

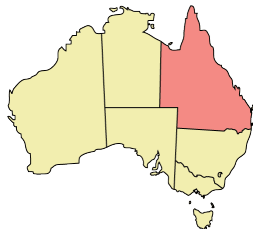
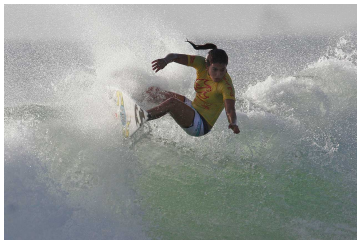
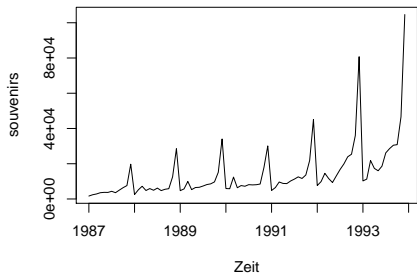
## Kapitel 11

# **Explorative Zeitreihenanalyse**

# Inhalt

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

# Umsätze in Maroochydore, Queensland, AUS



# Glättung und Prognose von Umsätzen

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. Insbesondere 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 1988 bis Ende 1993.

**Fragen:** Wie entwickelt sich der langfristige Trend?

Besitzen die Verkaufszahlen eine regelmäßiges Muster, das sich jedes Jahr wiederholt? Liegt ein saisonales Verhalten vor?

Wie können Trend und Saison zur Prognose der zukünftigen Entwicklung verwendet werden?

# Beispiele für Reihen

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:

- Umsatzentwicklung eines Strandshops in Queensland in AUS Dollar, Monatsdaten, Jänner 1988 - Dezember 1993
- Wöchentliche Thermostatverkäufe in Stück, 52 Wochen
- Einkommenswachstum in Iowa in Prozent, 2. Quartal 1948 – 4. Quartal 1978
- Bierproduktion in Australien in Megaliter, Monatsdaten, Jänner 1991 – August 1995

# Beispiele für Reihen: Quellen

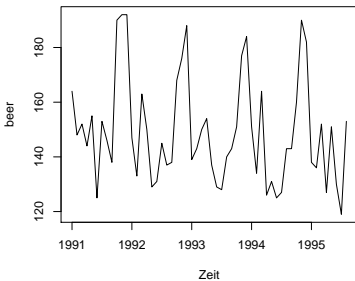
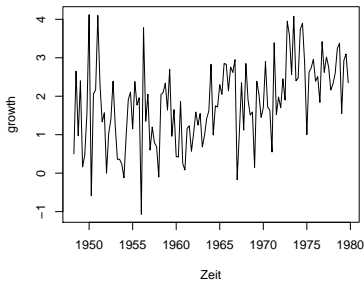
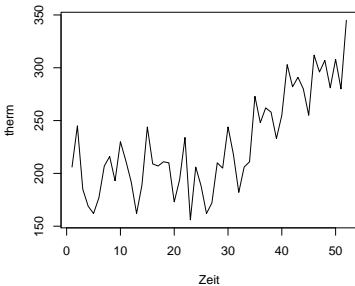
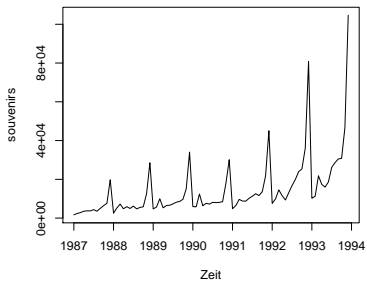
## Quellen:

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

Abraham and Ledolter(1983): *Statistical Methods for Forecasting*, Wiley, Table 3.6, p. 111 und Series 1, p. 419.

Makridakis, Wheelwright and Hyndman(1998), p. 23.

# Zeitreihenplots



# Beschreiben von Zeitreihen

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



# Zeitreihenmuster

- **Horizontales Muster:** Ein horizontales Muster zeigt über die gesamte Beobachtungsperiode ein konstantes Mittel. Zusätzlich verlangen wir, dass die Varianz konstant ist.
- **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 insbesondere 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 Periodenlänge 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.

[ Explorative Zeitreihenanalyse ]

# Die klassische Zeitreihenzerlegung

# Zerlegung einer Reihe in 3 Komponenten

Wir setzen eine Reihe  $Y_t$  aus 3 Komponenten zusammen:

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

Das Modell kann **additiv** oder **multiplikativ** sein:

$$Y_t = T_t + S_t + E_t \quad \text{oder} \quad Y_t = T_t \cdot S_t \cdot E_t$$

- **Additiv**: Der saisonale Effekt ist keine Funktion des Niveaus der Reihe. Vgl. beer Reihe.
- **Multiplikativ**: Der saisonale Effekt wird durch  $T_t$  verstärkt. Vgl. souvenirs Reihe.

Durch Logarithmierung können wir das multiplikative Modell in das additive überführen:  $\log(Y_t) = \log(T_t) + \log(S_t) + \log(E_t)$ .

# Berechnung der Komponenten (additives Modell)

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  $(Y_t - T_t)$ .

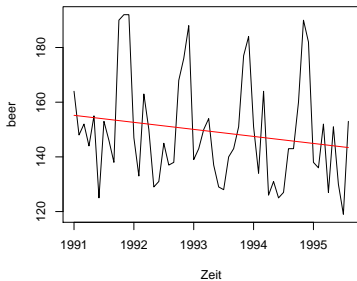
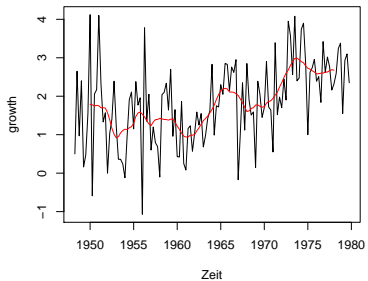
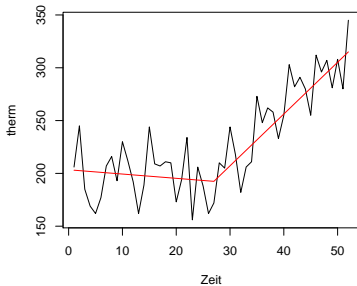
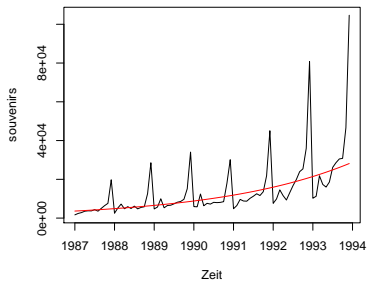
Als Rest ergibt sich die irreguläre Komponente,  $E_t$ .

