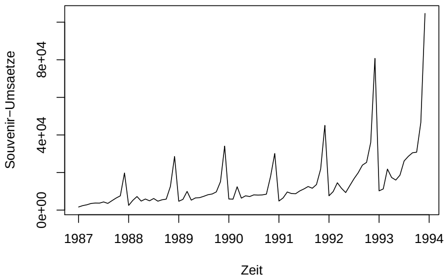
Kapitel 11

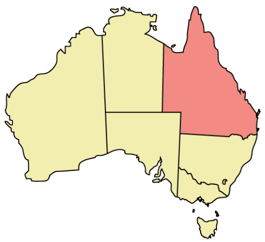
Explorative Zeitreihenanalyse

Einleitung: Souvenir-Verkäufe in Maroochydore

Souvenir-Verkäufe in Maroochydore

Glättung und Prognose von Umsätzen





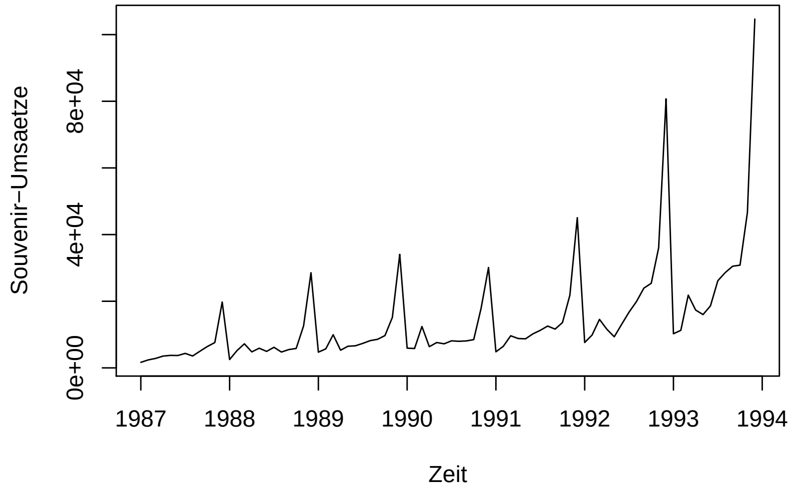
Daten

Im Januar 1987 eröffnet ein Souvenirgeschäft auf dem Kai des Strandbadeorts Maroochydore in Queensland, Australien. Die Verkäufe variieren über die Saison zusammen mit der Anzahl der Touristen. Insbesonders zu Weihnachten und zum örtlichen Surf-Festival, das jeden März seit 1988 veranstaltet wird, kommen die Urlauber an den Strand. Über die Jahre werden Geschäftsräume, Belegschaft und Produktpalette erweitert. Hier untersuchen wir die monatliche Zeitreihe der Umsätze von Anfang 1987 bis Ende 1993.

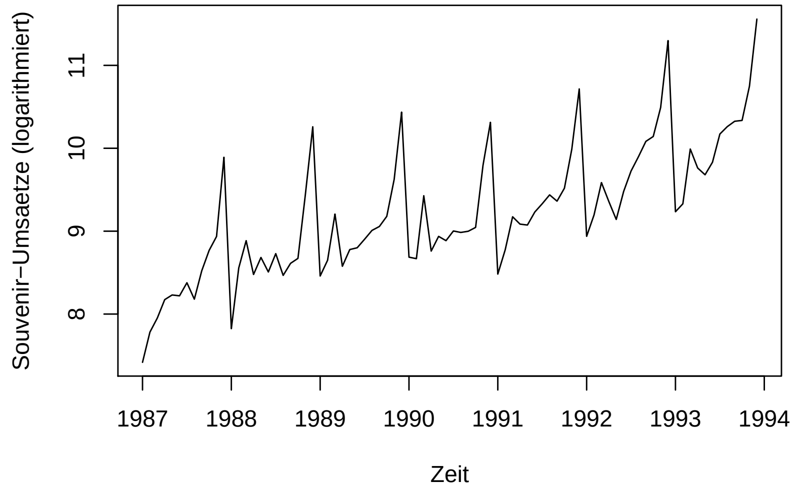
Fragen: Wie entwickelt sich der langfristige Trend? Besitzen die Verkaufszahlen eine regelmäßiges Muster, das sich jedes Jahr wiederholt? Wie können die Verkäufe für folgende Jahre vorhergesagt werden?

Quellen: Makridakis, Wheelwright and Hyndman(1998). *Forecasting, Methods and Applications*, Wiley, Beispiel 5.8, Seite 235f.

Daten



Daten



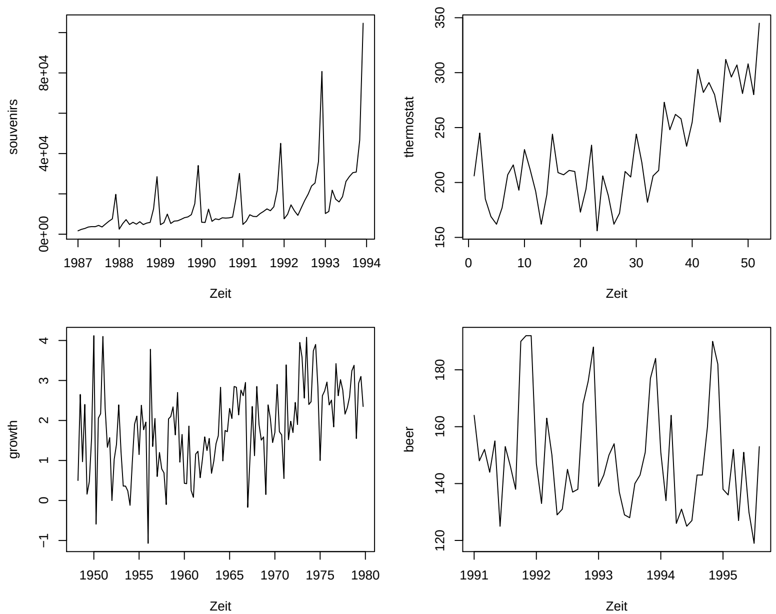
Beispiele für Zeitreihen

Im Gegensatz zu Querschnittsdaten stellen Zeitreihen der Zeit nach geordnete Beobachtungen dar. Viele Daten in der Wirtschaft fallen in dieser Form an.

Die mittel- bis langfristige Entwicklung kann recht unterschiedlich sein:

- `souvenirs`: Umsatzentwicklung von Souvenir-Verkäufen in Queensland (in Australischen Dollar), Monatsdaten, Januar 1987 – Dezember 1993.
- `thermostat`: Wöchentliche Thermostatverkäufe in Stück, 52 Wochen.
- `growth`: Einkommenswachstum in Iowa in Prozent, 2. Quartal 1948 – 4. Quartal 1978.
- `beer`: Bierproduktion in Australien in Megaliter, Monatsdaten, Januar 1991 – August 1995.

Zeitreihenplots



Beschreibung von Zeitreihen

- `souvenirs`: Die Strandumsätze steigen exponentiell bzw. geometrisch mit starken saisonalen Ausschlägen.
- `thermostat`: Die Thermostatnachfrage zeigt in der ersten Hälfte der Beobachtungsperiode einen leichten, linearen Rückgang und steigt in der zweiten Hälfte stark (linear) an.
- `growth`: Die Wachstumsreihe hat ebenfalls keinen stabilen Trend. Die mittelfristige Entwicklung wird von starken unregelmäßigen Schwankungen überlagert.
- `beer`: Die Bierproduktion verläuft leicht abfallend. Die saisonalen Schwankungen (hoher Bierkonsum im australischen Sommer, niedriger im Winter) sind stark ausgeprägt.

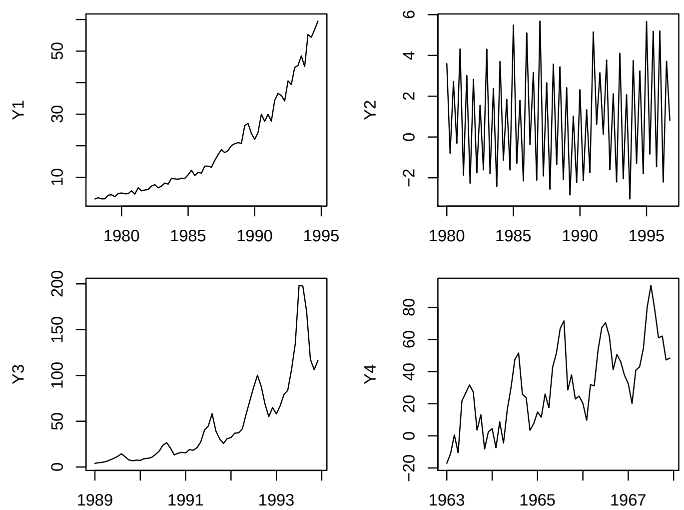
Zeitreihenmuster

- **Horizontales Muster:** Ein horizontales Muster zeigt über die gesamte Beobachtungsperiode ein konstantes Mittel und eine konstante Varianz.
- **Linearer Trend:** Reihen, die über die Zeit mit konstanter Steigung wachsen oder fallen, bezeichnet man als linearen Trend, sofern die Schwankungen um die Reihe konstant sind.
- **Geometrisches Wachstum, exponentieller Trend:** Liegt Wachstum vor, das sich ähnlich wie $\exp(x)$ entwickelt, spricht man von einem geometrischen oder exponentiellen Trend. Dieses Verhalten zeigt sich insbesonders im Zusammenhang mit Inflation oder Wirtschaftswachstum.

Zeitreihenmuster

- **Saison:** Saison beschreibt ein regelmäßig wiederkehrendes Muster im Rhythmus der Frequenz der Reihe. Das langfristige Verhalten wird dadurch nicht beeinflusst. Bei Monatsdaten liegt eine Frequenz von $s = 12$ Monaten, bei Quartalsdaten von $s = 4$ Quartalen vor.
- **Irreguläres Muster:** Ein Verlauf der ein völlig unregelmäßiges, unsystematisches bzw. zufälliges Bild zeigt, ähnlich den Residuen im Regressionsmodell (siehe Kapitel 9 und 10), wird als irregulär bezeichnet.

Beispiel: Zeitreihenmuster



Beispiel: Zeitreihenmuster

(a) Die Reihe Y1 wächst geometrisch. Die Saison wird ebenfalls exponentiell verstärkt.

(b) Die Reihe Y2 steigt linear ohne saisonale Effekte zu zeigen.

(c) Die Reihe Y4 steigt linear und weist zugleich saisonale Effekte auf.

(d) Die Reihe Y3 zeigt ein horizontales Muster ohne Saison.

(e) Die Reihe Y2 zeigt ein horizontales Muster mit Saison.

