Gliederung der Vorlesung

1) Einführung und Grundlagen
2) Klassische Zeitreihenanalyse
    a) Trendbestimmung
    b) Glättung von Zeitreihen durch Filter (lokale Trendbestimmung)
    c) Saisonbereinigung
    d) Exponentielles Glätten, Prognosen
3) Modellierung von Reihen durch stochastische Prozesse
    a) Autoregressive (AR) Modelle
    b) Moving Average (MA) Modelle
    c) ARMA Modelle
    d) ARIMA Modelle

Das klassische Komponentenmodell

Komponentenmodelle für Zeitreihen gehen aus von einer Zerlegung der Form

$$x_t = m_t + s_t + u_t \qquad \text{(additives Modell)}$$

$$x_t = m_t \cdot s_t + u_t \qquad \text{(quasimultiplikatives Modell)}$$

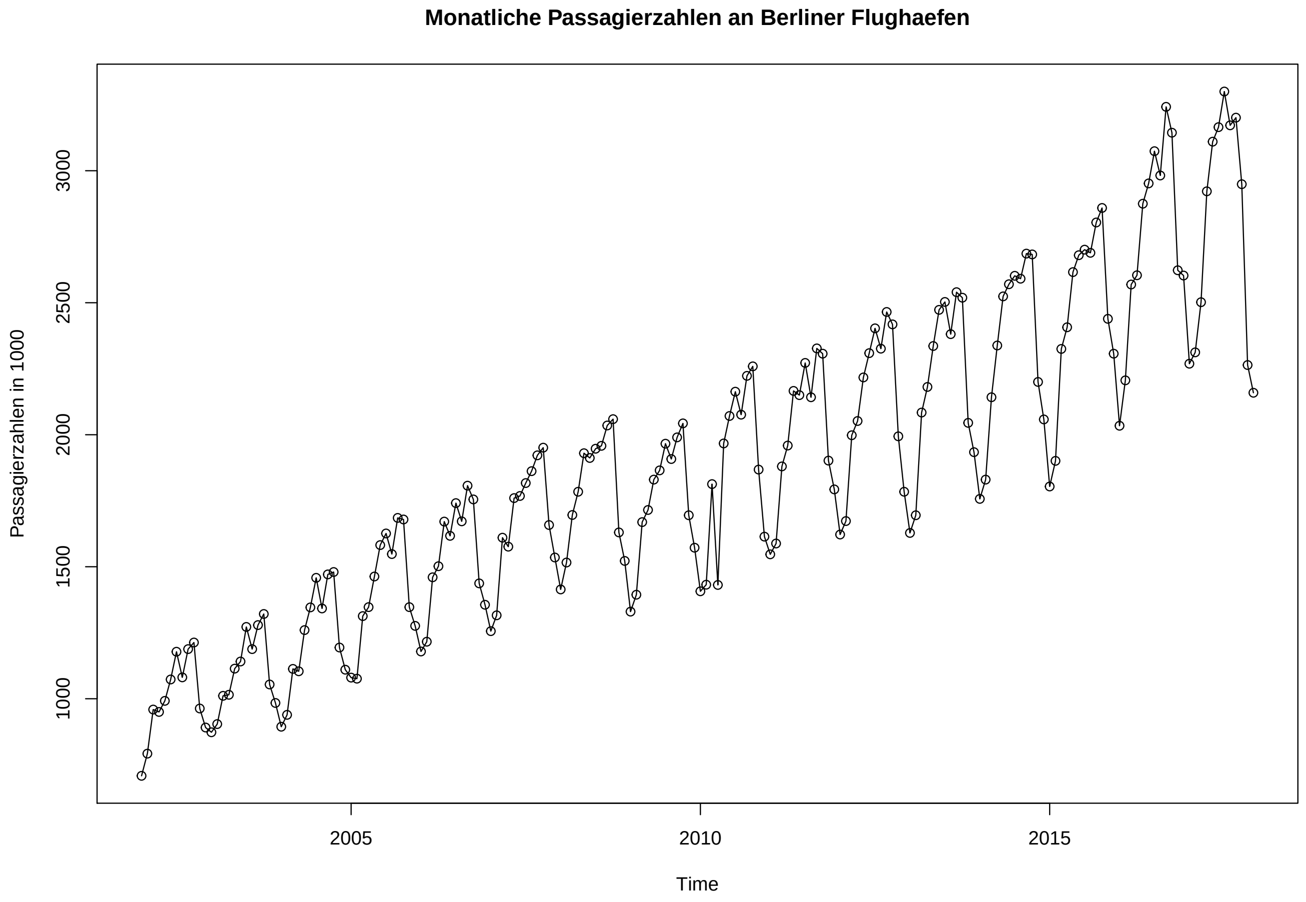
$$x_t = m_t \cdot s_t \cdot u_t \qquad \text{(multiplikatives Modell)}$$