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

[ Explorative Zeitreihenanalyse ]

# Trend

# Der Trend der Reihen



# Linearer und exponentieller Trend

- Ein **linearer Trend** mit dem Intercept  $\beta_0$ , der Steigung  $\beta_1$  hat die Form

$$Y_t = \beta_0 + \beta_1 \cdot t + \epsilon_t, \quad t = 1, 2, \dots, n$$

$t$  ist die Trendvariable,  $t = 1, 2, 3, \dots$ , und  $\epsilon_t$  die irreguläre Komponente. Das Modell kann mit der Methode der kleinsten Quadrate, KQ, geschätzt werden. (Siehe Kapitel 9)

- **Exponentielles Wachstum** mit der Wachstumsrate  $b$

$$Y_t = a \cdot e^{b \cdot t} \cdot \eta_t$$

kann durch Logarithmieren in eine lineare Funktion übergeführt werden und mit KQ geschätzt werden.  $\beta_0 = \log(a)$ .

$$\log(Y_t) = \beta_0 + b \cdot t + \epsilon_t$$



# Der Trend in der Reihe souvenirs

Wir logarithmieren souvenirs und schätzen einen linearen Trend.

$\log(\text{souvenirs}) \sim t$

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	8.1702	0.11322	72.2	5.69e-76
t	0.0247	0.00231	10.7	3.52e-17

$R^2=0.582$

Das Modell für  $\log(\text{souvenirs})$  lautet

$$\log(\text{souvenirs})_t = 8.17 + 0.025 \cdot t + \epsilon_t$$

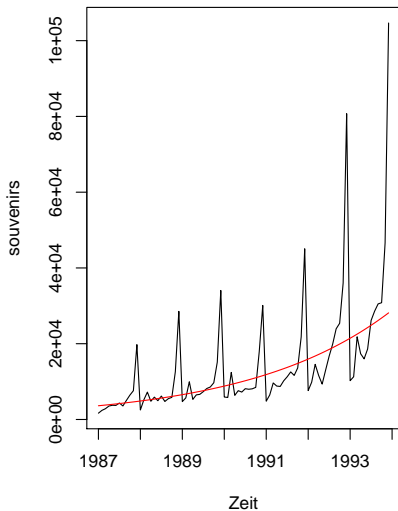
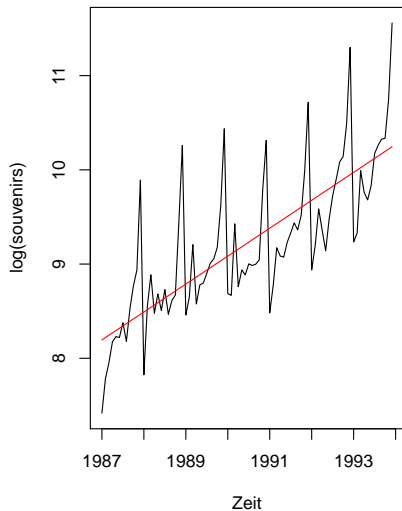
Das Modell für souvenirs lautet nun

$$\text{souvenirs}_t = 3534.1 \cdot e^{0.025 \cdot t} \cdot \eta_t$$

Der Faktor  $a$  ergibt sich aus  $\exp(8.1702) = 3534.1$ .

$R^2$  bezeichnet das multiple Bestimmtheitsmaß.

# Der Trend in der Reihe souvenirs



## Gleitender Durchschnitt, $k$ MA

Ist der Trend nicht stabil, kann ein **gleitender Durchschnitt (moving average, MA)** verwendet werden, um die mittelfristige Entwicklung sichtbar zu machen.

Wir wählen  $k = 2 \cdot m + 1$  (*ungerade*) benachbarte Beobachtungen, berechnen daraus das arithmetische Mittel und ordnen diesen Wert dem mittleren Beobachtungszeitpunkt zu. Dies wiederholen wir für alle Zeitpunkte, für die das möglich ist.

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

$t = m + 1, m + 2, \dots, n - m.$

Wir nennen  $T_t$  den  $(2m + 1)$ -gliedrigen Durchschnitt.

# 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 5 Gewichte  $\{1/8, 1/4, 1/4, 1/4, 1/8\}$  summieren sich zu 1.

- Für Monatsdaten,  $s = 12$ :

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

Die Summe der 13 Gewichte  $\{1/24, 1/12, \dots, 1/12, 1/24\}$  ist 1.

Z.B. für den Juli-Wert gehen die Jännermonate zweier aufeinanderfolgender Jahre mit je 50% in die Berechnung ein.

# Rechenbeispiele

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

- Der 3-gliedrige gleitende Durchschnitt für  $t = 2$  ist

$$(1 + 2 + 1)/3 = 4/3$$

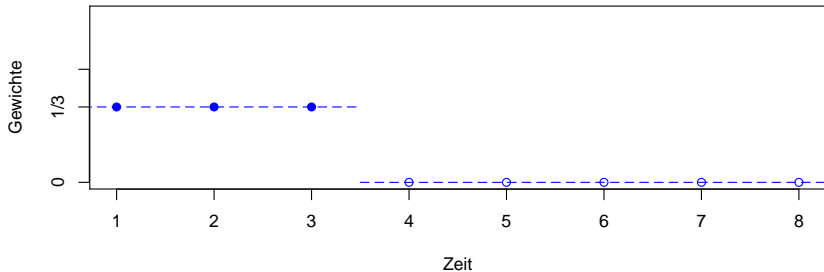
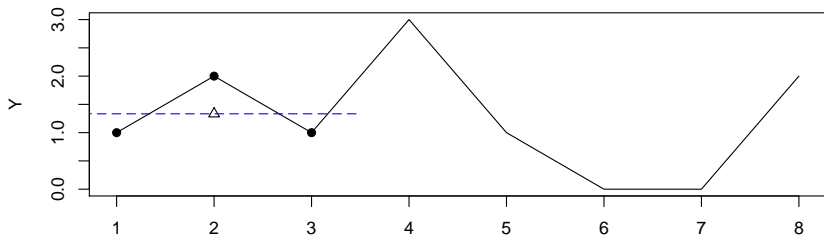
- Der 5-gliedrige gleitende Durchschnitt für  $t = 5$  ist

$$(1 + 3 + 1 + 0 + 0)/5 = 5/5$$

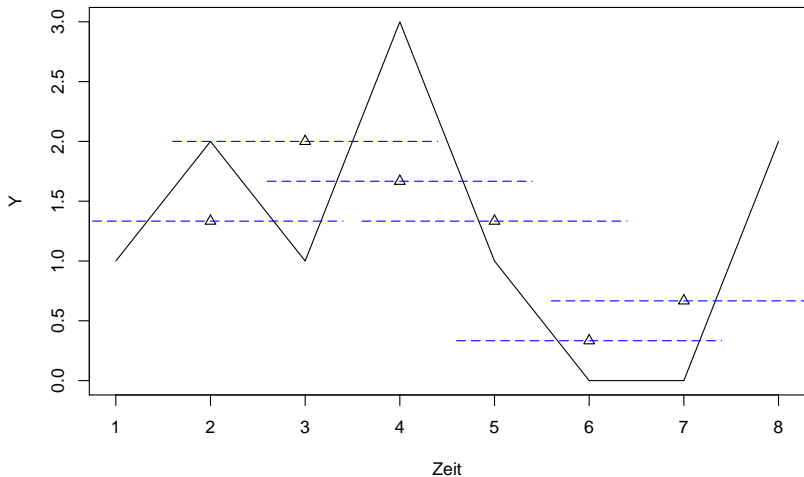
- Nehmen wir an, es liegen Quartalsdaten vor,  $s = 4$ . Der gewichtete moving average für  $t = 6$  ist

