11. Zeitreihen mit Trend und Saisonalität

In diesem Abschnitt geht es um ZR, die in eine Trend-, eine Saison- und eine Restkomponente zerlegt werden können. (Das Niveau sei in der Trendkomponente enthalten.)

Beispiele für solche ZR sind in Abb. 23 und Abb. 24 zu sehen. Es ist eine deutliche Saisonalität in den Daten erkennbar. Die bisherigen Verfahren zur ZR-Analyse sind nicht geeignet so ein saisonales Verhalten zu modellieren. Daher werden in diesem Abschnitt neue Verfahren hergeleitet, die sowohl Trend als auch Saisonalität berücksichtigen.

11. Zeitreihen mit Trend und Saisonalität

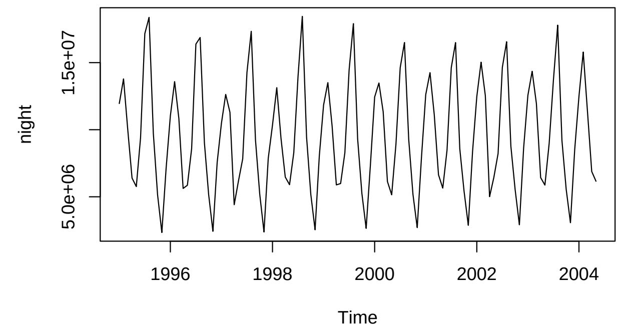


Abbildung 23: Zeitreihenplot der Nächtigungs-Daten.

11. Zeitreihen mit Trend und Saisonalität

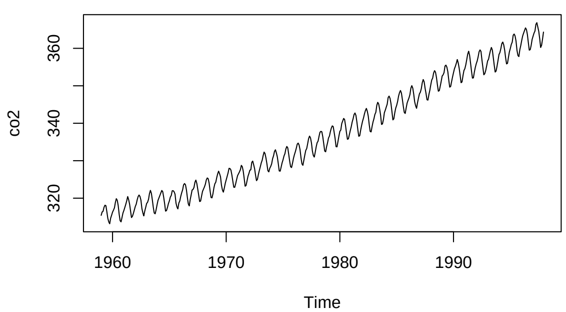


Abbildung 24: Zeitreihenplot der $C0_2$-Daten.

11. Zeitreihen mit Trend und Saisonalität

In Abschnitt 11.1. und 11.2. werden wir Methoden kennen lernen, die eine ZR gemäß dem additiven klassischen Komponentenmodell in eine Trend-, eine Saison- und eine Restkomponente zerlegen.

In Abschnitt 11.3. wird das Holt-Winters Verfahren zur Prognose von ZR, die Trend und Saisonalität aufweisen, behandelt. In diesem iterativen Algorithmus wird Niveau, Trend und Saisonalität rekursiv berechnet. Die Saisonkomponenten werden hier nicht über die Jahre konstant gehalten (wie es einer Annahme im klassischen Komponentenmodell entspricht), sondern auch sie verändern sich über die Jahre.

11. Zeitreihen mit Trend und Saisonalität

Bei Daten, die einen nichtlinearen Trend und nicht konstante Saisonalität aufweisen, kann eine Transformation sinnvoll sein.

In Abb. 25 ist so ein Beispiel zu sehen (in R unter `AirPassengers` verfügbar). Der Trend und die saisonalen Schwankungen nehmen mit der Zeit zu. Ein Modell, das einen linearen Trend und eine über die Jahre konstante Saisonalität annimmt, wäre für diese Daten nicht passend.

Die Daten können über den Logarithmus transformiert werden. In Abb. 26 sehen wir den Zeitreihenplot der logarithmierten Daten. Der Trend ist hier annähernd linear und die saisonalen Abweichungen bleiben über die Jahre annähernd gleich.

