

Aufgabenblatt 4 – Lösungen

Aufgabe 1 (Glättungsverfahren)

- a) Sie haben in der Vorlesung mit der gleitenden, der polynomialen und der exponentiellen Glättung mehrere Verfahren kennengelernt, um alle Komponenten außer dem Trend aus einer Zeitreihe zu entfernen. Bitte beschreiben Sie Unterschiede und Gemeinsamkeiten.

- Bei der gleitenden Glättung wird die Beobachtung y_t durch die gewichtete Summe $\sum_{j=-p}^f g_j y_{t+j}$ ersetzt. Die ersten p und die letzten f Beobachtungen gehen verloren. Oft verwendet man konstante Gewichte $g_j = (p + f + 1)^{-1}$.
- Die polynomiale Glättung betrachtet k Beobachtungen links und rechts von y_t und passt daran ein Polynom des Grades p mittels Regression an. Dann wird y_t durch den prognostizierten Wert des Polynoms ersetzt. Polynomiale Glättung ist eine gleitende Glättung mit nicht-konstanten Gewichten.
- Die exponentielle Glättung ist eine polynomiale Glättung, die alle Beobachtungen einbezieht. Der Approximationsfehler wird geometrisch abklingend mit der Distanz τ durch $(1 - \alpha)^\tau$ gewichtet. Wenn das Polynom den Grad $p = 0$ hat, nennen wir das Verfahren ‘einfach’, für $p = 1$ ‘zweifach’.

- b) Die Glättungsverfahren haben Parameter, die von dem Anwender gewählt werden müssen. Welchen Einfluss hat die Wahl auf das Ergebnis der Glättung?

- Bei der gleitenden Glättung bestimmen p und f , wie weit in die Vergangenheit und Zukunft geschaut wird. Je weiter, desto kleiner ist die Prognosevarianz. Je kürzer, desto besser ist die Anpassung an einen Trend. Bei konstanten Gewichten und $p = f$ bleibt ein linearer Trend erhalten. Ist $p > f$, so hinkt der geglättete Wert dem tatsächlichen hinterher (und umgekehrt), siehe c).
- Bei der polynomialen Glättung muss zusätzlich der Polynomgrad p gewählt werden. Je größer p , desto besser die Anpassung, aber auch desto höher die Prognosevarianz.
- Bei der exponentiellen Glättung muss zusätzlich der Gewichtungssparameter α bestimmt werden. Je kleiner α , desto größer ist der Einfluss vergangener Beobachtungen.

- c) Sie finden im Lernraum der PÜ den Datensatz ‘Geburten in Deutschland’ mit monatlichen Geburtenzahlen in Deutschland. Bitte glätten Sie diese Zeitreihe mittels backward smoothing, forward smoothing und moving averages.

Siehe R Code.

Aufgabe 2 (Exponentielle Glättung)

- a) Sie haben in der Vorlesung die rekursive Formel

$$a_T = \alpha y_T + (1 - \alpha)a_{T-1}$$

für die Berechnung der einfachen exponentiellen Glättung kennengelernt. Bitte glätten Sie damit die Zeitreihe aus Aufgabe 1c). Welchen Einfluss hat die Wahl für α und a_1 ?

Siehe R Code.

- b) Führen Sie ebenfalls die zweifache exponentielle Glättung durch, wofür Sie die folgenden Rekursionsformeln kennengelernt haben:

$$N_T := a_T + b_T T$$

$$N_T = \alpha y_T + (1 - \alpha)(N_{T-1} + b_{T-1})$$

$$b_T = (1 - \beta)b_{T-1} + \beta(N_T - N_{T-1})$$

Probieren Sie wieder verschiedene Werte für a_1 (beziehungsweise N_1), b_1 , α und β aus.

Siehe R Code.

- c) Die R Funktion `HoltWinters()` hilft bei der Bestimmung der Parameter. Nutzen Sie die dort vorgeschlagenen Werte und prognostizieren Sie die Geburtenzahl für Juni 2023, basierend auf der einfachen sowie der zweifachen exponentiellen Glättung.

Siehe R Code. Einfache prognostiziert 52861, zweifache 50641 Kinder.

Aufgabe 3 (Saisonmodellierung mittels Regression)

Wir betrachten erneut die Zeitreihe aus Aufgabe 1c), aber hier nur ab 2010.

- a) Wie interpretieren Sie die vorliegende Saison? Welche Werte haben die Beobachtungen mit den Indizes 2022.2 und 2022.Q2 basierend auf der Notation in der Vorlesung?

Im Juli gibt es mehr Geburten als im Februar, also werden im Winter mehr Kinder gezeugt als im Sommer. Die Beobachtung 2022.2 ist die Geburtenzahl im Februar 2022 und beträgt 53727. Die Beobachtung 2022.Q2 sind die Geburten im zweiten Quartal 2022, also die Summe von 2022.4, 2022.5, 2022.6; sie beträgt 187198.

b) Bitte passen Sie an die Geburtenzeitreihe das lineare Modell

$$\text{Geburten}_{J,i} = b_0 + (12[J - 2010] + i)b_1 + (12[J - 2010] + i)^2 b_2 + \sum_{j=2}^{12} D_{J,i,j} \cdot S_j + u_{J,i}$$

an. Die Notationen sind analog zur Vorlesung.

Siehe R Code.

c) Wiederholen Sie die Anpassung aus b), aber nun mit allen 12 Monatsdummies, dafür ohne die Konstante. Können Sie die Ergebnisse vergleichen?

Siehe R Code. Beide Modelle sind äquivalent.