann man aus dieser unendlich viele andere ableiten durch Addition einer Funktion $\varphi(x)$, die der Bedingung

$$\int_{(J)} \varphi(x)\,dx = 0 \qquad \text{genügt.}$$

Weitere Bedingungen zur näheren Bestimmung des Trend sind nicht ausgesprochen worden, es bleibt also die Wahl für jeden einzelnen Fall frei, unter Wahrung der Bedingungsgleichung 2).

Die Methode, die bisher zur Bestimmung des Trend zur Anwendung gelangt ist, besteht darin, dass man in gleichen Abständen die Jahresdurchschnitte als Ordinate aufträgt und mit Hilfe der Methode der kleinsten Quadrate eine Kurve, welche meistens eine Gerade ist, bestimmt, die sich den so enthaltenen Punkten anpassen soll. Damit wird aber der Zusammenhang des Trend mit der ursprünglichen Darstellung aufgehoben. Der Jahresdurchschnitt ist in Wirklichkeit der Flächeninhalt eines Rechtecks, mit dem Jahresabschnitt als Grundlinie, dieser Inhalt ist gleich dem der gegebenen Kurve für das betreffende Jahr. Man sollte also mit der Methode der kleinsten Quadrate die Gleichung der Integralkurve des Trend bestimmen und aus dieser durch Ableitung die Gleichung des Trend selbst erhalten. So würde eine Kurve bestimmt, die möglichst gut die Bedingung 2) zu erfüllen hätte.

Gegen den so bestimmten Trend lassen sich verschiedene Einwände erheben.

1. Der Verlauf der Kurve für ein Jahr kann durch die späteren Jahre in unzulässiger Weise beeinflusst werden. Wenn z. B. wegen Zollmassnahmen eines fremden Staates der Export eines Produktes einen starken Rückgang erleidet, so wird in den früheren Jahren schon ein Sinken der Kurve auftreten, das gar nicht der Wirklichkeit entspricht.

2. Die Bestimmung der Koeffizienten mit der Methode der kleinsten Quadrate führt zu einer Kurve, die über das betrachtete Intervall hinaus nicht extrapoliert werden darf.

Die Darstellung einer Funktion durch ein Polynom, dessen Koeffizienten nach und nach mit der Methode der kleinsten Quadrate berechnet werden, ist nämlich keineswegs identisch mit der Approximation durch die Taylorsche Reihe. Die Approximation mit der Taylorschen Reihe führt Kurven ein, welche in dem Punkt $x = 0$ mit der gegebenen Kurve eine Berührung immer steigender Ordnung hat. Der Konvergenzbereich der Reihe ist durch die Singularitäten der Funktion eindeutig bestimmt.

Die Methode der kleinsten Quadrate führt aber nicht zu einer solchen Entwicklung; denn hat man die Glieder bis zur Potenz x^n bestimmt und will man ein Glied mit x^{n+1} einführen, so werden dadurch alle bereits berechneten Koeffizienten verändert. Wir haben eine Reihenentwicklung nach Polynomen, z. B. den Tchebijcheffschen Polynomen, und die Approximation geht folgendermassen vor sich.

Alle Wurzeln der Polynome sind reell und befinden sich in dem Intervalle, auf welches die Methode der kleinsten Quadrate angewandt wird, folglich besitzt auch ein Polynom vom Grade n im ganzen $n-1$ Maxima und Minima, die sich auch alle an dem betreffenden Intervalle befinden. Wir haben eine Reihe von Funktionen, die immer zahlreichere und engere Wellen aufweisen und durch deren Überlagerung und Kompensation die gewünschte Annäherung erzielt wird. Dieses Annäherungsverfahren, welches übrigens dem durch Fouriersche Reihen ähnlich ist, kommt aber ausserhalb des Intervalles nicht mehr zur Anwendung, da dort die den Polynomen entsprechenden Kurven eine Variation in konstanter Richtung aufweisen. Aus diesem Grunde ist die Reihenentwicklung nach Polynomen nur innerhalb des Intervalles gültig.

Die Methode der kleinsten Quadrate bietet also bei weitem nicht die Vorzüge, die man sich wohl von ihrer Anwendung versprochen hat. Bei so unklaren Verhältnissen wie den vorliegenden ist das beste, die Berechnung möglichst einfach zu gestalten. Man wird also als Trend die Stufenkurve nehmen, welche durch die horizontalen Seiten der Rechtecke der Jahresdurchschnitte, die nicht auf der x-Achse liegen. Die Rechnung gestaltet sich dann wie folgt:

Setzt man

$$3)\quad y = b_0 + (b_1 \cos x + a_1 \sin x) + (b_2 \cos 2x + a_2 \sin 2x) + (b_3 \cos 3x + a_3 \sin 3x) + \\ + (b_4 \cos 4x + a_4 \sin 4x) + (b_5 \cos 5x + a_5 \sin 5x) + b_6 \cos 6x,$$

so enthält dieser Ausdruck unbestimmte Koeffizienten in genügender Anzahl, um zu erlauben, die Kurve durch die 12, den einzelnen Monaten entsprechenden Punkte zu legen. Wendet man zur Bestimmung der Koeffizienten die Methode der kleinsten Quadrate an und bezeichnet man mit y_0, y_1, y_2 — y_{11} die Monatswerte, so sind die Konstanten durch die Formeln gegeben

$$4)\quad \begin{cases} b_0 = \sum\limits_{i=0}^{11} \dfrac{y_i}{12} & b_6 = \sum\limits_{i=0}^{11} (-1)^i \dfrac{y_i}{12} \\ b_s = \sum\limits_{i=0}^{11} \dfrac{y_i \cos s x_i}{6} & a_s = \sum\limits_{i=0}^{11} \dfrac{y_i \sin s x_i}{6} \\ x_0 = 0 \quad x_2 = 30^0 \quad x_2 = 60^0 \;\ldots\ldots\; x_{11} = 330^0. \end{cases}$$

Diese Darstellung besitzt die bemerkenswerte Eigenschaft, dass, wenn nur eine beschränkte Anzahl von Klammern berücksichtigt wird, die reduzierte Formel, z. B.

$$b_0 = (b_1 \cos x + a_2 \sin x),$$

ebenfalls der Bestimmung nach der Methode der kleinsten Quadrate für diese Annäherung entspricht.

Die Perioden der Ausdrücke in Klammern sind, der Reihenfolge nach, 1 Jahr, $^1/_2$ Jahr, $^1/_3$ Jahr, $^1/_4$ Jahr, $^1/_5$ Jahr und $^1/_6$ Jahr. In den meisten Fällen wird es genügen, wenn man die von Saisonschwankungen bereinigten Zahlen nach der Formel

$$5)\qquad \bar{y} = b_0 + (b_5 \cos 5x + a_5 \sin 5x) + b_6 \cos 6x$$

berechnet. Zur Kontrolle müssen die Summen der unbereinigten und der bereinigten Zahlen gleich sein. Will man nach der Bereinigung eine Darstellung des allgemeinen Verlaufs erhalten, so berechnet man eine angenäherte Darstellung der Stufenkurve am besten mit Hilfe der Legendreschen Polynome, welche eine sehr einfache Berechnung gestatten. Es wird gut sein, sich zu diesem Zwecke Zahlentabellen anzulegen, z. B. eine solche der Integrale der ersten Potenzen von x zwischen zwei aufeinander folgenden Zahlen. Die Kurve gibt einen befriedigenden Verlauf, wenn sie die bereinigten Kurvenwellen in genügender Anzahl schneidet.

Die Werte der trigonometrischen Funktionen sind ein für alle Male bestimmt, die Rechnung gestaltet sich also denkbar einfach. Das Resultat ist, natürlich, bei der Unbestimmtheit der Begriffe nur eine Lösung unter unendlich vielen. Ich möchte im folgenden noch eine andere Methode darlegen.

Es seien im ganzen n Jahre zu untersuchen, man verfügt also über $12\,n$ Monatsangaben und will nun, statt eines einzelnen Jahres, die ganze Reihe durch die Anfangsglieder einer Fourierschen Entwicklung darstellen. Wir müssen, damit die entsprechende Kurve durch die gegebenen Punkte geht, einen Ausdruck mit $12\,n$ unbestimmten Koeffizienten benützen, also von der Form.

$$6)\quad y = b_0 + (b_1 \cos x + a_1 \sin x) + (b_2 \cos 2x + a_2 \sin 2x) + \ldots + (b_2 \cos s x + a_s \sin s x) \\ + \ldots + (b_{6n-1} \cos (6n-1)\,x + a_{6n-1} \sin (6n-1)\,x) + b_{6n} \cos 6nx.$$

Die Methode der kleinsten Quadrate gibt für die Koeffizienten die Werte:

$$7)\quad \left\{ \begin{array}{ll} b_0 = \sum\limits_{i,k} \dfrac{y_i^{(k)}}{12n} & b_{6n} = \sum\limits_{i,k} \dfrac{(-1)^i\, y_i^{(k)}}{12n} \\ b_s = \sum\limits_{i,k} \dfrac{y_i^{(k)} \cos s\, x_i^{(k)}}{6n} & a_s = \sum\limits_{i,k} \dfrac{y_i^{(k)} \sin s\, x_i^{(k)}}{6n} \end{array} \right.$$

Der obere Index K bedeutet die Ordnungszahl des Jahres und durchläuft die Werte von 1 bis n; der Index i entspricht dem Monat des betreffenden Jahres und durchläuft die Werte von 0 bis 11. Die Zahl $y_2^{(5)}$ bedeutet also z. B. den Wert für den Monat März des fünften Jahres. Die Differenz zwischen zwei aufeinanderfolgenden Werten von x beträgt $\frac{2\pi}{12\,n}$. Die Saisonschwankungen sollen als Periode eine Jahresdauer zulassen. Dies ist in der Formel 6) für die Ausdrücke

$(b_s \cos s\, x + a_s \sin s\, x)$ der Fall, in welchen s die Werte n, $2n$, $3n \ldots 6n$ annimmt. Die Wellen, die den anderen Ausdrücken für $s > n$ entsprechen, verschieben sich von einem Jahr zum anderen und weisen nicht die Stabilität auf, die man von den Saisonschwankungen jedenfalls erwarten muss.