11. Zeitreihen mit Trend und Saisonalität

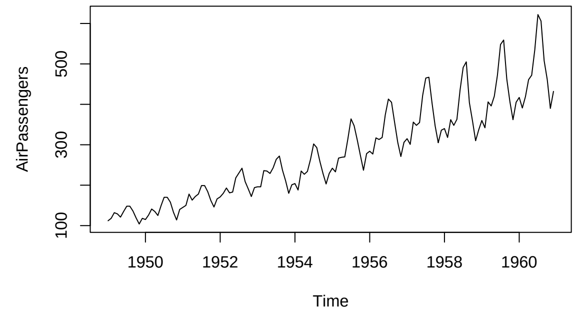


Abbildung 25: Zeitreihenplot der Flugpassagier-Daten.

11. Zeitreihen mit Trend und Saisonalität

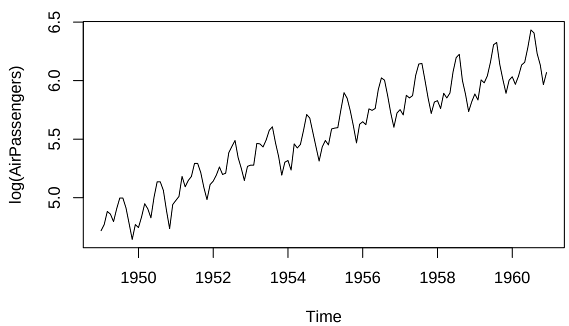


Abbildung 26: Logarithmierten Flugpassagier-Daten.

11.1. Methode der kleinen Trends

Diese Methode zerlegt eine ZR in eine Trend-, eine Saison- und eine Restkomponente.

Es sei eine ZR $x_{j,k}$ gegeben. Der erste Index steht für das Jahr $j = 1, \ldots, J$ und der zweite Index steht für die Saison $k = 1, \ldots, p$. p ist daher die Periodenlänge z.B. 12 bei Monatsdaten, 4 bei Quartalsdaten.

$$x_{j,k} = \hat{m}_j + \hat{s}_k + \varepsilon_{j,k}$$

$\hat{m}_j$ ist der für das Jahr j geschätzte Trend.

$\hat{s}_k$ ist die saisonale Komponente für Saison k.

$\varepsilon_{j,k}$ ist die Abweichung zwischen dem wahren Wert der ZR und der Schätzung: $\varepsilon_{j,k} = x_{j,k} - (\hat{m}_j + \hat{s}_k)$.

11.1. Methode der kleinen Trends

Schätzung der Trendkomponente für jedes Jahr $j = 1, \ldots, J$:

$$\hat{m}_j = \frac{1}{p} \sum_{k=1}^{p} x_{j,k}$$

Jeweils für ein ganzes Jahr wird ein konstanter Wert angenommen.

Schätzung der Saisonkomponente für jede Saison $k = 1, \ldots, p$ einzeln aus den um den Trend bereinigten Daten (z.B. aus allen bereinigten Jännerwerten, dann alle Februarwerte,..):

$$\hat{s}_k = \frac{1}{J} \sum_{j=1}^{J} (x_{j,k} - \hat{m}_j)$$

11.1. Methode der kleinen Trends

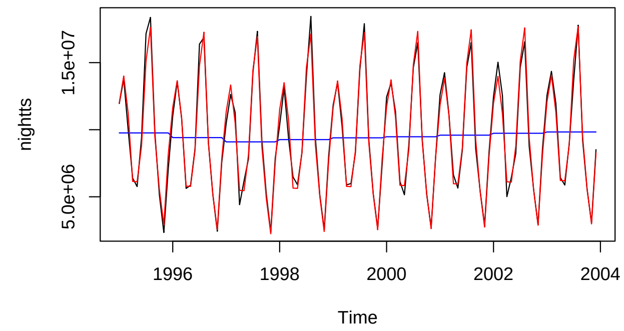


Abbildung 27: Methode d. kl. Trends für Nächtigungs-Daten.