Beispiel: Zeitreihenmuster

(a) Die Reihe Y1 wächst geometrisch. Die Saison wird ebenfalls exponentiell verstärkt. falsch

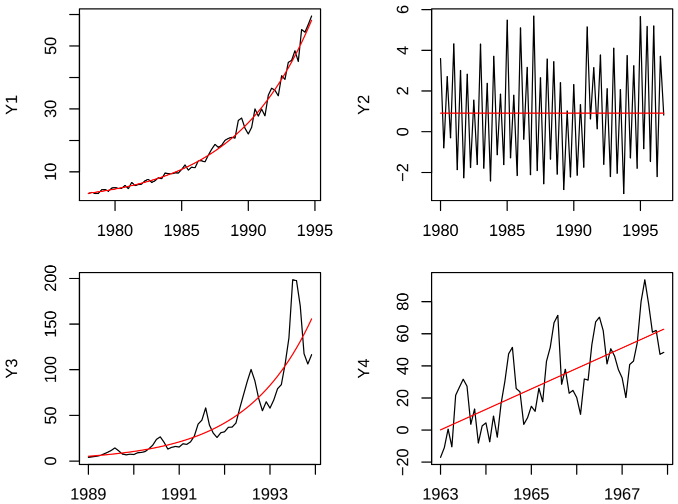
(b) Die Reihe Y2 steigt linear ohne saisonale Effekte zu zeigen. falsch

(c) Die Reihe Y4 steigt linear und weist zugleich saisonale Effekte auf. richtig

(d) Die Reihe Y3 zeigt ein horizontales Muster ohne Saison. falsch

(e) Die Reihe Y2 zeigt ein horizontales Muster mit Saison. richtig

Beispiel: Zeitreihenmuster



[Explorative Zeitreihenanalyse]

Die klassische Zeitreihenzerlegung

Zeitreihenzerlegung

Wir setzen eine Reihe Y_t aus 3 Komponenten zusammen:

- T_t = Trend und/oder Zyklus
- S_t = Saison
- E_t = irreguläre Komponente, Fehler

Die klassische Zeitreihenzerlegung ist **additiv**:

$$Y_t = T_t + S_t + E_t.$$

In diesem Modell wirken Trend und Saison unabhängig voneinander, d.h. beeinflussen einander nicht. Beispiel: `beer`.

Zeitreihenzerlegung

Manche Reihen lassen sich aber besser durch eine **multiplikative** Zerlegung beschreiben:

$$Y_t = T_t \cdot S_t \cdot E_t.$$

In diesem Modell wirken Trend und Saison *nicht* unabhängig, der saisonale Effekt wird durch den Trend verstärkt. Beispiel: `souvenirs`.

Durch Logarithmierung kann das multiplikative Modell in das additive überführt werden:

$$\begin{aligned} \log(Y_t) &= \log(T_t \cdot S_t \cdot E_t) \\ &= \log(T_t) + \log(S_t) + \log(E_t). \end{aligned}$$

Deshalb wird im folgenden nur das additive Modell behandelt und die zu untersuchenden Zeitreihen werden bei Bedarf logarithmiert.

Berechnung der Komponenten

Wir ermitteln die systematischen Komponenten T_t und S_t in 3 Schritten:

1. Berechnung der Trendkomponente T_t:
   Z.B. linearer, exponentieller Trend, mittelfristiger Trend durch Glättung
2. Berechnung der trendbereinigten (detrended) Reihe $TB_t = (Y_t - T_t)$.
3. Ermittlung der Saisonkomponente S_t, aus $TB_t = (Y_t - T_t)$.

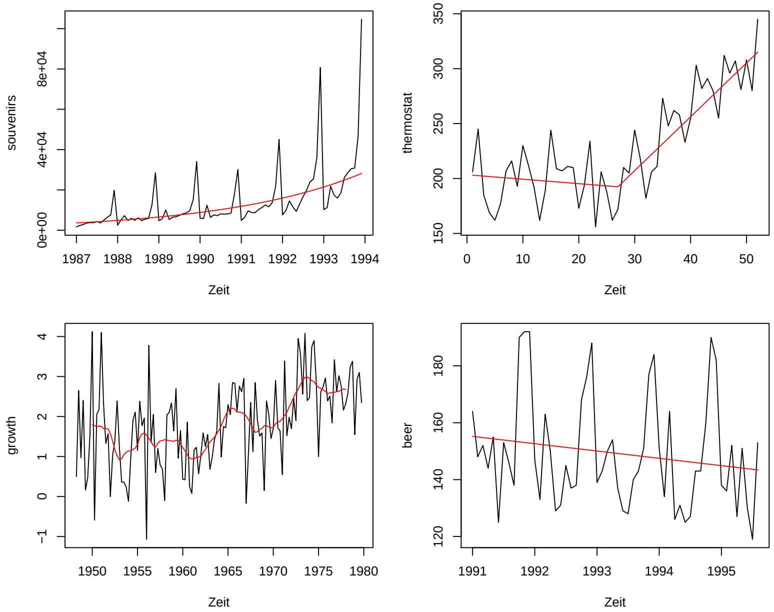
Als Rest ergibt sich die irreguläre Komponente, E_t.

$$E_t = Y_t - (T_t + S_t)$$

[Explorative Zeitreihenanalyse]

Trend

Der Trend der Reihen



Linearer und exponentieller Trend

Viele Zeitreihen lassen sich durch einen **linearen Trend** beschreiben: $T_t = \beta_0 + \beta_1 \cdot t$ mit Achsenabschnitt β_0 und Steigung β_1.

Als Modell für die gesamte Zeitreihe Y_t bedeutet dies:

$$Y_t = \beta_0 + \beta_1 \cdot t + \epsilon_t,$$

wobei $t = 1, 2, \dots, n$ die Trendvariable ist und ϵ_t die restliche Streuung (irreguläre Komponente und ggf. auch Saison) enthält. Dies ist ein einfaches Regressionsmodell, das mit der Methode der kleinsten Quadrate (KQ) geschätzt werden kann (siehe Kapitel 9).

Für das Modell log(`souvenirs`) $\sim$ `t`:

```
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  8.17023    0.11322    72.2   <2e-16
t            0.02470    0.00231    10.7   <2e-16
```


Linearer und exponentieller Trend

Das Modell für log(`souvenirs`) lautet damit

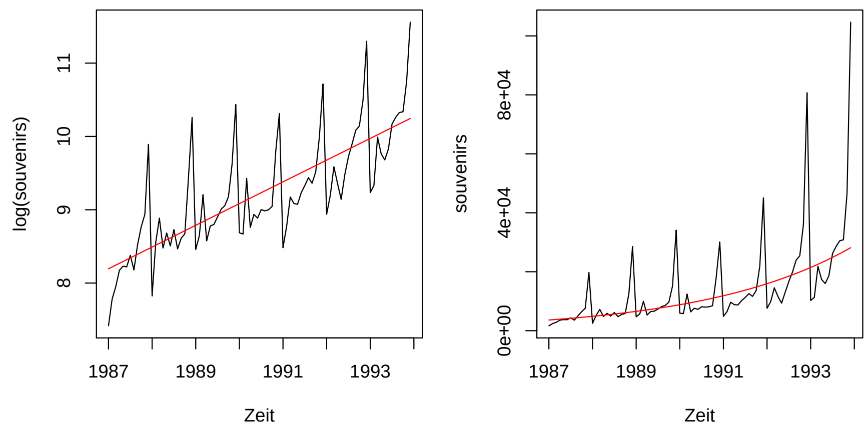
$$\log(\texttt{souvenirs}_t) = 8.170 + 0.025 \cdot t + \epsilon_t.$$

Um dies in ein Modell für `souvenirs` überzuführen, muss lediglich die Exponentialfunktion angewendet werden:

$$\begin{aligned} \texttt{souvenirs}_t &= e^{8.170 + 0.025 \cdot t + \epsilon_t} \\ &= e^{8.170} \cdot e^{0.025 \cdot t} \cdot e^{\epsilon_t} \\ &= 3534.145 \cdot e^{0.025 \cdot t} \cdot \eta_t. \end{aligned}$$

Ein solches Modell nennt man auch **exponentiellen Trend**.

Linearer und exponentieller Trend



Gleitender Durchschnitt

Liegt kein stabiler linearer Trend vor, kann ein **gleitender Durchschnitt** (**moving average, MA**) verwendet werden, um die mittelfristige Entwicklung sichtbar zu machen.

Der Trend T_t zum Zeitpunkt t berechnet sich als Durchschnitt der m Beobachtungen links bzw. rechts von t sowie Y_t selbst:

$$T_t = \frac{1}{2m+1}(Y_{t-m} + \ldots + Y_t + \ldots + Y_{t+m})$$

Insgesamt wird also das Mittel von $k = 2 \cdot m + 1$ (*ungerade*) benachbarten Beobachtungen berechnet. Dies ist möglich für $t = m+1, m+2, \ldots, n-m$.

Wir nennen T_t den k-gliedrigen gleitenden Durchschnitt oder kurz k**-MA**.

Gewichteter gleitender Durchschnitt

Saisonale Daten hingegen weisen oft eine *gerade* Periodenlänge auf: 4 für Quartalsdaten, 12 für Monatsdaten. Wir verwenden hier statt eines k-MA einen geeigneten **gewichteten gleitenden Durchschnitt**.

- Für Quartalsdaten, $s = 4$:

$$T_t = (0.5 \cdot Y_{t-2} + Y_{t-1} + Y_t + Y_{t+1} + 0.5 \cdot Y_{t+2})/4$$

  Die Summe der 5 Gewichte $\{1/8, 1/4, 1/4, 1/4, 1/8\}$ ist 1.

- Für Monatsdaten, $s = 12$:

$$T_t = (0.5 \cdot Y_{t-6} + Y_{t-5} + \ldots + Y_t + \ldots + Y_{t+5} + 0.5 \cdot Y_{t+6})/12$$

  Die Summe der 13 Gewichte $\{1/24, 1/12, \ldots, 1/12, 1/24\}$ ist 1. Z.B. für den Juli-Wert gehen die Januarmonate zweier aufeinanderfolgender Jahre je zur Hälfte in die Berechnung ein.

Beispiel: 3-MA

Gegeben ist die Reihe Y_t der Länge $n = 8$.

t	1	2	3	4	5	6	7	8
Y_t	1	2	1	3	1	0	0	2
T_t								

Berechnen Sie den 3-MA für $t = 2$.