Integriert man den Ausdruck 6) über ein ganzes Jahr, so sind die Integrale der Ausdrücke $(b_s \cos s\, x + a_s \sin s\, x)$ gleich Null, wenn s ein Vielfaches von n ist. Die Kurve, die man erhält, indem man diese besonderen Glieder in der Formel 6) unterdrückt, erfüllt also die für den Trend aufgestellte Bedingung 2). Die Berechnung der Koeffizienten, wenn s ein Vielfaches von n ist, gestaltet sich genau so, nach den Formeln 4), wie im zuerst untersuchten Falle, die Monatswerte werden durch die arithmetischen Mittel der Monatswerte für die einzelnen Jahre ersetzt, die Werte der trigonometrischen Funktionen bleiben genau dieselben. Aus den unbereinigten Zahlen erhält man dann durch Subtraktion den entsprechenden Punkt des Trend. Die Saisonschwankungen können ebenso gewählt werden wie bei der ersten Methode und haben für alle Jahre genau die gleichen Monatsbeträge.

Dieses Verfahren ist natürlich besonders schön, weil alle Koeffizienten nach der Methode der kleinsten Quadrate bestimmt sind.

Ich habe mich im vorhergehenden bemüht, die gestellte Aufgabe möglichst allgemein zu behandeln. Für die Schwankungen habe ich die in der Mathematik üblichen Definitionen beibehalten. Man kann sich des Eindrucks nicht erwehren, dass der ganze mathematische Aufwand für Konjunkturforschungen nur ein Deckmantel ist, um das Fehlen bestimmter Definitionen und die Unsicherheit der Begriffe zu verhüllen und den Untersuchungen einen trügerischen Schein wissenschaftlicher Strenge und Genauigkeit zu verleihen. Es ist in Wirklichkeit möglich, unendlich viele Lösungen der Aufgabe über die Saisonschwankungen zu finden, darunter auch jede gewünschte aufzustellen.

Eine besondere Besprechung der Personschen Methode erübrigt sich. Eine ausgezeichnete Besprechung und Kritik gibt die Abhandlung von *O. Anderson* «Zur Problematik der empirisch-statistischen Konjunkturforschung». (Veröffentlichung der Frankfurter Gesellschaft für Konjunkturforschung.) Ich möchte nur zu dieser folgende Bemerkungen machen. Auf Seite 28 ist zu lesen: «Was speziell Parabeln anbetrifft (zur Darstellung des Trend), so erscheint uns die Ansicht G. U. Yules wohl begründet, wonach eine ganze rationale Funktion $y = a + bt + ct^2 + dt^3 + \ldots$ in zwei Hinsichten für den Gebrauch unbequem sei; sie zwinge uns erstens eine Reihe von ganz willkürlichen Werten für die Parameter a, b, c, $d \ldots$ zu wählen, zweitens sei sie keine „natürliche Funktion, da sie gleichzeitig mit der Zunahme von t ins Unendliche wachse“.»

Diese Begründung scheint mir nicht ganz richtig, denn ich glaube, dass diese Parabelgleichung von den Begründern der Methode vielmehr als eine Annäherung durch die Anfangsglieder einer Taylorschen Reihenentwicklung aufgefasst worden ist. Mit wachsendem x werden die Konvergenzverhältnisse schlechter, und die Anzahl gewählter Glieder kann zu klein sein, dies besagt aber nichts gegen die Form der Darstellung an und für sich. Nimmt man z. B. die n ersten Glieder der Reihenentwicklung für e^{-x}, so wird dieses Polynom für unendlich grosse positive x unendlich gross, positiv oder negativ, die ganze Reihe konvergiert aber gleichwohl, und ihre Summe nähert sich beliebig dem Werte 0. Was ebenfalls die Willkürlichkeit der Werte für a, b, c — angeht, so könnte der gleiche Einwand z. B. für Reihenentwicklungen erhoben werden, deren Koeffizienten von Bernouillischen Zahlen abhängen, wenn diese Eigentümlichkeit nicht bekannt ist und nur die Zahlenwerte gegeben sind. So wäre z. B. $y = tg\, x$ keine «natürliche Funktion», wenn man die Entwicklung dieser Funktion vom Standpunkt Yules beurteilt.

Ich habe im vorhergehenden versucht, begreiflich zu machen, warum eine Extrapolation des Trend überhaupt unzulässig ist und wie die Begründung gerade in der Anwendung der Methode der kleinsten Quadrate zu suchen ist. Ich habe auch gezeigt, wie bei alleiniger Anwendung der Methode der kleinsten Quadrate eine Darstellung möglich ist, in welcher innerhalb eines jeden Jahres die gleichen Schwankungen unverändert wiederkehren. Da keine festgelegten Begriffe vorliegen, ist kein Grund vorhanden, einer Darstellungsart einen Vorzug vor anderen einzuräumen.

Damit soll natürlich keinswegs die Konjunkturforschung angefochten werden, sondern nur das mathematische Gewand, in welches man sie kleiden will.