$$(0.5 \cdot 3 + 1 + 0 + 0 + 0.5 \cdot 2)/4 = 3.5/4$$

# Graphik: 3-gliedriger gleitender Durchschnitt

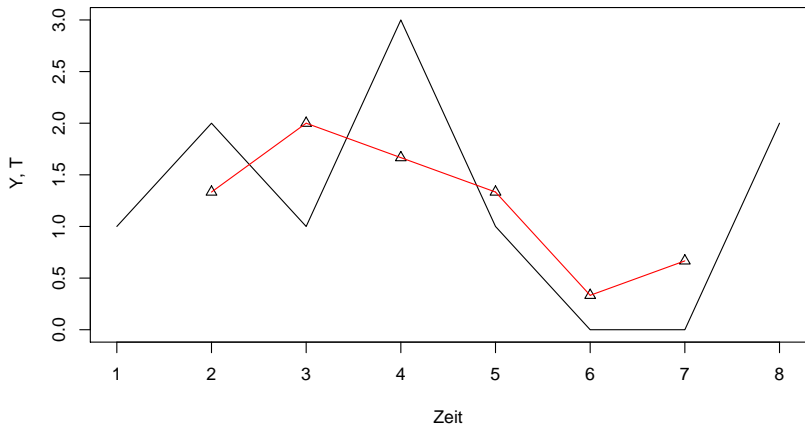


# Graphik: 3-gliedriger gleitender Durchschnitt



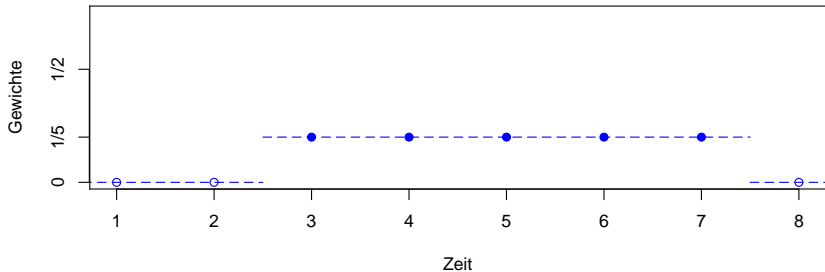
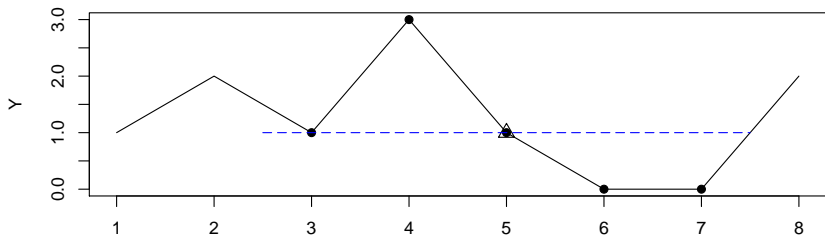
# Graphik: 3-gliedriger gleitender Durchschnitt

Y ... SCHWARZ, T ... ROT

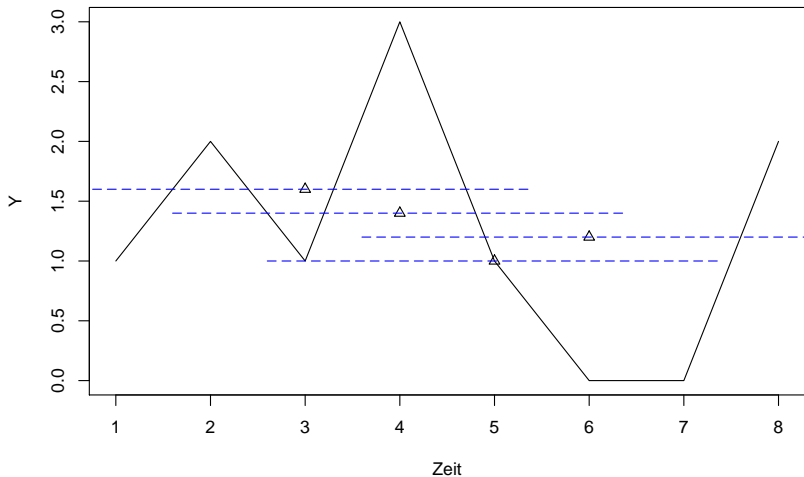




# Graphik: 5-gliedriger gleitender Durchschnitt

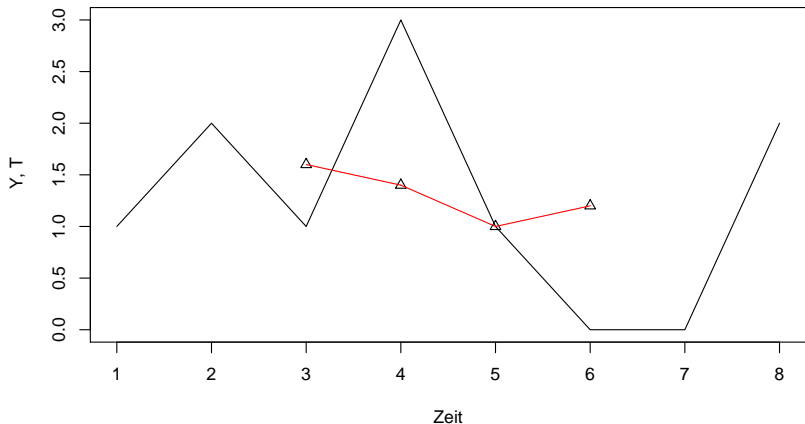


# Graphik: 5-gliedriger gleitender Durchschnitt

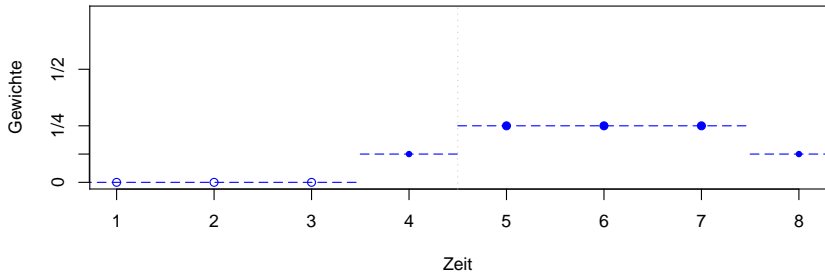
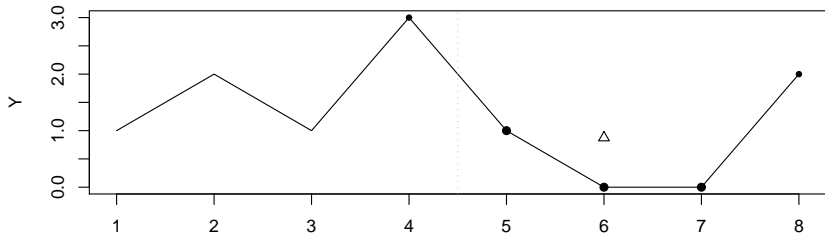