11.1. Methode der kleinen Trends

Die Methode der kleinen Trends für die Nächtigungs-Daten:

Der ZR-Plot der Nächtigungs-Daten Abb. 23 zeigt ein nahezu konstantes Niveau über jeweils ein Jahr und eine deutliche Saisonalität, die sich aber über die Jahre hin kaum ändert. Eine Zerlegung gemäß dem klassischem Komponentenmodell kann daher mit der Methode der kleinen Trends erfolgen.

In Abb. 27 sieht man die Niveauschätzungen (blaue Linie) und die Schätzung der ZR (rot), die sehr nahe an den wahren Werten liegt.

11.1. Methode der kleinen Trends

Zur Anwendung der Methode der kleinen Trends

- Da für ein ganzes Jahr die Trendkomponente $\hat{m}_j$ als konstant angenommen wird, ist diese Zerlegung nur für Daten sinnvoll, die kaum einen jährlichen Trend aufweisen.

- Da die Saisonkomponenten $\hat{s}_k$ für die einzelnen Saisonen über die Jahre immer gleich bleiben, bietet das Modell eine geringere Flexibilität als andere Verfahren (z.B. Holt-Winters).

11.2. Saison-Trend Zerlegung mittels Moving Average

Auch dies ist eine Methode zur Trend- und Saisonbereinigung. Die ZR x_t, $t = 1, \ldots, T$ wird gemäß dem klassischen Komponentenmodell in den Trend $\hat{m}_t$, die Saison $\hat{s}_t$ und die Restkomponente ε_t zerlegt:

$$x_t = \hat{m}_t + \hat{s}_t + \varepsilon_t$$

11.2. Saison-Trend Zerlegung mittels Moving Average

Die Trendkomponente $\hat{m}_t$ wird mit Hilfe des gleitenden Durchschnitts (moving average (MA)) für $t = q+1, \ldots, T-q$ berechnet: für ungerade Periode $p = 2q + 1$:

$$m_t = \frac{1}{p}(x_{t-q} + x_{t-q+1} + \cdots + x_{t+q-1} + x_{t+q})$$

für gerade Periode $p = 2q$:

$$m_t = \frac{1}{p}(0.5x_{t-q} + x_{t-q+1} + \cdots + x_{t+q-1} + 0.5x_{t+q})$$

11.2. Saison-Trend Zerlegung mittels Moving Average

Wie bei der Methode der kleinen Trends werden die saisonalen Komponenten $\hat{s}_k$ für jede Saison k einzeln aus den um den Trend bereinigten Daten geschätzt (z.B. aus allen bereinigten Jännerwerten, dann alle Februarwerte,..). Zeitpunkte:

$$\hat{s}_k = \frac{1}{J-2}\sum_{j=2}^{J-1}(x_{k+(j-1)p} - \hat{m}_{k+(j-1)p})$$

Anmerkungen: Die Bedingung $\sum_{k=1}^{p} \hat{s}_k = 0$ muss nicht notwendigerweise erfüllt sein, in diesem Fall werden die $\hat{s}_k$ entsprechend umgerechnet und die $\hat{m}_t$ nochmals geschätzt.
Da beim MA in $\hat{m}_t$ im ersten und im letzten Jahr Zeitpunkte wegfallen, werden diese beiden Jahre hier in der Summe weggelassen.

11.2. Saison-Trend Zerlegung mittels Moving Average

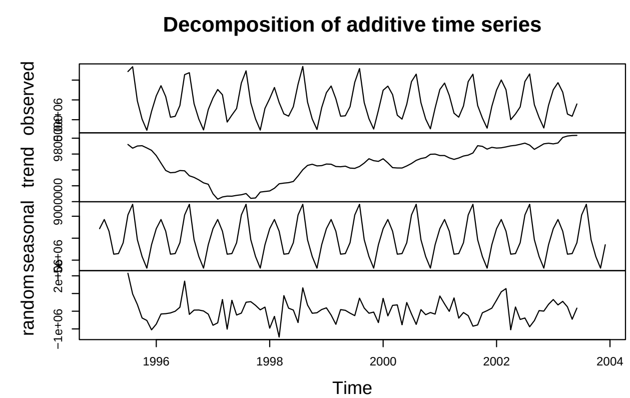


Abbildung 28: Saison-Trend Zerl. Nächtigungs-Daten mit MA.

