



## Saisonbereinigung *vs.* Harmonische Analyse

Hier zeige ich, dass die Saisonbereinigung von Zeitreihen nichts anderes als eine spezielle Anwendung der harmonischen Analyse (Fourieranalyse) ist.

*Hinweis:* Komplexe Zahlen werden vorausgesetzt; mit Sinüssen und Cosinüssen ist alles umständlicher, geht aber genauso: Einfach  $e^{\pm ix} = \cos x \pm i \sin x$  ansetzen und Additionstheoreme für  $\sin(a) \sin(b)$ ,  $\cos(a) \cos(b)$ , die Beziehung  $\sin^2 + \cos^2 = 1$  sowie einiges mehr an Trigo-Beziehungen verwenden.

### Gegeben:

Zeitreihe mit einer periodischen Saisonkomponente, o.E.d.A. als Jahreszeitenabhängigkeit mit einer Periode von  $\tau = 12$  Monaten dargestellt:

$$y_i \equiv y_{tj}, \quad i = 12t + j, \quad t = 1, \dots, n, \quad j = 1, \dots, 12$$

Eine Abspaltung der glatten Komponente ist *nicht* nötig.

### Gesucht:

Abschätzung der Saisonkomponente  $S_j$ , so dass gilt

$$y_{12t+j} = R_{12t+j} + S_j,$$

mit  $R_{12t+j}$  dem saisonbereinigten nichtperiodischen Rest, der alle anderen Anteile enthält.

### Ansatz:

$S_j$  enthält einfach alle Fourierkomponenten mit Periode 12 oder ganzen Bruchteilen davon, d.h. mit Kreisfrequenzen  $\omega_k = 2\pi k/12$ ,  $k = 1, \dots, 11$ :

$$S_j = \sum_{k=1}^{11} F_k e^{2\pi i \left(\frac{kj}{12}\right)}, \quad (1)$$

mit  $i^2 = -1$ . *Hinweise:*

- Wegen der Bedeutung  $i$  als imaginäre Einheit haben wir den Index  $i$  der Vorlesung hier in  $t$  umgetauft
- Die Beiträge von  $F_k$  und  $F_{k+12j}$  sind für beliebige ganzzahlige  $j$  identisch.
- Es gibt nur 11, nicht 12 Beiträge, da die Komponente mit  $k = 0$  (oder  $k = 12$ ) ein "Gleichanteil" ist, den wir ja beim Saisonanteil nicht wollen.
- Man hätte  $k$  auch von 1 bis 6 sowie von -5 bis -1 gehen lassen. Dann wird das Sampling-Theorem von Shannon sichtbar (Fourierkomponenten bis zur halben Abtastfrequenz beschreiben die Zeitreihe vollständig), aber die Buchhaltung ist für unsere Zwecke unübersichtlicher.

Setzt man die (aus einer Mathe-Formelsammlung entnommenen) **Fourierkoeffizienten**

$$F_k = \frac{1}{12n} \sum_{l=1}^{12n} y_l e^{-2\pi i \left(\frac{kl}{12}\right)} \quad (2)$$

in (1) ein, erhält man

$$\begin{aligned} S_j &= \frac{1}{12n} \sum_{k=1}^{11} \sum_{l=1}^{12n} y_l e^{\frac{2\pi i k(j-l)}{12}} \\ &= \frac{1}{12n} \sum_{l=1}^{12n} y_l \sum_{k=1}^{11} e^{\frac{2\pi i k(j-l)}{12}} \end{aligned} \quad (3)$$

Wir betrachten nun  $\sum_{k=1}^{11} e^{\frac{2\pi i k(j-l)}{12}}$ . Stellt man sich die komplexen Zahlen als Zeiger in der komplexen Ebene anschaulich vor, wird sofort einsichtig, dass sich die Anteile der Summe wegheben, wenn nicht  $(j-l)$  ein ganzzahliges Vielfaches von 12 ist. Andernfalls stehen nämlich die Zeiger auf allen "Stunden" von 1 bis 12 und "1h" hebt sich mit "7h" weg, "2h" mit "8h" etc. Ist dagegen  $l = j, j+12, j+2*12$  etc, stehen die Zeiger immer auf "3h", was einer reellen "1" entspricht. Also

$$\sum_{k=1}^{11} e^{\frac{2\pi i k(j-l)}{12}} = \sum_{t=\dots, -1, 0, 1, 2, \dots} \delta_{l, j+12t}$$

mit der üblichen  $\delta$ -Funktion  $\delta_{jk} = 1$ , falls  $j = k$  und  $=0$  sonst.

Setzt man dies in (3) ein, sieht man, da  $l$  von 1 bis  $12n$  läuft, dass die Summe über  $t$  von 1 bis  $n$  beschränkt ist und man erhält

$$S_j = \frac{1}{12n} \sum_{l=1}^{12n} y_l \sum_{k=1}^{11} e^{\frac{2\pi i k(j-l)}{12}} = \frac{1}{n} \sum_{l=1}^{12n} \sum_{t=1}^n y_j \delta_{l, j+12t} - \frac{1}{12n} \sum_{l=1}^{12n} y_l = \frac{1}{n} \sum_{t=1}^n y_{j+12t} - \bar{y}. \quad (4)$$

und mit  $y_{j+12t} = y_{tj}$  schließlich

$$S_j = \frac{1}{n} \sum_{t=1}^n y_{tj} - \bar{y}. \quad (5)$$

Das Ergebnis (5) ist nichts anderes als der gewöhnliche Ausdruck für den Saisonanteil (einschließlich des "Korrekturglieds"). Man sieht hier, dass (im Ggs zu den Aussagen in vielen Büchern) die Ausgangsdaten nicht frei von Trends oder Konjunkturanteilen sein müssen! Das Verschwinden des Jahresmittels von  $S_j$  ist *per constructionem* gewährleistet. Die Periodizität ebenfalls.

Das Ergebnis ist allerdings *nicht* identisch zum in Abschnitt 19.4 vorgestellten Verfahren, sondern eliminiert automatisch (wie man zeigen kann) aus den Daten einen linearen Trend.

- Vorteil gegenüber der allgemeinen Saisonbestimmung: Es gibt saisonbereinigte Daten auch für den Anfang und das Ende der Zeitreihe (wegen der notwendigen gleitenden Mittel liegt der letzte Wert bei der allgemeinen Saisonbestimmung um eine halbe Periode (6 Monate bei den Arbeitslosenzahlen!) zurück).
- Nachteil: Nichtlinearitäten der glatten Komponente werden der Periodizität zugeordnet und verfälschen das Ergebnis.