# Graphik: 5-gliedriger gleitender Durchschnitt

Y ... SCHWARZ, T ... ROT

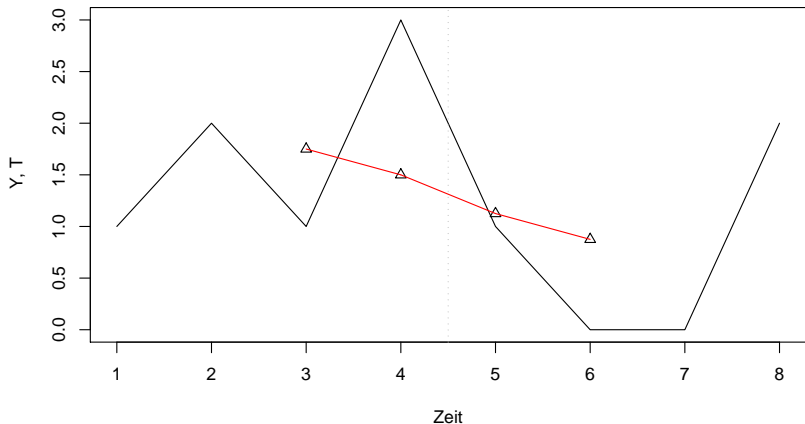


# Graphik: gewichteter gleitender Durchschnitt



# Graphik: gewichteter gleitender Durchschnitt

Y ... SCHWARZ, T ... ROT



# Stärke der Glättung eines $k$ MA

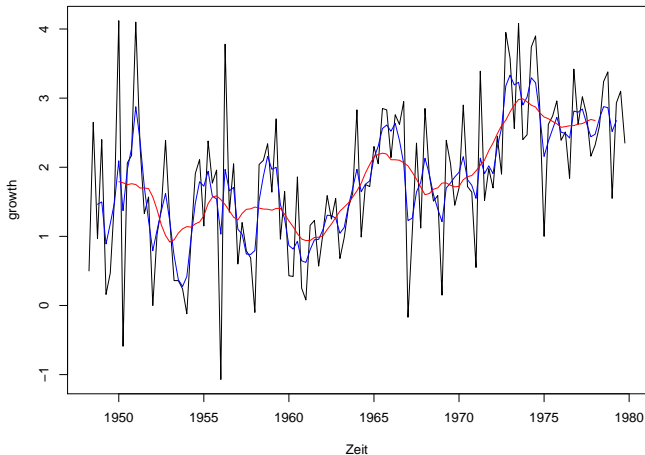
Die Stärke der Glättung eines gleitenden Durchschnitts hängt von der Wahl von  $k$  ab:

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

# Glättung von growth mit gleitendem Durchschnitt

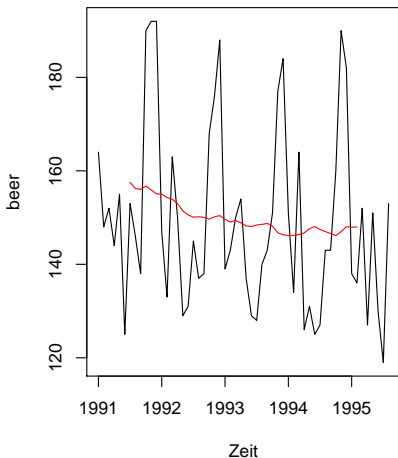
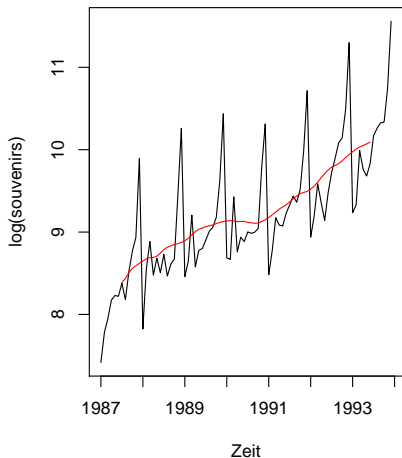
BLAU:  $k = 5 = 2 \cdot 2 + 1$  ( $m = 2$ )

ROT:  $k = 15 = 2 \cdot 7 + 1$  ( $m = 7$ )



# Glättung saisonaler Reihen

Wir glätten  $\log(\text{souvenirs})$  und  $\text{beer}$  mittels des saisonalen gewichteten Durchschnitts,  $s = 12$ .





# Einfache exponentielle Glättung für Reihen ohne Trend und ohne Saison

Die geglättete Reihe heißt nun  $L_t$ , wie level (Niveau). Sie wird rekursiv als gewichtetes Mittel aus der aktuellen Beobachtung und der Glättung der Vorperiode berechnet.

$$L_t = \alpha \cdot Y_t + (1 - \alpha) \cdot L_{t-1}, \quad 0 < \alpha \leq 1$$

- Ist  $\alpha = 1.00$ , ist die geglättete Reihe gleich der ursprünglichen. Es wird nicht geglättet.
- Ist  $\alpha$  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  $\alpha$  klein, z.B. 0.10, so kann der aktuelle  $Y$ -Wert die Glättung nur wenig beeinflussen. Es wird stark geglättet.

# Rechenbeispiel zur einfachen exponentiellen Glättung

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

$$L_t = 0.6 \cdot Y_t + 0.4 \cdot L_{t-1}$$

Der Anfangswert für  $L_1$  sei  $L_1 = Y_1 = 1$ .

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

$$L_2 = 0.6 \cdot Y_2 + 0.4 \cdot L_1 = 0.6 \cdot 2 + 0.4 \cdot 1 = 1.6$$

$$L_3 = 0.6 \cdot Y_3 + 0.4 \cdot L_2 = 0.6 \cdot 1 + 0.4 \cdot 1.6 = 1.24$$

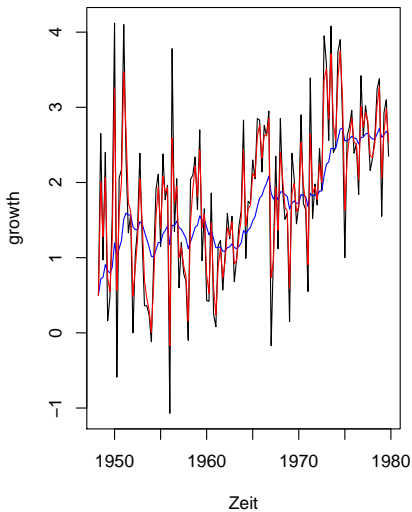
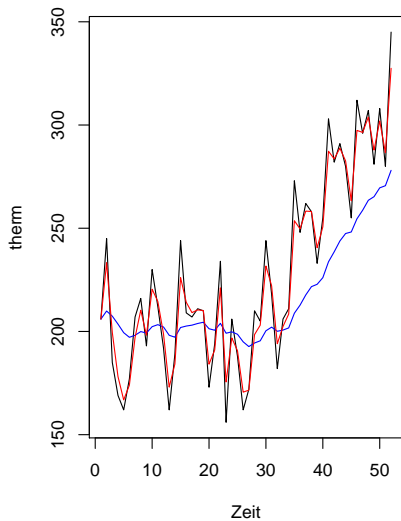
...

$$L_8 = 0.6 \cdot Y_8 + 0.4 \cdot L_7 = 0.6 \cdot 2 + 0.4 \cdot 1.131 = 1.652$$

# Exponentielle Glättung von therm und growth

BLAU:  $\alpha = 0.1$

ROT:  $\alpha = 0.7$



# Exponentielle Glättung von therm und growth

Zu therm:

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 aber gut 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.