Dabei bezeichnen

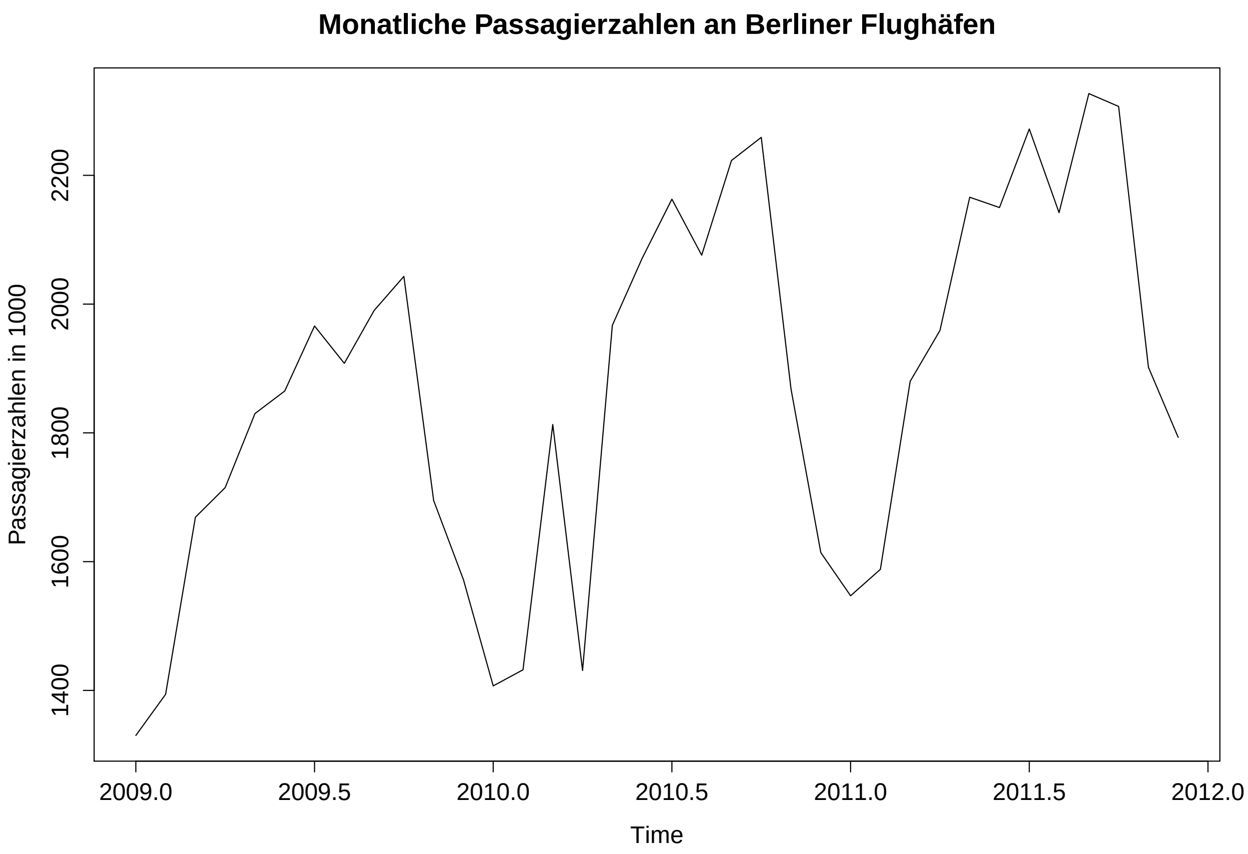
- (m_t) den Trend: eine langfristige systemische Veränderung des mittleren Niveaus der Zeitreihe
- (s_t) die Saison: eine regelmäßige zyklische Schwankung mit Periode p, d.h. $s_{t+p} = s_t$ und $\sum_{k=1}^{p} s_k = 0$.
- (u_t) den Rest: nicht zu erklärende Einflüsse oder Störungen

Manchmal noch Konjunkturkomponente, Kalenderkomponente...