Beispiel: 3-MA

Gegeben ist die Reihe Y_t der Länge $n = 8$.

t	1	2	3	4	5	6	7	8
Y_t	1	2	1	3	1	0	0	2
T_t		1.333						

Berechnen Sie den 3-MA für $t = 2$.

$$\begin{aligned} T_2 &= (1+2+1)/3 \\ &= 4/3 \\ &= 1.333 \end{aligned}$$

Beispiel: 3**-MA**

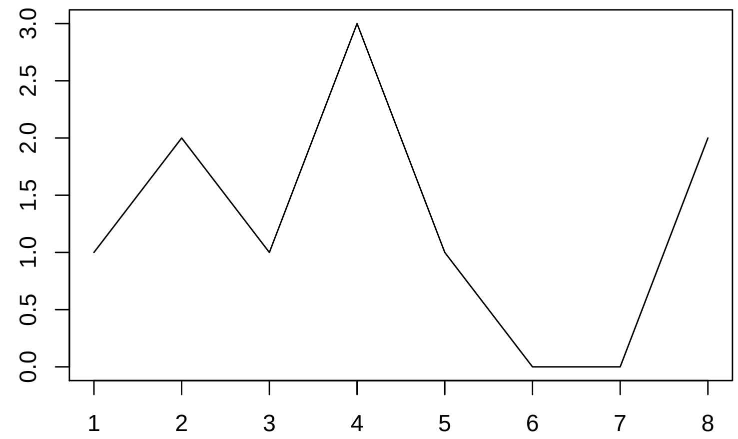
Gegeben ist die Reihe Y_t der Länge $n = 8$.

t	1	2	3	4	5	6	7	8
Y_t	1	2	1	3	1	0	0	2
T_t		1.333	2.000	1.667	1.333	0.333	0.667	

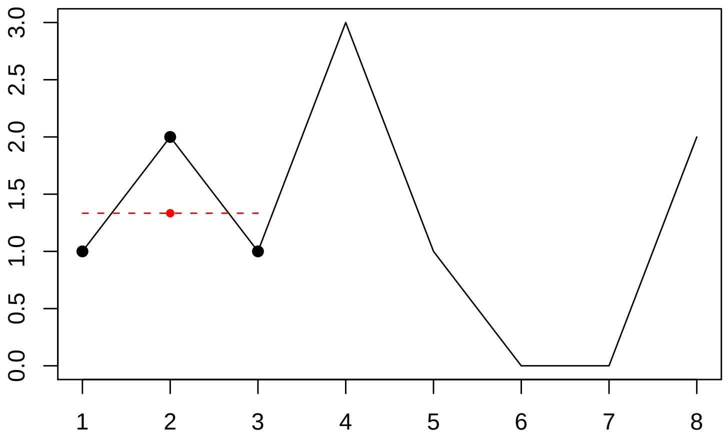
Berechnen Sie den 3-MA für $t = 2$.

$$\begin{aligned} T_2 &= (1 + 2 + 1)/3 \\ &= 4/3 \\ &= 1.333 \end{aligned}$$

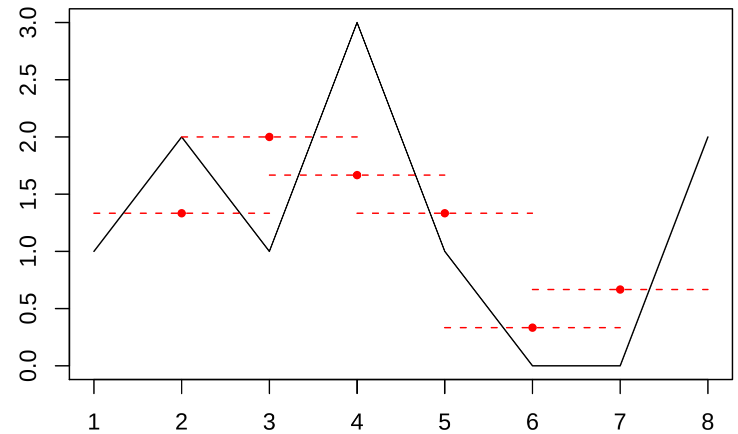
Beispiel: 3**-MA**



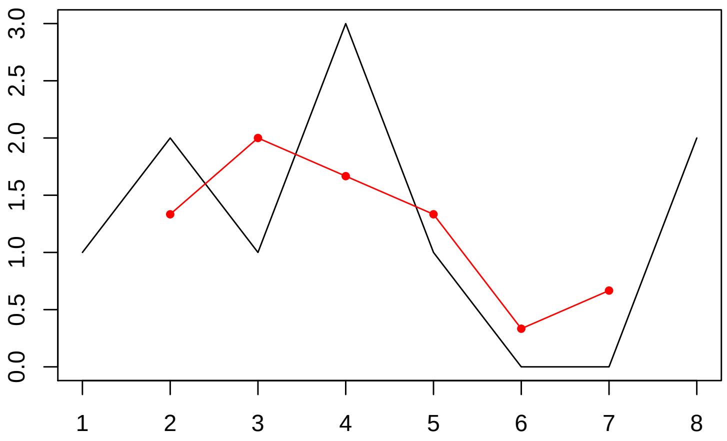
Beispiel: 3**-MA**



Beispiel: 3**-MA**



Beispiel: 3-MA



Beispiel: 5-MA

Gegeben ist die Reihe Y_t der Länge $n = 8$.

t	1	2	3	4	5	6	7	8
Y_t	1	2	1	3	1	0	0	2
T_t								

Berechnen Sie den 5-MA für $t = 3$.

Beispiel: 5-MA

Gegeben ist die Reihe Y_t der Länge $n = 8$.

t	1	2	3	4	5	6	7	8
Y_t	1	2	1	3	1	0	0	2
T_t			1.600					

Berechnen Sie den 5-MA für $t = 3$.

$$\begin{aligned} T_3 &= (1+2+1+3+1)/5 \\ &= 8/5 \\ &= 1.600 \end{aligned}$$

Beispiel: 5-MA

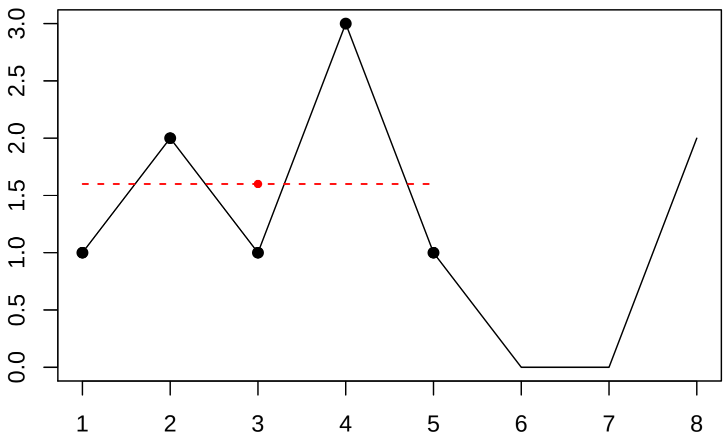
Gegeben ist die Reihe Y_t der Länge $n = 8$.

t	1	2	3	4	5	6	7	8
Y_t	1	2	1	3	1	0	0	2
T_t			1.600	1.400	1.000	1.200		

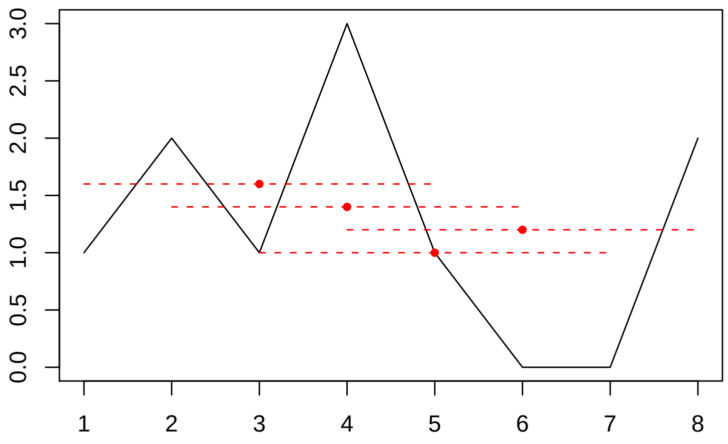
Berechnen Sie den 5-MA für $t = 3$.

$$\begin{aligned} T_3 &= (1+2+1+3+1)/5 \\ &= 8/5 \\ &= 1.600 \end{aligned}$$

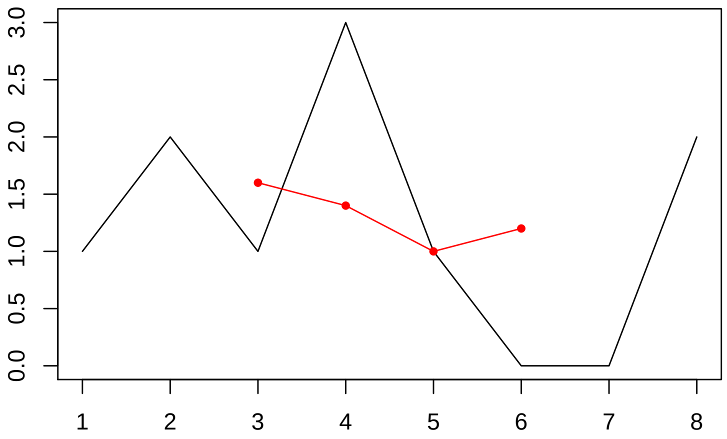
Beispiel: 5-MA



Beispiel: 5-MA



Beispiel: 5-MA



Beispiel: Gewichteter gleitender Durchschnitt

Gegeben ist die Reihe Y_t der Länge $n = 8$ und $s = 4$.

t	1	2	3	4	5	6	7	8
Y_t	1	2	1	3	1	0	0	2
T_t								

Berechnen Sie den gewichteten gleitenden Durchschnitt für $t = 3$.

Beispiel: Gewichteter gleitender Durchschnitt

Gegeben ist die Reihe Y_t der Länge $n = 8$ und $s = 4$.

t	1	2	3	4	5	6	7	8
Y_t	1	2	1	3	1	0	0	2
T_t			1.750					

Berechnen Sie den gewichteten gleitenden Durchschnitt für $t = 3$.

$$\begin{aligned} T_3 &= (0.5 \cdot 1 + 2 + 1 + 3 + 0.5 \cdot 1)/4 \\ &= 7/4 \\ &= 1.750 \end{aligned}$$

Beispiel: Gewichteter gleitender Durchschnitt

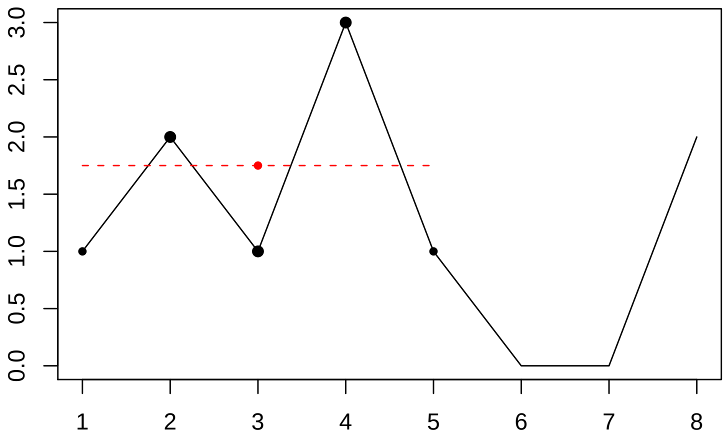
Gegeben ist die Reihe Y_t der Länge $n = 8$ und $s = 4$.