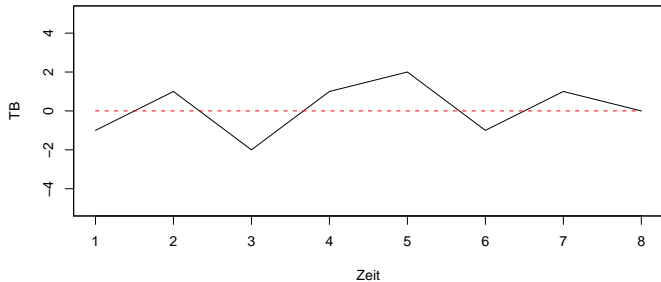
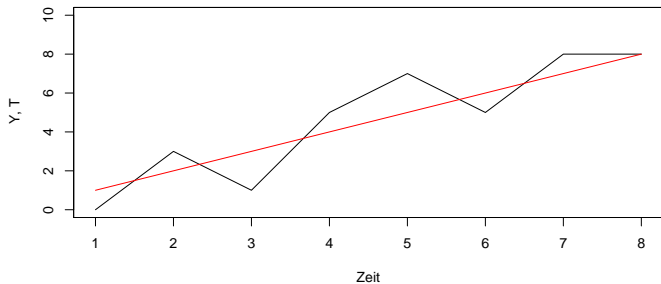
$$TB_t = Y_t - T_t$$

## Rechenbeispiel:

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	-2	1	0

# Trendbereinigung: Graphik



[ Explorative Zeitreihenanalyse ]

# Saison

# Saisonkomponente für Quartalsdaten

Wir berechnen die **Saisonkomponente** sehr einfach. Hier am Beispiel von Quartalsdaten mit  $s = 4$ . Die Zeit  $t$  wird zur besseren Lesbarkeit als

$$t = j.q$$

notiert.  $j$  bezeichnet das Jahr und  $q$  das jeweilige Quartal.

Es wird nun der durchschnittliche saisonale Effekt,  $S_q$ , für das Quartal  $q$  (fix) unter Zugrundelegung der trendbereinigten Reihe  $TB_t = TB_{j.q}$  angegeben.

$$S_q = S_{j.q} = (TB_{1.q} + TB_{2.q} + \dots + TB_{J.q})/J \quad j = 1, \dots, J$$

$J$  bezeichnet die Anzahl der Jahre, für die eine Beobachtung des Quartals  $q$  zur Verfügung steht.

Die Saisonkomponente wird als Durchschnitt über alle Jahre berechnet und gilt in jedem Jahr gleichermaßen. Z.B.:  $S_3$  gilt im 3.Quartal jedes Jahres.



# Saisonkomponente für Monatsdaten

Wir berechnen die **Saisonkomponente** für Monatsdaten analog.

$s = 12$ . Die Zeit  $t$  wird nun als  $t = j.m$  notiert.  $j$  bezeichnet das Jahr,  $m$  den betreffenden Monat.

Der Effekt im Monat  $m = m_0$  (fix) ergibt sich als der durchschnittliche Effekt des Monats  $m$  in der trendbereinigten Reihe,  $TB_t$ , über alle Jahre.

$$S_m = S_{j.m} = (TB_{1.m} + TB_{2.m} + \dots + TB_{J.m})/J \quad j = 1, \dots, J$$

$J$  bezeichnet die Anzahl der Jahre, für die eine Beobachtung des Monats  $m = m_0$  zur Verfügung steht.

Die durchschnittlichen Effekte in jedem Monat  $m = 1, \dots, 12$  wiederholen sich in jedem Jahr.

$$S_t = (S_1, \dots, S_{12}, \quad S_1, \dots, S_{12}, \quad \dots, \quad S_1, \dots, S_{12})$$

# Rechenbeispiel

Gegeben ist die trendbereinigte Quartalsreihe  $TB_t$  der Länge  $n = 12$ .  
Sie besitzt 3 volle Jahre.  $s = 4$ ,  $J = 3$ ,  $q = 1, 2, 3, 4$ .

$t$	1	2	3	4	:	5	6	7	8	:	9	10	11	12
$j$	1	1	1	1	:	2	2	2	2	:	3	3	3	3
$q$	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

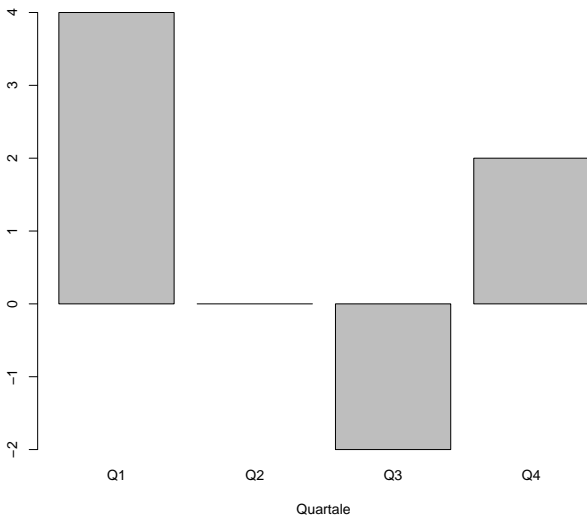
$$S_1 = (2 + 4 + 6)/3 = 4 \qquad S_2 = (1 + 1 - 2)/3 = 0$$