11.2. Saison-Trend Zerlegung mittels Moving Average

Die Saison-Trend Zerlegung der Nächtigungs-Daten mittels MA:

In Abb. 28 sieht man der Zerlegungsformel $x_t = \hat{m}_t + \hat{s}_t + \varepsilon_t$ entsprechend zeilenweise die ZR x_t, $\hat{m}_t$, $\hat{s}_t$ und ε_t.

11.3. Holt-Winters Verfahren

Das Holt Verfahren wird nun so erweitert, dass auch ZR, die saisonale Schwankungen enthalten, prognostiziert werden könnnen.

Es gibt 2 Varianten:

eine für eine additive und eine für eine multiplikative saisonale Komponente. Die Periode der saisonalen Komponente $\hat{s}_t$ habe die Länge p.

Die h-Schritt Prognosen für eine Periode zum Zeitpunkt T lauten $(h = 1, 2, \ldots, p)$:

$$\textbf{Additiv}: \quad \hat{x}_T(h) = \hat{a}_{T+1} + \hat{b}_{T+1} \cdot h + \hat{s}_{T+h-p}$$

$$\textbf{Multiplikativ}: \quad \hat{x}_T(h) = (\hat{a}_{T+1} + \hat{b}_{T+1} \cdot h) \cdot \hat{s}_{T+h-p}$$

11.3. Holt-Winters Verfahren

Additives versus Multiplikatives Modell

Das Holt Winters Verfahren kann Saisonalität auf 2 verschiedene Arten modellierten:

- Additive Saisonalität: In Abb. 23 ist die Amplitude des saisonalen Verhaltens in etwa gleich bleibend. Daher wird für diese Daten die saisonale Komponente in additiver Form im Holt-Winters Verfahren berücksichtigt.

- Multiplikative Saisonalität In Abb. 29 ist eine mit der Zeit wachsende Amplitude erkennbar. Daher ist für diese Daten das Holt-Winters Verfahren mit multiplikativer Saisonalität passend.

11.3. Holt-Winters Verfahren

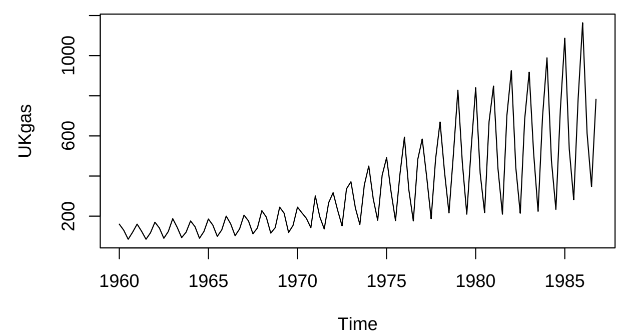


Abbildung 29: Die UKgas-Daten.

11.3. Holt-Winters Verfahren

Für die additive Variante lautet die Rekursion für $t = p + 1, \ldots, T$:

$$\hat{a}_t = \alpha(x_{t-1} - \hat{s}_{t-1-p}) + (1-\alpha)(\hat{a}_{t-1} + \hat{b}_{t-1})$$
$$\hat{b}_t = \beta(\hat{a}_t - \hat{a}_{t-1}) + (1-\beta)\hat{b}_{t-1}$$
$$\hat{s}_{t-p} = \gamma(x_{t-p} - \hat{a}_{t-p+1}) + (1-\gamma)\hat{s}_{t-2p}$$