t	1	2	3	4	5	6	7	8
Y_t	1	2	1	3	1	0	0	2
T_t			1.750	1.500	1.125	0.875		

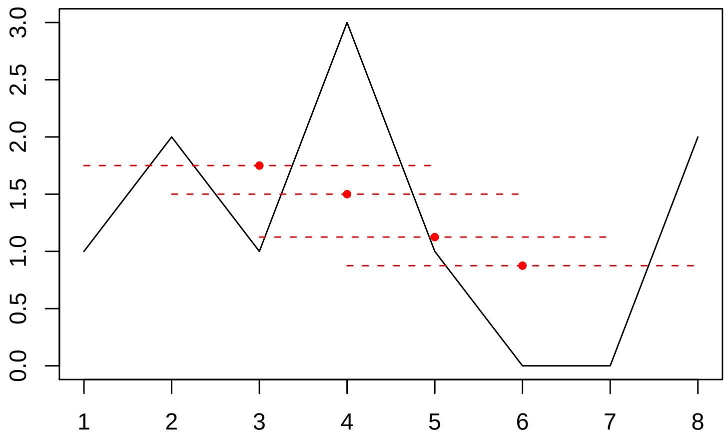
Berechnen Sie den gewichteten gleitenden Durchschnitt für $t = 3$.

$$\begin{aligned} T_3 &= (0.5 \cdot 1 + 2 + 1 + 3 + 0.5 \cdot 1)/4 \\ &= 7/4 \\ &= 1.750 \end{aligned}$$

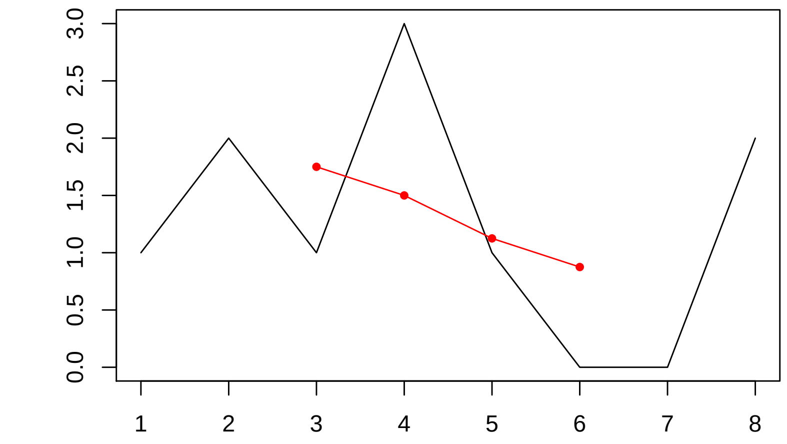
Beispiel: Gewichteter gleitender Durchschnitt



Beispiel: Gewichteter gleitender Durchschnitt



Beispiel: Gewichteter gleitender Durchschnitt

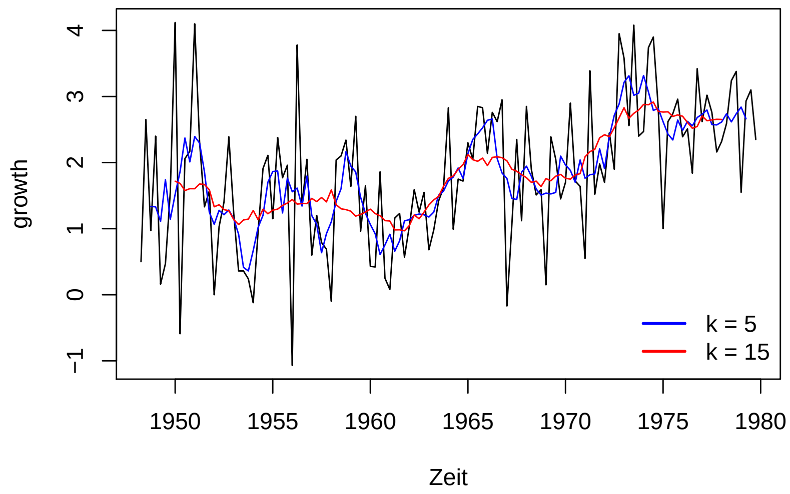


Stärke der Glättung eines k-MA

Die Stärke der Glättung eines gleitenden Durchschnitts hängt von der Wahl von k (bzw. m) ab:

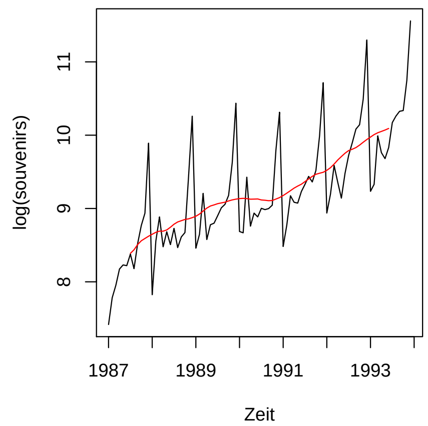
- Für große k (bzw. m) wird Y_t stark geglättet,
- für kleine k (bzw. m) wird Y_t wenig geglättet.

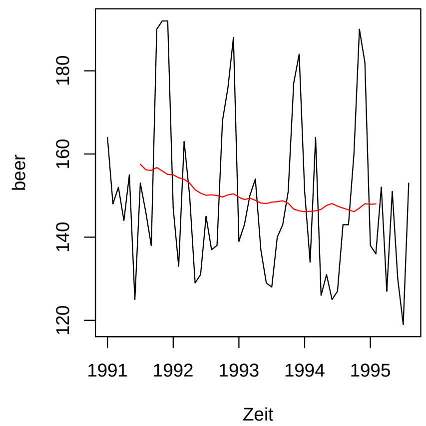
Glättung mit k-MA



Glättung saisonaler Reihen

Wir glätten log(`souvenirs`) und `beer` mittels des saisonalen gewichteten Durchschnitts, $s = 12$.





Einfache exponentielle Glättung

Die geglättete Reihe T_t wird rekursiv als gewichtetes Mittel aus der aktuellen Beobachtung und der Glättung der Vorperiode berechnet.

$$T_t = \alpha \cdot Y_t + (1 - \alpha) \cdot T_{t-1}$$

Der Glättungsparameter α liegt zwischen 0 und 1, $0 < \alpha \leq 1$.

- Ist $\alpha = 1.00$, ist die geglättete Reihe gleich der ursprünglichen. Es wird nicht geglättet.
- Ist α groß, z.B. 0.75, so ist der aktuelle Wert sehr wichtig, der Einfluss des geglätteten Wertes der Vorperiode (in dem auch die gesamte Vergangenheit der Reihe eingeht) gering. Es wird wenig geglättet.
- Ist α klein, z.B. 0.10, so kann der aktuelle Y-Wert die Glättung nur wenig beeinflussen. Es wird stark geglättet.

Beispiel: einfache exponentielle Glättung

Gegeben ist die Reihe Y_t der Länge $n = 8$ und $\alpha = 0.6$.

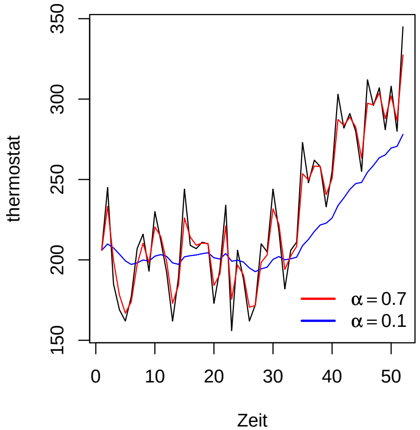
$$T_t = 0.6 \cdot Y_t + 0.4 \cdot T_{t-1}$$

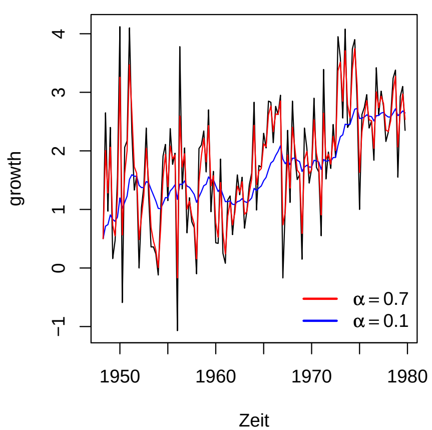
Der Anfangswert für T_1 sei $T_1 = Y_1 = 1$.

t	1	2	3	4	5	6	7	8
Y_t	1	2	1	3	4	0	1	2
T_t	1	1.6	1.24	2.296	3.318	1.327	1.131	1.652

$T_2 = 0.6 \cdot Y_2 + 0.4 \cdot T_1 = 0.6 \cdot 2 + 0.4 \cdot 1 = 1.6$
$T_3 = 0.6 \cdot Y_3 + 0.4 \cdot T_2 = 0.6 \cdot 1 + 0.4 \cdot 1.6 = 1.24$
$\ldots$
$T_8 = 0.6 \cdot Y_8 + 0.4 \cdot T_7 = 0.6 \cdot 2 + 0.4 \cdot 1.131 = 1.652$

Exponentielle Glättung von `thermostat` und `growth`





Exponentielle Glättung von `thermostat` und `growth`

Zu `thermostat`:
Die Glättung mit $\alpha = 0.1$ liegt im ansteigenden Bereich systematisch unterhalb der Beobachtungen. Daher ist sie nicht geeignet. Die andere Variante glättet kaum, verfolgt deutlich den Pfad.

Zu `growth`:
Die Glättung mit $\alpha = 0.1$ erfasst weitgehend die wesentlichen Bewegungen. Die Variante $\alpha = 0.7$ glättet kaum, da sie jeden Ausschlag fast zu Gänze nachbildet.

Trendbereinigung

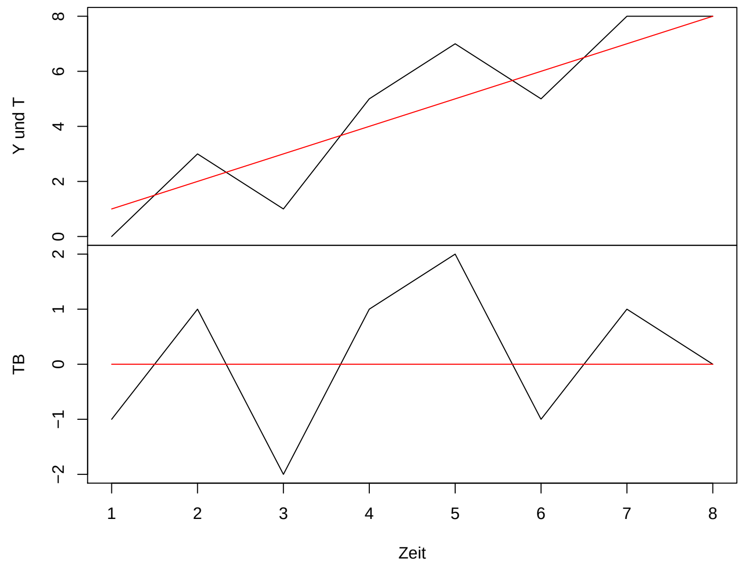
Die trendbereinigte Reihe, TB_t, ergibt sich aus Y_t nach Abzug des Trends.

$$TB_t = Y_t - T_t$$

Beispiel: Berechnung von TB
Gegeben ist Y_t und die bereits berechnete Trendkomponente T_t.

t	1	2	3	4	5	6	7	8
Y_t	0	3	1	5	7	5	8	8
T_t	1	2	3	4	5	6	7	8
TB_t	-1	1	-2	1	2	-1	1	0

Trendbereinigung



[Explorative Zeitreihenanalyse]

Saison

Saisonkomponente

Zur Berechnung der Saisonkomponenten gehen wir von der trendbereinigten Reihe TB_t aus. Wir rechnen die Saisonkomponente für jede Saison als den Durchschnitt der trendbereinigten Werte der jeweiligen Saison über alle Jahre.