$$S_3 = (-3 - 1 - 2)/3 = -2 \qquad S_4 = (3 + 2 + 1)/3 = 2$$

Die Saisonkomponente ist

$$S_t = S_{j.q} = S_q$$

# Bsp.: Durchschnittliche saisonale Effekte, $S_q$

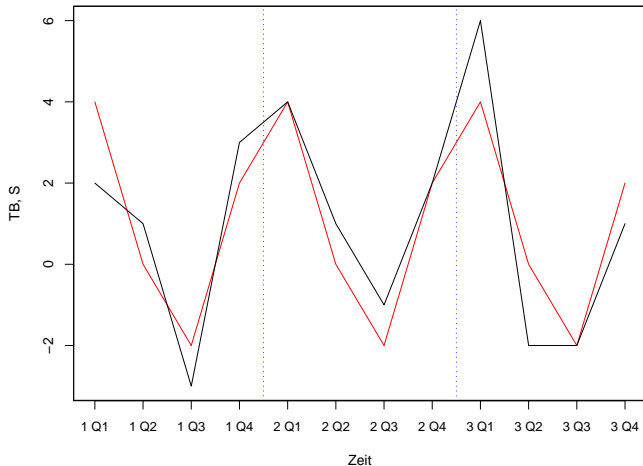


# Bsp.: Graphik der Saisonkomponente, $S_t = S_{j.q}$

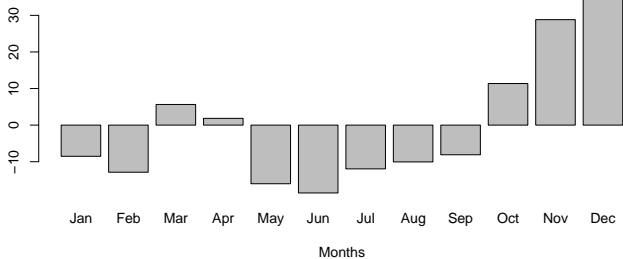
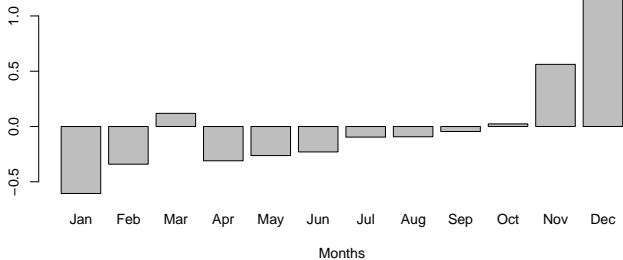
SCHWARZ ...  $TB_t$

ROT ...  $S_t$

Graphik der Saisonkomponente S



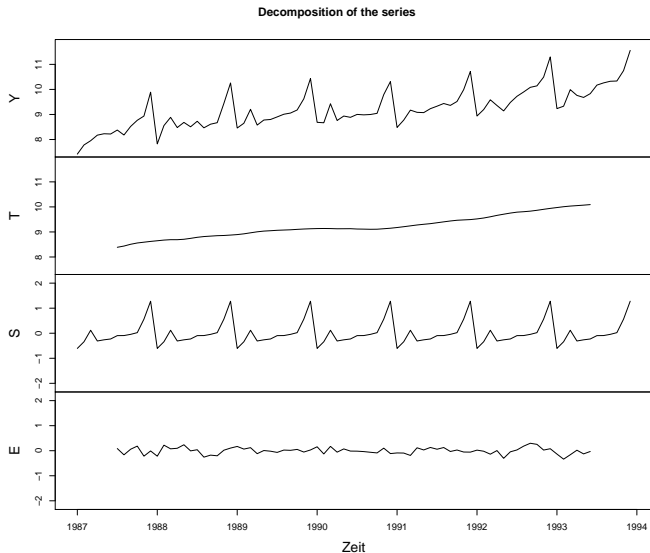
# Saisonmuster von $\log(\text{souvenirs})$ und beer, $S_m$



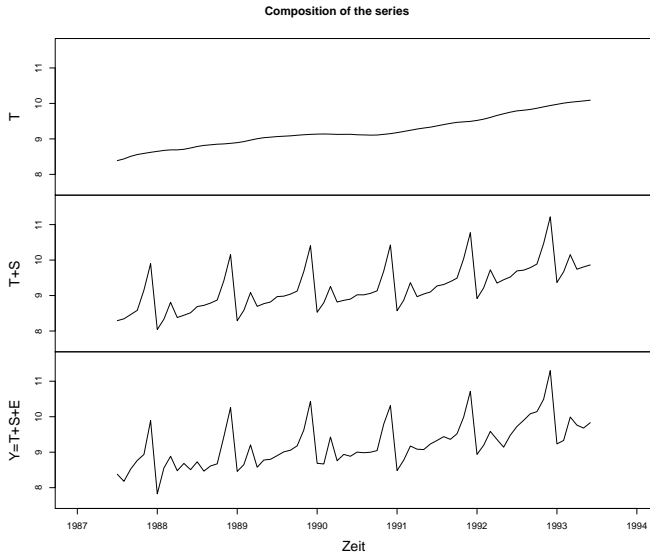
[ Explorative Zeitreihenanalyse ]

# **Zusammenfassung der Komponentenzerlegung**

# Komponentenzerlegung von $\log(\text{souvenirs})$

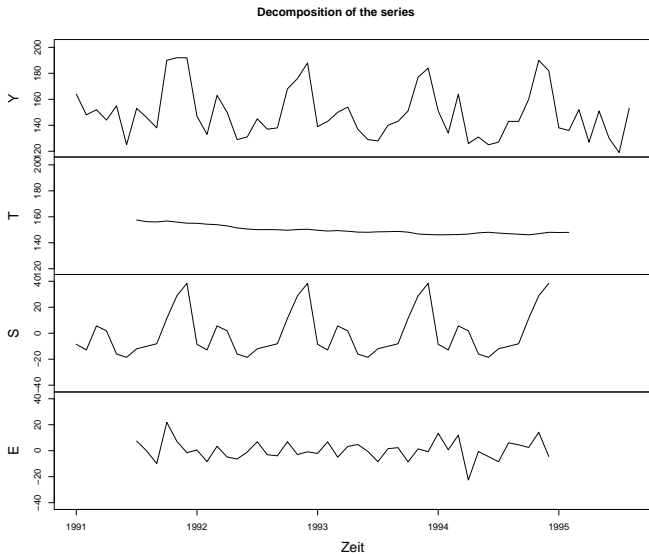


# Zusammensetzung von $\log(\text{souvenir})$

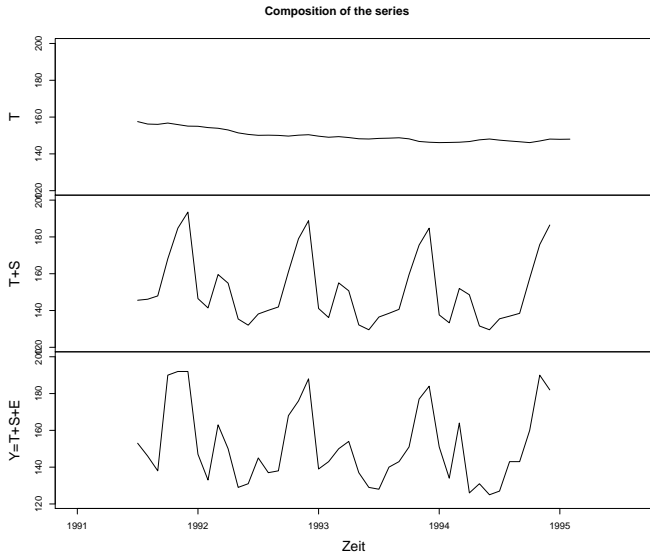




# Komponentenzerlegung von beer



# Zusammensetzung von beer



[ Explorative Zeitreihenanalyse ]

# Prognose

# Beobachtungen und Vorhersagen

Beobachtungsperiode						:	Prognoseperiode		
$Y_1$	$Y_2$	$Y_3$	$\dots$	$Y_{n-1}$	$Y_n$	:	$Y_{n+1}$	$\dots\dots$	$Y_{n+r}$
historische Daten						:	zukünftige Beobachtungen		

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

Bezeichnungen:

$Y_n(r)$  bezeichnet die Vorhersage für die zukünftige Beobachtung  $Y_{n+r}$ ,  $r$ -Schritt Prognose

$n$  gibt den Vorhersagezeitpunkt an,

$r$  den Vorhersagehorizont.