Die Einschritt-Vorhersagen zum Zeitpunkt $t-1$ ergeben sich daher aus Summe der Schätzwerte von Trend, Niveau und Saison:

$$\hat{x}_{t-1}(1) = \hat{a}_t + \hat{b}_t + \hat{s}_{t-p}.$$

11.3. Holt-Winters Verfahren

Für die multiplikative Variante lautet die Rekursion für $t = p + 1, \ldots, T$:

$$\hat{a}_t = \alpha(x_{t-1}/\hat{s}_{t-1-p}) + (1-\alpha)(\hat{a}_{t-1} + \hat{b}_{t-1})$$
$$\hat{b}_t = \beta(\hat{a}_t - \hat{a}_{t-1}) + (1-\beta)\hat{b}_{t-1}$$
$$\hat{s}_{t-p} = \gamma(x_{t-p}/\hat{a}_{t-p+1}) + (1-\gamma)\hat{s}_{t-2p}$$

Die Einschritt-Vorhersagen zum Zeitpunkt $t-1$ ergeben sich daher aus den Schätzwerte von Trend, Niveau und Saison:

$$\hat{x}_{t-1}(1) = (\hat{a}_t + \hat{b}_t) \cdot \hat{s}_{t-p}.$$

11.3. Holt-Winters Verfahren

Man benötigt Startwerte für die Niveau- und die Trendkomponente für den Zeitpunkt $p+1$, und für die saisonalen Komponenten für die Zeitpunkte $t = 1, \ldots, p$. Es gibt verschiedene Möglichkeiten solche Startwerte festzusetzen. In R können Startwerte explizit vorgegeben werden. Ansonsten wird dafür eine Saison-Trend Zerlegung mittels moving average verwendet.

11.3. Holt-Winters Verfahren

Die Prognosen beim additiven Holt-Winters Verfahren über den Zeitraum der ZR hinaus lauten:

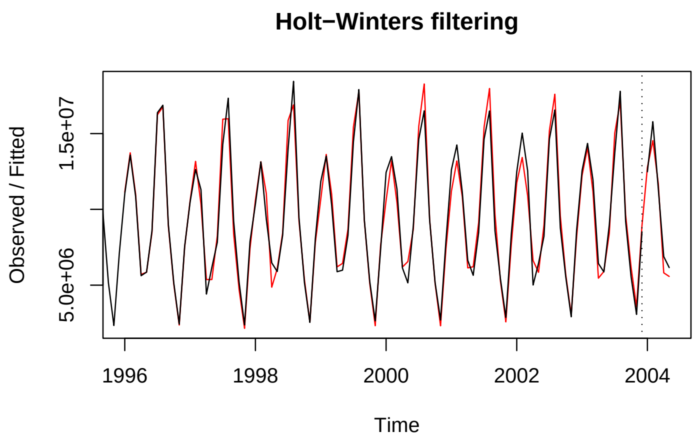
Niveau $\hat{a}_{T+1} = \alpha(x_T - \hat{s}_{T-p}) + (1-\alpha)(\hat{a}_T + \hat{b}_T)$

Trend $\hat{b}_{T+1} = \beta(\hat{a}_{T+1} - \hat{a}_T) + (1-\beta)\hat{b}_T$

Saison $\hat{s}_t = \gamma(x_t - \hat{a}_t) + (1-\gamma)\hat{s}_{t-p},$ t=T-p+1,. . . ,T

$$\hat{x}_T(1) = \hat{a}_{T+1} + \hat{b}_{T+1} + \hat{s}_{T+1-p}$$
$$\hat{x}_T(h) = \hat{a}_{T+1} + \hat{b}_{T+1} \cdot h + \hat{s}_{T+h-p}, \quad h = 2, \ldots, p$$
$$\hat{x}_T(h) = \hat{a}_{T+1} + \hat{b}_{T+1} \cdot h + \hat{s}_{T+h-2p}, \quad h = p+1, \ldots, 2p$$
$$usw.$$

11.3. Holt-Winters Verfahren



11.3. Holt-Winters Verfahren

Zur Anwendung des Holt-Winters Verfahrens:

- Das additive Holt-Winters Verfahren modelliert eine ZR als Summe einer Niveau-, einer lineare Trend- und einer Saisonkomponente und ist daher geeignet ZR zu prognostizieren, die einen annähernd linearen Trend und eine annähernd konstante Saisonalität aufweisen.
- Das multiplikative Holt-Winters Verfahren ist geeignet eine ZR zu modellieren, deren Amplitude anwächst.