Beispiel: Wie haben Monatsdaten über 10 Jahre gegeben. Dann berechnen wir die Januarsaison als Durchschnitt aller trendbereinigten Januarwerte über alle 10 Jahre. Analog für Februar, . . . , Dezember.

Man erhält daher für jede Saison *einen* Saisonwert, also z.B. 12 Monatswerte bei Monatsdaten oder 4 Quartalswerte für Quartalsdaten.

Beispiel: Saisonkomponenten

Gegeben ist die trendbereinigte Quartalsreihe TB_t der Länge 12. Sie besitzt 3 volle Jahre.

	1998				1999				2000			
Quartal	1	2	3	4	1	2	3	4	1	2	3	4
TB_t	2	1	-3	3	4	1	-1	2	6	-2	-2	1
S_t												

Die Quartalseffekte sind

$$S_1 = (2+4+6)/3 = 4 \qquad S_2 = (1+1-2)/3 = 0$$
$$S_3 = (-3-1-2)/3 = -2 \qquad S_4 = (3+2+1)/3 = 2$$

Beispiel: Saisonkomponenten

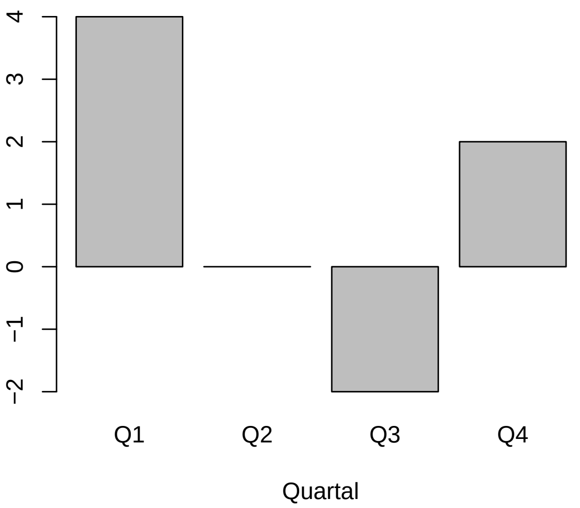
Gegeben ist die trendbereinigte Quartalsreihe TB_t der Länge 12. Sie besitzt 3 volle Jahre.

	1998				1999				2000			
Quartal	1	2	3	4	1	2	3	4	1	2	3	4
TB_t	2	1	-3	3	4	1	-1	2	6	-2	-2	1
S_t	4	0	-2	2	4	0	-2	2	4	0	-2	2

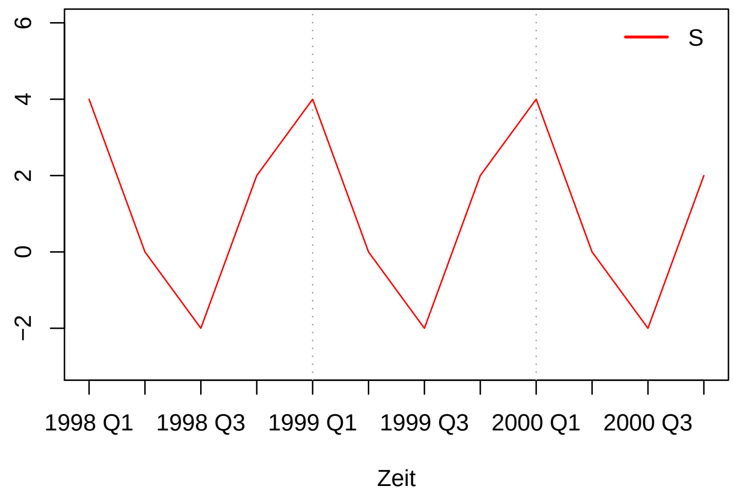
Die Quartalseffekte sind

$$S_1 = (2+4+6)/3 = 4 \qquad S_2 = (1+1-2)/3 = 0$$
$$S_3 = (-3-1-2)/3 = -2 \qquad S_4 = (3+2+1)/3 = 2$$

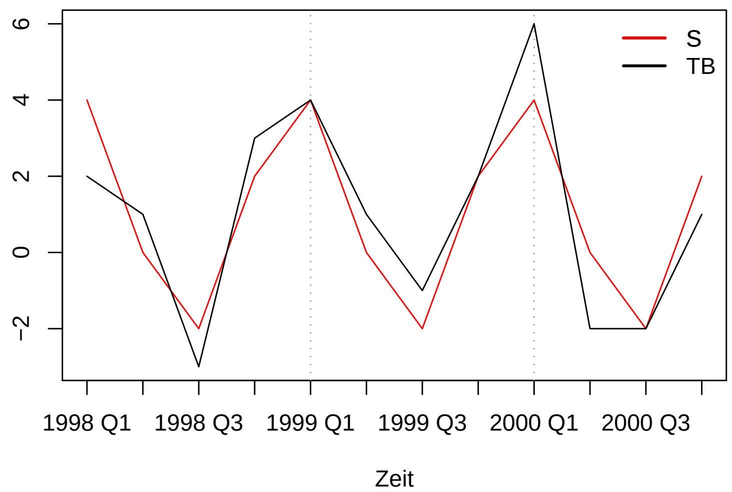
Beispiel: Saisonkomponenten



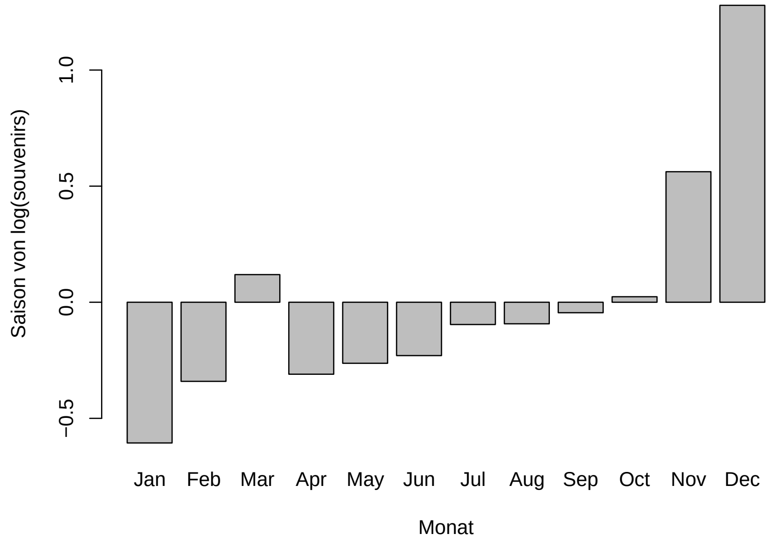
Beispiel: Saisonkomponenten



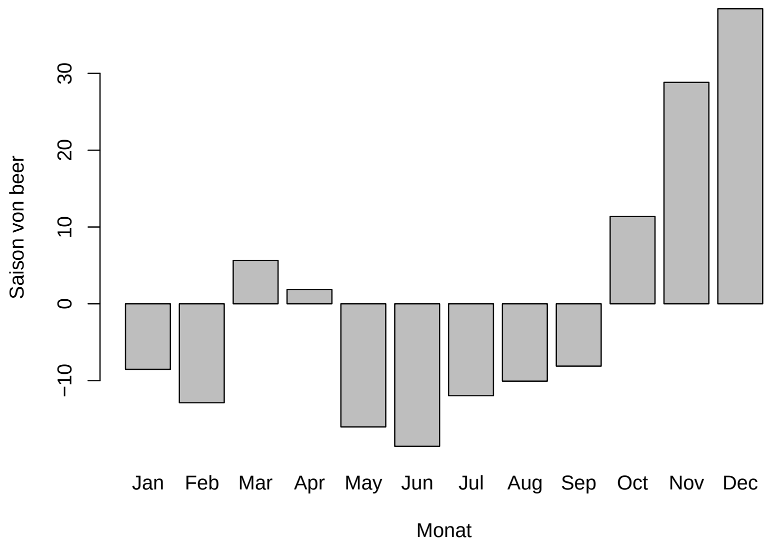
Beispiel: Saisonkomponenten



Saisonkomponente von log(`souvenirs`)



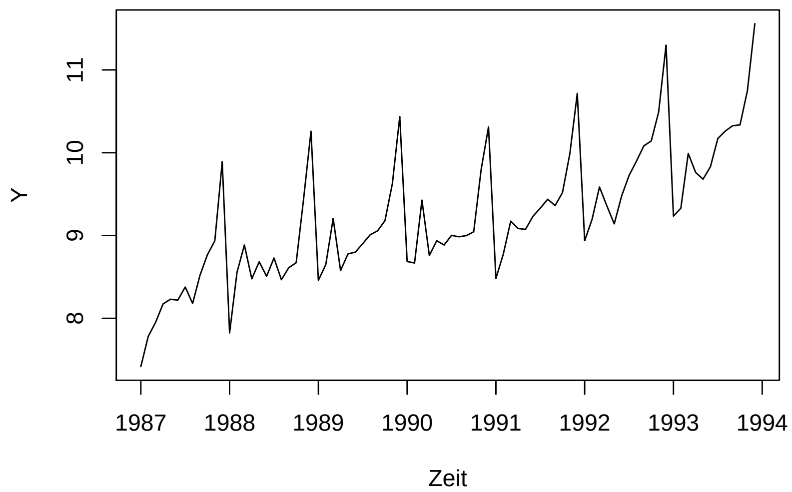
Saisonkomponente von `beer`



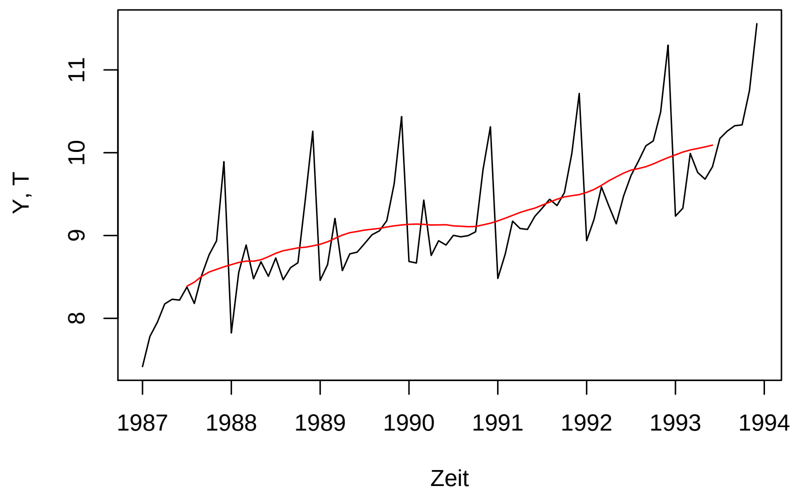
[Explorative Zeitreihenanalyse]