# Beobachtungen und Vorhersagen

$Y_n(r)$ : Ist  $r$  klein sprechen wir von einer **kurzfristigen**, sonst von einer mittel- oder **langfristigen** Prognose.

$Y_{2008/9}(3)$  Vorhersage für 2008/12 (Dez 2008) auf Basis der Beobachtungen bis einschließlich 2008/9, eine 3-Schritt Prognose.

Vorhersagen für Okt 2008:

$Y_{2008/6}(4)$  Vorhersage für 2008/10 (Okt 2008) auf Basis der Beobachtungen bis einschließlich 2008/6, eine 4-Schritt Prognose.

$Y_{2008/9}(1)$  Vorhersage für 2008/10 (Okt 2008) auf Basis der Beobachtungen bis einschließlich 2008/9, eine 1-Schritt Prognose.

# Prognosefehler

Die Differenz zwischen tatsächlicher, zukünftiger Beobachtung und der Vorhersage bezeichnen wir als Prognosefehler.

$$r\text{-Schritt Prognosefehler} = Y_{n+r} - Y_n(r)$$

Zum Bewerten der Prognosen  $Y_n(j)$ ,  $j = 1, \dots, r$  einer Reihe eignet sich der **mean square error**, MSE,

$$MSE = \frac{1}{r} \sum_{j=1}^r [Y_{n+j} - Y_n(j)]^2$$

oder der **root mean square error**, RMSE,

$$RMSE = \sqrt{MSE}$$

# Rechenbeispiel zum MSE und RMSE

Angenommen wir verwenden 2 verschiedene Verfahren, A und B, um eine Reihe  $Y_t$ ,  $t = 1, \dots, n$ , vorherzusagen und notieren uns die Vorhersagewerte. Nach 3 Perioden (und 3 weiteren Beobachtungen) wollen wir feststellen, welche Methode die besseren Vorhersagen geliefert hat. Wir verwenden dazu den RMSE.  $r = 3$ .

$t$	1	...	$n-1$	$n : n+1$	$n+2$	$n+3$
$Y_t$	5	...	-1	4 : 3	2	3
$j$				: 1	2	3
Verf. A $Y_n(j)$				: 3	4	2
Verf. B $Y_n(j)$				: 2	3	3

$$RMSE_A = \sqrt{(3-3)^2 + (2-4)^2 + (3-2)^2} = 2.236$$

$$RMSE_B = \sqrt{(3-2)^2 + (2-3)^2 + (3-3)^2} = 1.414$$

Verfahren B hat den kleineren RMSE und ist daher vorzuziehen.

# Prognose von Reihen ohne Trend und ohne Saison: Einfache exponentielle Glättung

Wir verwenden nun das Verfahren der einfachen exponentiellen Glättung zur Prognose einer Reihe mit dem Glättungsparameter  $\alpha$ ,  $0 < \alpha \leq 1$ .

Das Modell lautet:

$$\text{Niveau: } L_t = \alpha \cdot Y_t + (1 - \alpha) \cdot L_{t-1}$$

$$\text{Prognose: } Y_t(r) = L_t$$

mit dem Anfangswert:  $L_1 = Y_1$ .

Die 1-Schritt Prognose, wie auch alle mehr-Schritt Prognosen sind gleich dem letzten geglätteten Wert.  $r = 1, 2, \dots$

$$Y_t(1) = L_t, \quad Y_t(2) = L_t, \quad \dots$$

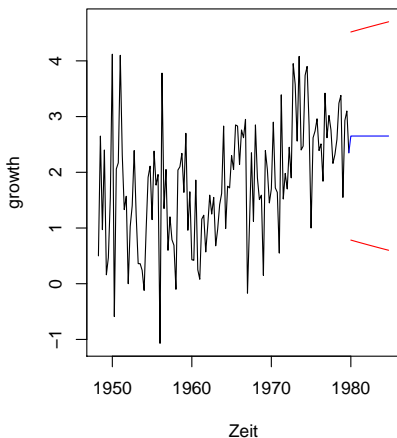
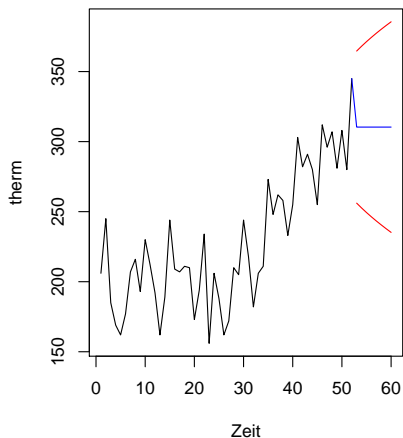
Den optimalen Wert für  $\alpha$  ermittelt man durch Minimierung der Summe der quadrierten 1-Schritt Prognosefehler,  $\sum_t [Y_t - Y_{t-1}(1)]^2$ .



# Prognose mit einf. exp. Glättung: therm und growth

$\alpha=0.362$

$\alpha=0.104$



BLAU ... Punktprognosen

ROT ... 95% Prognoseintervalle

# Prognose mit einf. exp. Glättung: Kommentar

Der Graph der Reihen wird mit den Punktprognosen  $Y_n(1), Y_n(2), \dots, Y_n(8)$  bzw.  $Y_n(20)$  fortgesetzt.

Die roten Linien geben die 95% Prognoseintervall-Grenzen an. Sie zeigen, dass bei beiden Reihen eine beträchtliche Unsicherheit über den zukünftigen Verlauf besteht.

- `therm`: Das Niveau wird zu 36.2% ( $\alpha$ ) durch den aktuellen Wert bestimmt. Die Graphik zeigt, dass die Prognose einen konstanten Verlauf fortschreibt. Der Anstieg in der zweiten Beobachtungshälfte wird dabei ignoriert. Für die Prognose geeigneter wird sich das Verfahren von Holt erweisen.
- `growth`: Für die Reihe `growth` spielt das Niveau vergangener Beobachtungen eine große Rolle ( $(1 - \alpha) = 0.896$ ). Auch hier scheint das Verfahren von Holt (siehe unten) vorzuziehen zu sein.