11.3. Holt-Winters Verfahren

- Die 3 Glättungsparameter α, β und γ liegen zwischen 0 und 1. α dient zur Glättung des Niveaus, β zur Glättung des linearen Trends und γ zur Glättung der Saisonalität.
- Die Glättungsparameter können auch wieder durch Minimierung der Fehlerquadratsumme zwischen Einschritt-Vorhersage und wahrem Wert der ZR geschätzt werden.

12. Saison-Trend Modellierung in R

Die Verfahren zur Saison-Trendzerlegung und das Holt-Winters Verfahren sind in R mit Hilfe der Funkionen `decompose`, `stl` und `HoltWinters` durchzuführen.

Die Erklärungen zur Anwendung stehen in "SaisonTrendinR.pdf" und können von der LV-Seite heruntergeladen werden.

Übungen 4

Lösen Sie die folgenden Beispiele mit Hilfe von R:

1. Folgen Sie der Anleitung in "SaisonTrendinR.pdf" und rechnen Sie die dortigen Beispiele nach.

2. Laden Sie die Daten "USAccDeaths" in das Arbeitsverzeichnis. (Befehl: `data(USAccDeaths)`, ist ein Datensatz, der im Package "stats" in R enthalten ist. Nähere Informationen können über `help(''USAccDeaths'')` abgefragt werden.)

   (a) Glätten Sie die ZR mit der Methode der kleinen Trends.
   (b) Führen Sie die Zerlegung mittels der R-Funktion `decompose` durch.

Übungen 4

   (c) Führen Sie eine klassische Saison-Trend Zerlegung mittels der R-Funktion `stl` durch.

   In welchem Monat ist die Saisonalität am größten, wann am geringsten?

3. In "SaisonTrendinR.pdf" wurden die Nächtigungsdaten der Jahre 1995-2003 mit Hilfe des Holt-Winters Verfahrens geschätzt und die ersten 5 Monate des Jahres 2004 vorhergesagt. Dort wurden schon 2 verschiedene Modellansätze dafür verwendet. Verwenden Sie nun als dritten Ansatz ein Modell, in dem Sie die Glättungsparameter von R optimieren lassen.

Übungen 4

   Welches der 3 Modelle würden Sie wählen? Wie verändert sich die Fehlerquadratsumme? Welche mittlere Fehlerquadratsumme im Prognosezeitraum ergibt sich für das dritte Modell?

4. Die ZR "USAccDeaths" enthält Monatsdaten für die Jahre 1973-1978. Führen Sie das Holt-Winters Verfahren für ein Modell mit Trend und eines ohne Trend durch. Wählen Sie die Glättungsparameter `alpha=0.3`, `beta=0.1`, `gamma=0.4`.

   (a) Welches Modell würden Sie wählen?

Übungen 4

   (b) Prognostizieren Sie die ersten 6 Monate des Jahres 1979 mit dem gewählten Modell. Vergleichen Sie diese Vorhersagen mit den wahren Werten aus dem `help(''USAccDeaths'')`.

5. Laden Sie die Daten "co2" in das Arbeitsverzeichnis. (Befehl: `data(co2)`, ist ein Datensatz, der im Package "stats" in R enthalten ist. Nähere Informationen können über `help(''co2'')` abgefragt werden.)

   Das Holt-Winters Verfahren soll verwendet werden, um die Entwicklung bis zum Jahr 2000 vorherzusagen. Zeichnen Sie die ZR und ihre Prognose. Welche Komponenten sind im Modell enthalten?