Zusammenfassung der Zeitreihenzerlegung

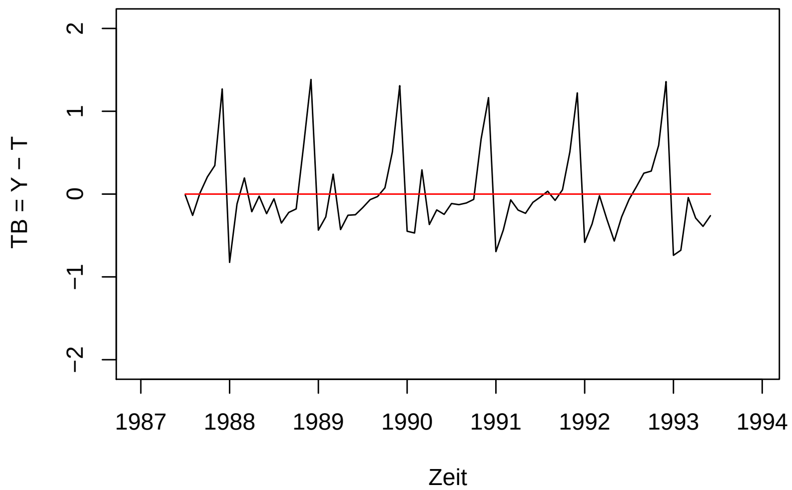
Zeitreihenzerlegung von log(souvenirs)



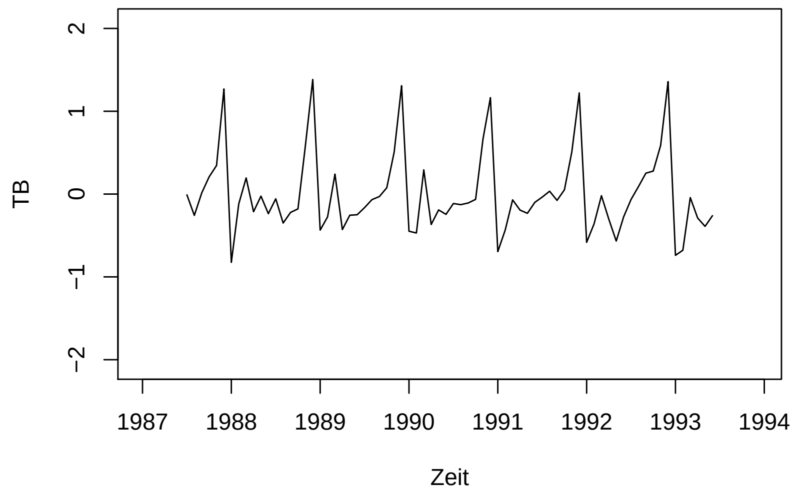
Zeitreihenzerlegung von log(souvenirs)



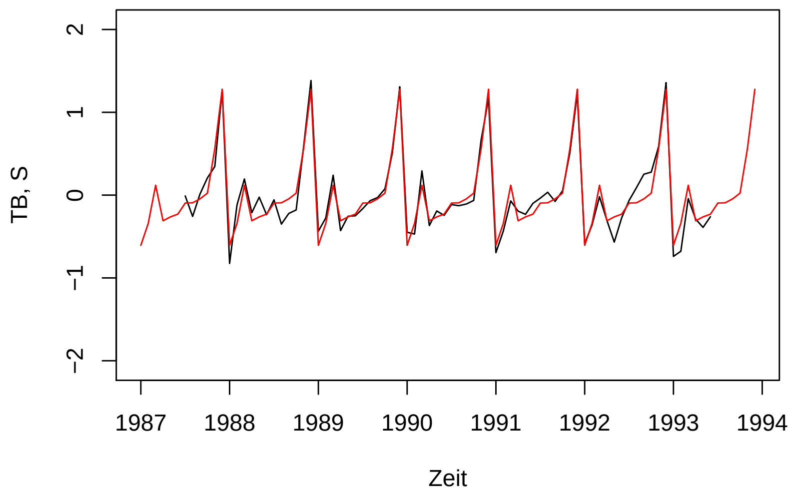
Zeitreihenzerlegung von log(souvenirs)



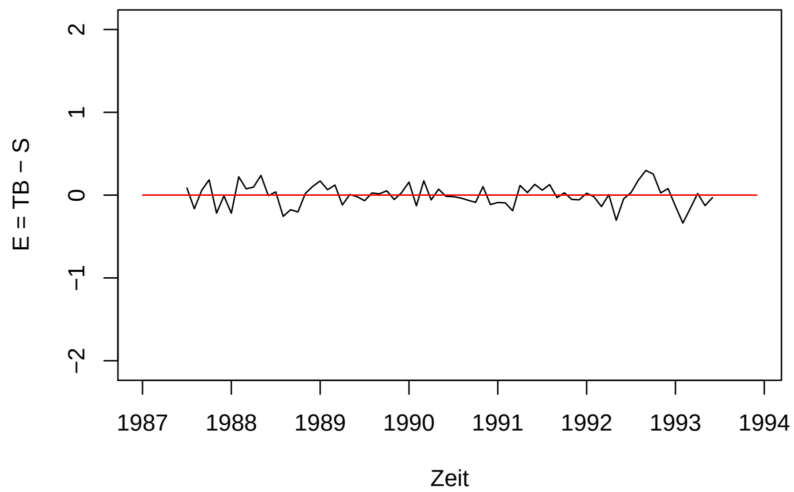
Zeitreihenzerlegung von log(souvenirs)



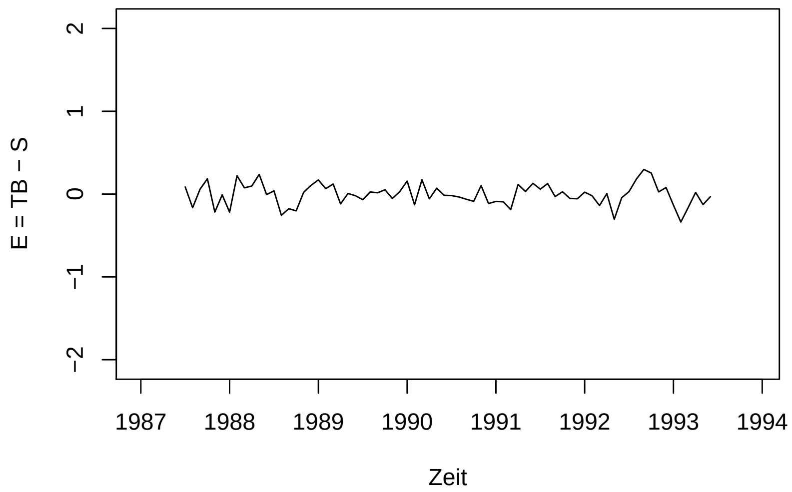
Zeitreihenzerlegung von log(souvenirs)



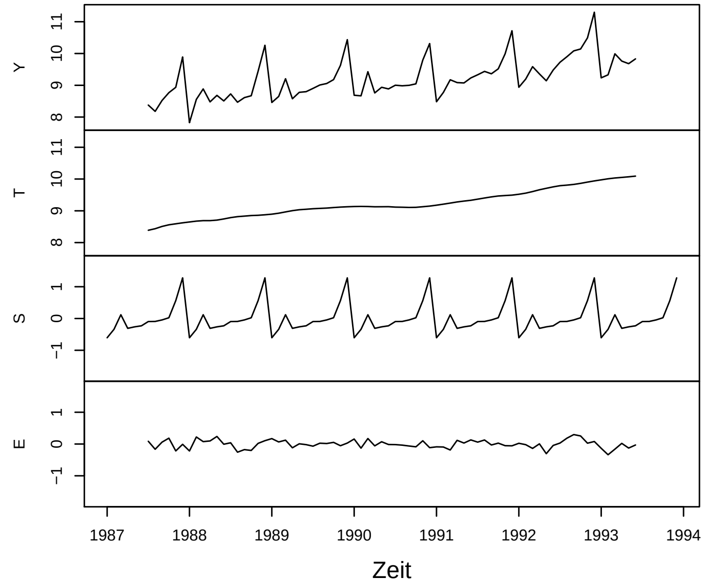
Zeitreihenzerlegung von log(souvenirs)



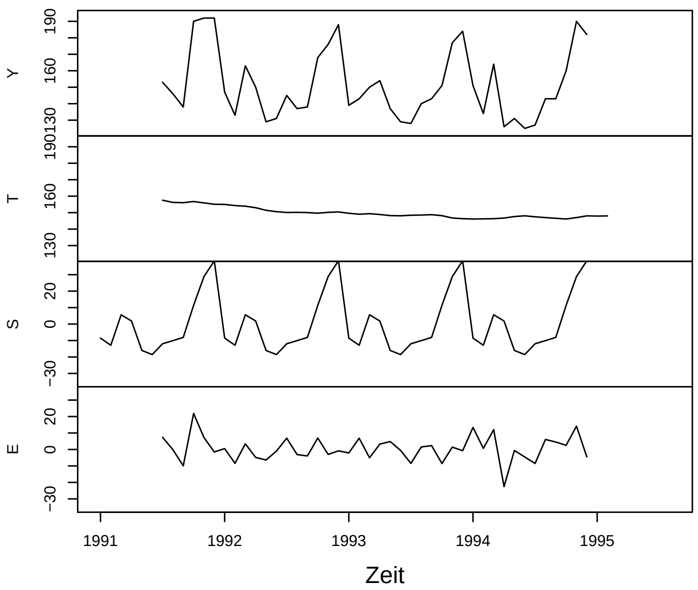
Zeitreihenzerlegung von log(souvenirs)



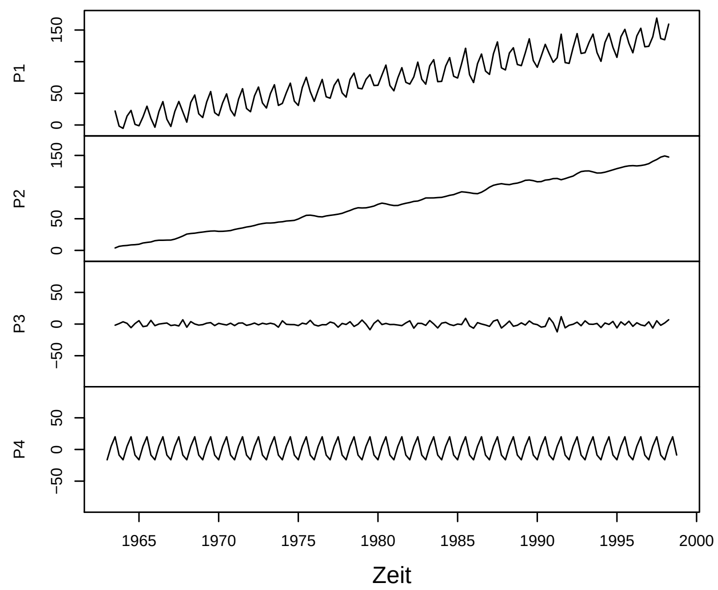
Zeitreihenzerlegung von log(souvenirs)



Zeitreihenzerlegung von `beer`



Beispiel: Zeitreihenzerlegung



Beispiel: Zeitreihenzerlegung

Die Komponenten der Zeitreihenzerlegung sind zufällig angeordnet. Welche der folgenden Zuordnungen sind richtig?