## Prognose bei stabilem Trend ohne Saison

Liegt eine Reihe mit einem stabilen Trend vor, können wir die Trend-Regression für mittel- bis langfristige Prognosen verwenden. Für  $\log(\text{souv}_t)$  haben wir die Gleichung (ohne saisonale Effekte zu berücksichtigen) bereits geschätzt.

$$\log(\text{souv}_t) = 8.1702 + 0.0247 \cdot t + \hat{\epsilon}_t, \quad t = 1, \dots, 84$$

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

$$\widehat{\log(\text{souv}_t)} = 8.1702 + 0.0247 \cdot t$$

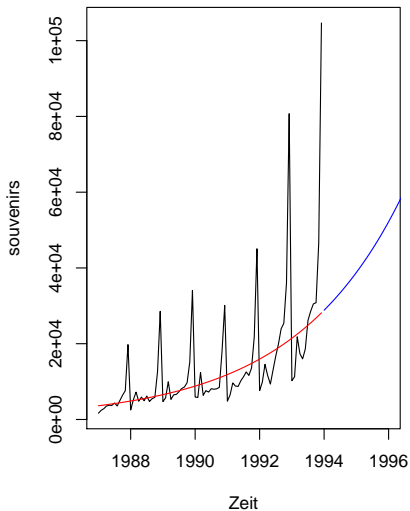
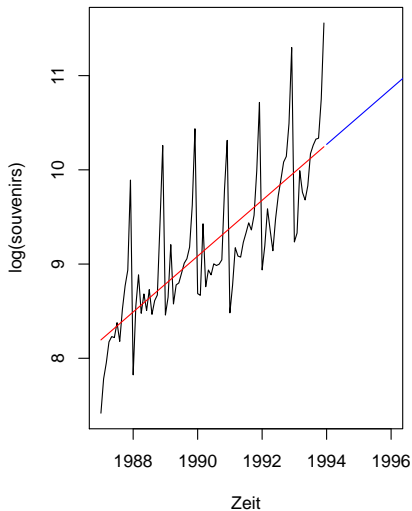
in die wir für  $t = 85, 86, \dots, 120$  einsetzen.

# Prognose bei stabilem Trend ohne Saison

Die Vorhersagen für die Perioden 85, ..., 120 sind:

$t$	$j$	Periode		$\log(\text{souv}_{84})(j)$	$\text{souv}_{84}(j)$
85	1	Jan 94	$8.1702 + 0.0247 \cdot (85)$	10.2697	28845.2
86	2	Feb 94	$8.1702 + 0.0247 \cdot (86)$	10.2944	29566.6
		...	...	...	...
97	13	Jan 95	$8.1702 + 0.0247 \cdot (97)$	10.6551	38797.9
		...	...	...	...
120	38	Dez 96	$8.1702 + 0.0247 \cdot (120)$	11.1342	68473.4

# Prognose mit stabilem Trend: souvenirs



# Prognose für Trend ohne Saison: Das Modell von Holt

Das Verfahren von Holt verwendet exponentielle Glättungen für das Niveau der Reihe  $L_t$  und für die 1-Perioden Veränderung im Niveau,  $b_t$ . Dazu gibt es 2 Glättungsparameter:  $0 < \alpha, \beta \leq 1$ .

Das Modell lautet:

$$\text{Niveau: } L_t = \alpha \cdot Y_t + (1 - \alpha) \cdot (L_{t-1} + b_{t-1})$$

$$\text{Trend: } b_t = \beta \cdot (L_t - L_{t-1}) + (1 - \beta) \cdot b_{t-1}$$

$$\text{Prognose: } Y_t(r) = L_t + b_t \cdot r$$

mit den Anfangswerten:  $L_2 = Y_2$  und  $b_2 = L_2 - L_1 = Y_2 - Y_1$ .

Im Gegensatz zur einfachen Glättung verwendet man hier die geglätteten Werte,  $L_t$  und  $b_t$ , in der Prognose für die folgende Periode.

Dieses Verfahren modelliert additives Wachstum. Liegt ein exponentielles Wachstum vor, ersetzt man  $Y_t$  durch die logarithmierte Reihe.

# Holt: Die Gleichungen

- Die Prognosegleichung für  $r = 1$ :

$$Y_t(1) = L_t + b_t \cdot 1$$

Für die 1-Schritt Prognose wird zum prognostizierten Niveau die prognostizierte Veränderung addiert.

- Die Trendgleichung stellt eine exponentielle Glättung der Veränderung der Niveaus ( $L_t - L_{t-1}$ ) dar.

$$b_t = \beta \cdot (L_t - L_{t-1}) + (1 - \beta) \cdot b_{t-1}$$

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

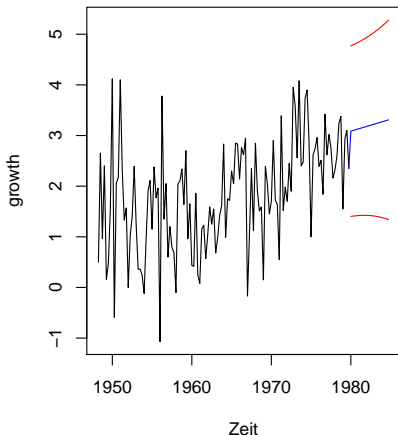
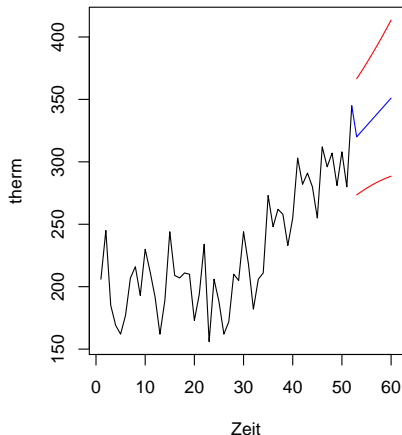
$$L_t = \alpha \cdot Y_t + (1 - \alpha) \cdot (L_{t-1} + b_{t-1})$$

Die optimalen Werte für  $\alpha$  und  $\beta$  ermittelt man durch Minimierung der Summe der quadrierten 1-Schritt Prognosefehler,  $\sum_t [Y_t - Y_{t-1}(1)]^2$ .

# Prognose mit Holt: therm und growth

$\alpha=0.245, \beta=0.093$

$\alpha=0.078, \beta=0.074$



BLAU ... Punktprognosen

ROT ... 95% Prognoseintervalle



## Prognose mit Holt: Kommentar

Der Graph der Reihen wird mit den Punktprognosen  $Y_n(1), Y_n(2), \dots, Y_n(8)$  bzw.  $Y_n(20)$  fortgesetzt.

Die roten Linien geben die 95% Prognoseintervall-Grenzen an. Sie zeigen, dass bei beiden Reihen eine beträchtliche Unsicherheit über den zukünftigen Verlauf besteht.

- `therm`: Das Niveau wird zu 24.5% ( $\alpha$ ) durch den aktuellen Wert bestimmt, der Trend nur zu 9.3% ( $\beta$ ). Die Graphik oben zeigt, dass die Prognose im Wesentlichen den Trend der zweiten Hälfte der Beobachtungsperiode berücksichtigt.
- `growth`: Für die Reihe `growth` spielt das Niveau vergangener Beobachtungen eine große Rolle ( $(1 - \alpha) = 0.922$ ). Die Gewichtung der aktuellen lokalen Veränderungen (Trend) ist ähnlich gering wie bei der Reihe `therm` ( $\beta = 0.074$ ).

Der Trend beider Reihen wird durch das Holt Verfahren besser fortgeschrieben als durch die einfache exponentielle Glättung.

# Zusammenfassung

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