(a) Der Plot P1 zeigt die beobachtete Reihe.

(b) Der Plot P2 weist die irreguläre Komponente aus.

(c) Der Plot P3 gibt den mittelfristigen Verlauf der Reihe wieder (Trend).

(d) Der Plot P4 macht das sich wiederholende Saisonmuster sichtbar.

(e) Der Plot P3 zeigt die beobachtete Reihe.

Beispiel: Zeitreihenzerlegung

Die Komponenten der Zeitreihenzerlegung sind zufällig angeordnet. Welche der folgenden Zuordnungen sind richtig?

(a) Der Plot P1 zeigt die beobachtete Reihe. richtig

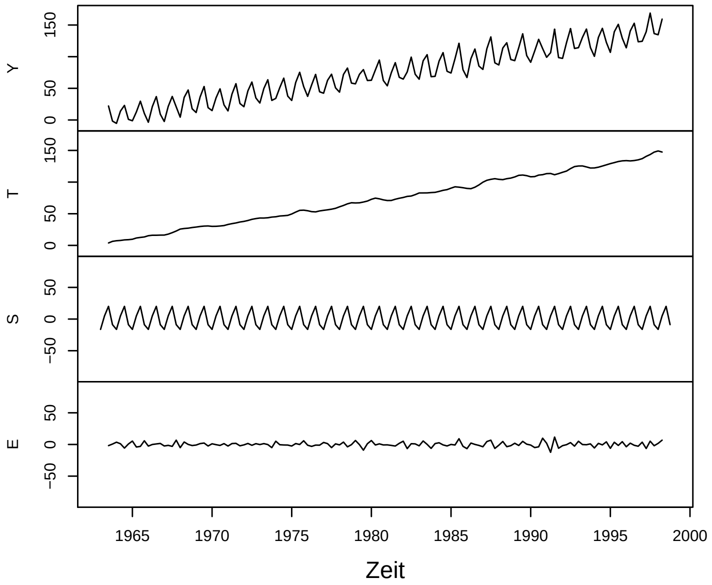
(b) Der Plot P2 weist die irreguläre Komponente aus. falsch

(c) Der Plot P3 gibt den mittelfristigen Verlauf der Reihe wieder (Trend). falsch

(d) Der Plot P4 macht das sich wiederholende Saisonmuster sichtbar. richtig

(e) Der Plot P3 zeigt die beobachtete Reihe. falsch

Beispiel: Zeitreihenzerlegung



[Explorative Zeitreihenanalyse]

Prognose

Beobachtungen und Prognosen

Beobachtungsperiode				⋮	Prognoseperiode		
Y_1	Y_2	$\dots$	Y_{n-1}	Y_n ⋮	Y_{n+1}	$\dots$	Y_{n+r}
historische Daten				⋮	zukünftige Beobachtungen		

Wir kennen die Daten für die Periode $t = 1, \dots, n$ und möchten Prognosen (oder Vorhersagen) für die zukünftigen Zeitpunkte $t = n + 1, \dots, n + r$ berechnen.

Bezeichnungen:

- $\hat{Y}_{n+r}$ bezeichnet die Prognose für die zukünftige Beobachtung Y_{n+r} (r-Schritt Prognose)
- n gibt den Prognosezeitpunkt an,
- r den Vorhersagehorizont.

Prognosefehler

Um die Qualität der Prognose zu beurteilen, kann man die Abweichungsquadratsummen zwischen der tatsächlichen, zukünftigen Beobachtung und der Prognose berechnen:

$$RSS = \sum_{j=1}^{r} (Y_{n+j} - \hat{Y}_{n+j})^2.$$

Üblicherweise berechnet man als Kenngrößen den Mittelwert der Abweichungsquadrate, den **mean squared error (*MSE*)**

$$MSE = \frac{1}{r} \sum_{j=1}^{r} (Y_{n+j} - \hat{Y}_{n+j})^2$$

oder die Wurzel des *MSE*, den **root mean squared error (*RMSE*)**

$$RMSE = \sqrt{MSE}.$$

Prognose mit einfacher exponentieller Glättung

Idee: Zum Zeitpunkt t erhält man die 1-Schritt-Prognose für den kommenden Zeitpunkt $t+1$ als gewichteten Durchschnitt aus dem derzeitigen Wert Y_t und dem erwarteten Wert $\hat{Y}_t$ (eigentlich: $\hat{Y}_{(t-1)+1}$) für den derzeitigen Zeitpunkt.

$$\text{Niveau:} \quad \hat{Y}_{t+1} = \alpha \cdot Y_t + (1-\alpha) \cdot \hat{Y}_t, \qquad 0 < \alpha \leq 1$$

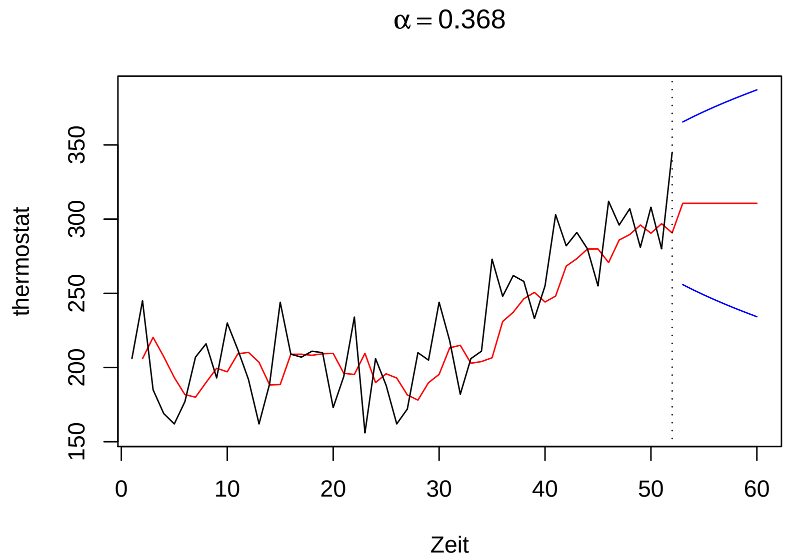
Wir starten mit dem Anfangswert: $\hat{Y}_1 = Y_1$ und können damit die 1-Schritt-Prognosen $\hat{Y}_2, \ldots, \hat{Y}_{n+1}$ berechnen.

Alle weiteren Prognosen über n hinaus sind auch gleich $\hat{Y}_{n+1}$:

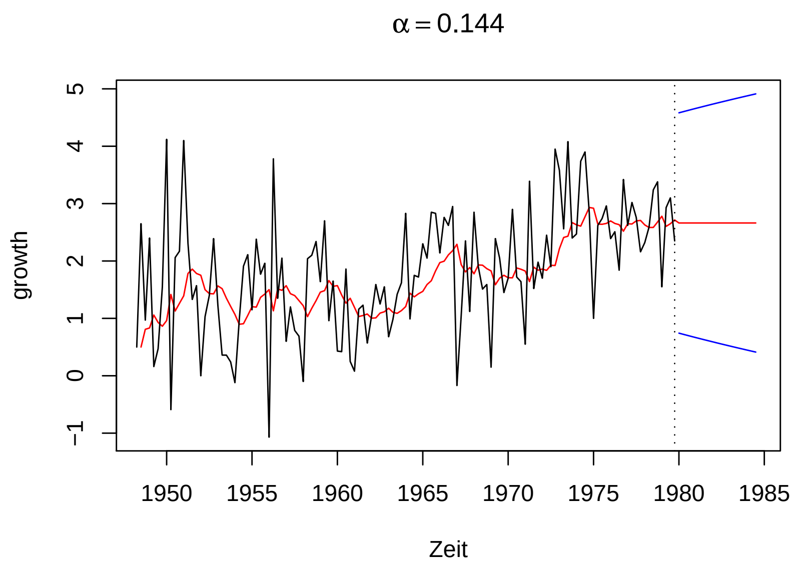
$$\hat{Y}_{n+2} = \cdots = \hat{Y}_{n+r} = \hat{Y}_{n+1}$$

Den optimalen Wert für α ermittelt man durch Minimierung der Summe der Fehlerquadratsumme im Beobachtungszeitraum: $\sum_t (Y_t - \hat{Y}_t)^2$.

Prognose mit einfacher exponentieller Glättung



Prognose mit einfacher exponentieller Glättung



Prognose mit einfacher exponentieller Glättung

Die Zeitreihen werden geglättet und mit den Punktprognosen $\hat{Y}_{n+1} = \hat{Y}_{n+2} = \ldots$ fortgesetzt. Zusätzlich werden 95% Prognoseintervalle angegeben. Sie zeigen, dass bei beiden Reihen eine beträchtliche Unsicherheit über den zukünftigen Verlauf besteht.

- `thermostat`: Das Niveau wird zum Anteil 36.8% (α) durch den aktuellen Wert bestimmt.
- `growth`: Für die Reihe `growth` spielt das Niveau vergangener Beobachtungen eine größere Rolle, hier gehen die aktuellen Werte nur zum Anteil 14.4% (α) ein.

Bei der einfachen exponentiellen Glättung wird ein konstanter Verlauf fortgeschrieben. Speziell bei der Reihe `thermostat` sollte aber wohl ein Trend mit einbezogen werden. Dies ist mit der Holt-Glättung möglich (einer Verallgemeinerung der einfachen exponentiellen Glättung).

Prognose mit einfacher linearer Regression

Liegt eine Reihe mit einem stabilen linearen Trend vor, können wir die Trend-Regression für Prognosen verwenden.
Für $Y_t = \log(\texttt{souvenirs}_t)$ haben wir die Gleichung (ohne saisonale Effekte zu berücksichtigen) für $t = 1, \ldots, 84$ bereits geschätzt:

$$Y_t = 8.170 + 0.025 \cdot t + \hat{\epsilon}_t.$$

Für die Prognose ersetzen wir ϵ_t durch $E(\epsilon_t) = 0$. Damit erhalten wir die Prognosefunktion

$$\hat{Y}_t = 8.170 + 0.025 \cdot t$$

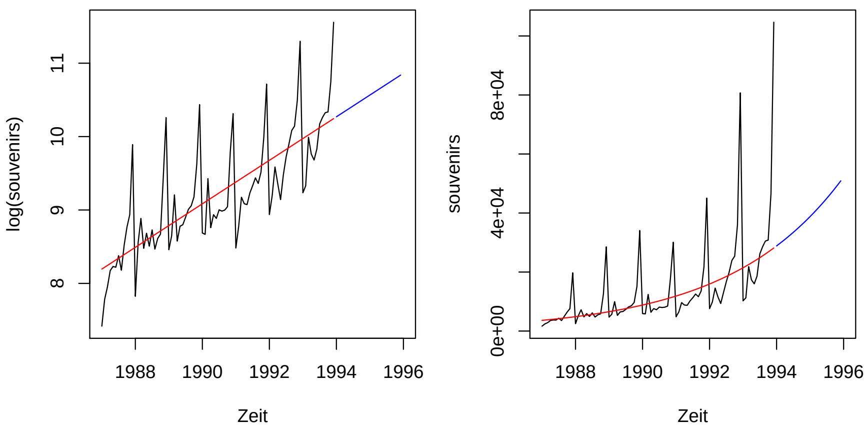
in die wir für $t = 85, 86, \ldots$ einsetzen.

Prognose mit einfacher linearer Regression

Die Prognosen für die Perioden $85, \ldots, 108$ sind:

t	r	Periode	$\hat{Y}_{84+r}$			$\exp(\hat{Y}_{84+r})$
85	1	Jan 1994	$8.170 + 0.025 \cdot 85$	=	10.270	28847.3
86	2	Feb 1994	$8.170 + 0.025 \cdot 86$	=	10.294	29568.7
	⋮			⋮		⋮
108	24	Dez 1995	$8.170 + 0.025 \cdot 108$	=	10.838	50913.5

Prognose mit einfacher linearer Regression



Prognose mit Holt-Glättung

In der Holt-Glättung gibt es zwei Rekursionen: eine exponentielle Glättungen für den Trend im Niveau der Reihe T_t und eine für die Steigung b_t. Dazu gibt es 2 Glättungsparameter: $0 < \alpha, \beta \leq 1$.

$$\begin{aligned}
\text{Niveau:} \quad & T_t &&= \alpha \cdot Y_t + (1-\alpha) \cdot \hat{Y}_t \\
& &&= \alpha \cdot Y_t + (1-\alpha) \cdot (T_{t-1} + b_{t-1}) \\
\text{Steigung:} \quad & b_t &&= \beta \cdot (T_t - T_{t-1}) + (1-\beta) \cdot b_{t-1} \\
\text{Prognose:} \quad & \hat{Y}_{t+1} &&= T_t + b_t \cdot 1, \quad t = 2, \ldots, n
\end{aligned}$$

Wir starten mit den Anfangswerten: $T_2 = Y_2$, $b_2 = T_2 - T_1 = Y_2 - Y_1$.

Prognosewerte über den letzten Beobachtungszeitpunkt n hinaus:

$$\hat{Y}_{n+1} = T_n + b_n, \hat{Y}_{n+2} = T_n + b_n \cdot 2, \ldots, \hat{Y}_{n+r} = T_n + b_n \cdot r$$

Dieses Verfahren modelliert additives Wachstum. Liegt ein exponentielles Wachstum vor, ersetzt man Y_t wieder durch $\log(Y_t)$.

Prognose mit Holt-Glättung

- Die Prognosegleichung für $t = 2, \dots, n$:

$$\hat{Y}_{t+1} = T_t + b_t \cdot 1$$

  Für die 1-Schritt Prognose wird zum prognostizierten Niveau die prognostizierte Veränderung addiert.
- Die Gleichung der Steigung stellt eine exponentielle Glättung der Veränderung der Niveaus $(T_t - T_{t-1})$ dar.

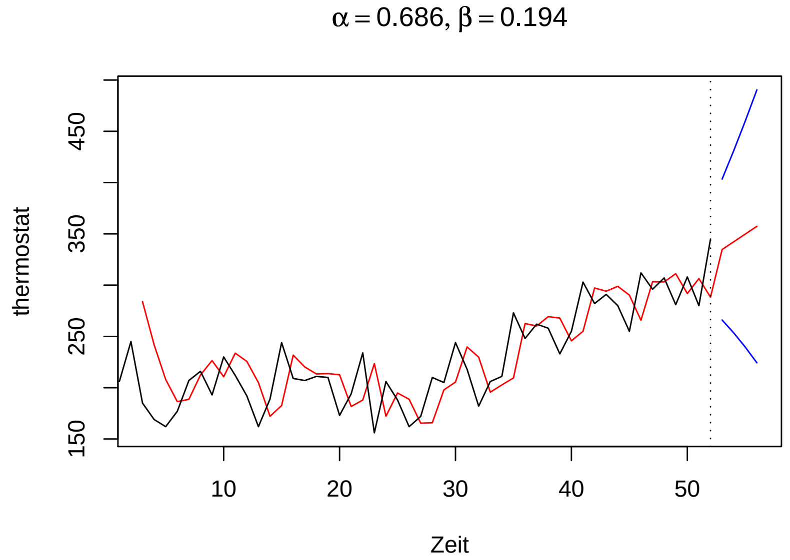
$$b_t = \beta \cdot (T_t - T_{t-1}) + (1 - \beta) \cdot b_{t-1}$$

- Die Niveaugleichung glättet Y_t, wobei $(T_{t-1} + b_{t-1})$ als Glättung für das Niveau in t verwendet wird.

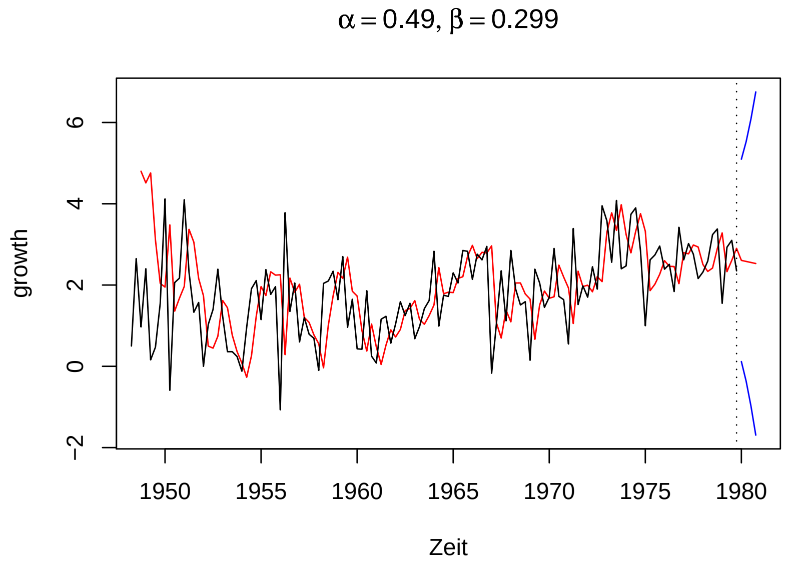
$$T_t = \alpha \cdot Y_t + (1 - \alpha) \cdot (T_{t-1} + b_{t-1})$$

Die optimalen Werte für α und β ermittelt man durch Minimierung der Summe der quadrierten 1-Schritt Prognosefehler, $\sum_t (Y_t - \hat{Y}_t)^2$.

Prognose mit Holt-Glättung



Prognose mit Holt-Glättung



Prognose mit Holt-Glättung

Die Grafiken zeigen wieder Punktprognosen und 95% Prognoseintervalle. Es besteht bei beiden Reihen eine beträchtliche Unsicherheit über den zukünftigen Verlauf.

- `thermostat`: Das Niveau wird zu 68.6% (α) durch den aktuellen Wert bestimmt, der Trend nur zu 19.4% (β). Der Trend wird hier klar fortgesetzt (im Gegensatz zur einfachen exponentiellen Glättung).
- `growth`: Das Niveau wird weniger stark durch aktuelle Werte bestimmt ($\alpha = 49.0\%$), der Trend dafür etwas stärker ($\beta = 29.9\%$). Der Trend verläuft in der Prognoseperiode aber fast horizontal.

Prognose: Zusammenfassung

Die verschiedenen Prognosemethoden sind für Zeitreihen mit unterschiedlichen Mustern geeignet:

Einfache exponentielle Glättung: horizontales Muster, ohne Trend und ohne Saison.

Linearer Trend: stabiler linearer Trend ohne Saison.

Holt-Glättung: Trend ohne Saison. (Verallgemeinerung von einfacher exponentieller Glättung)

Holt-Winters-Glättung: Trend und Saison. (Verallgemeinerung von Holt-Glättung)

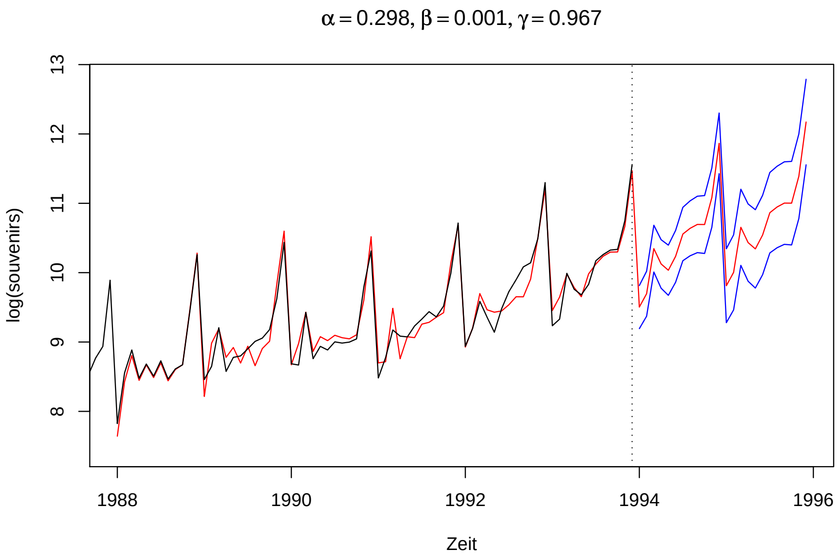
Souvenir-Verkäufe in Maroochydore

Die Holt-Winters-Glättung ist eine weitere Verallgemeinerung der exponentiellen Glättung (bzw. der Holt-Glättung), die zusätzlich zu Niveau (und Steigung) auch die Saison einbeziehen kann.

Formal bedeutet dies, dass es einen dritten Glättungsparameter γ für die Saisongleichung gibt. Weitere Details werden hier nicht diskutiert.

Als Illustration wird das Verfahren für die Reihe log(`souvenirs`) angewendet. Diese zeigt einen etwa linearen Trend und eine deutliche Saison.

Souvenir-Verkäufe in Maroochydore



Zusammenfassung

- Beschreiben von Zeitreihen
- Klassische Zeitreihenzerlegung
- Trend, Saison, irreguläre Komponente
- Trendmodelle, Glättung, Trendbereinigung
- Berechnung der Saisonkomponente
- Prognose, Maße für die Prognosegüte
- Prognose mit exponentieller Glättung und Holt-Glättung