Wir schauen uns an:

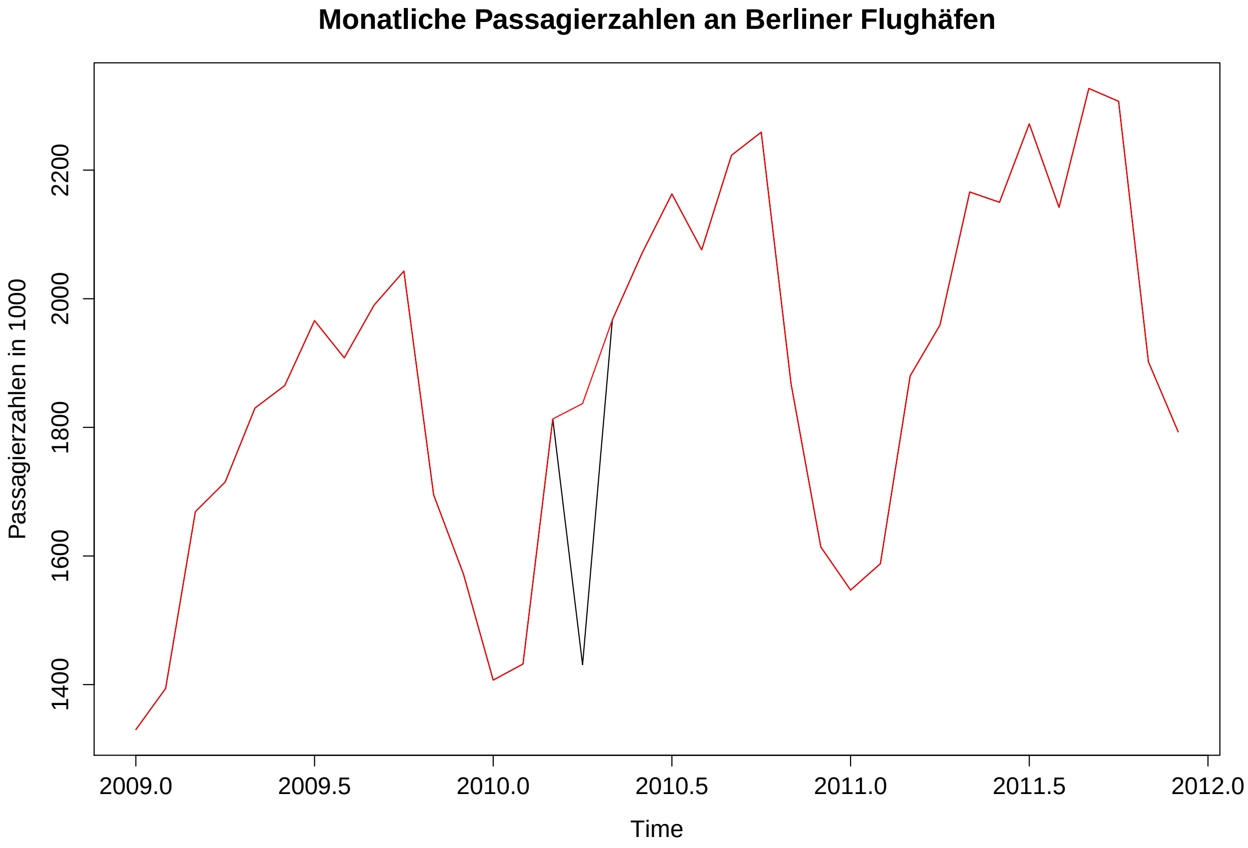


Wir schauen uns an:



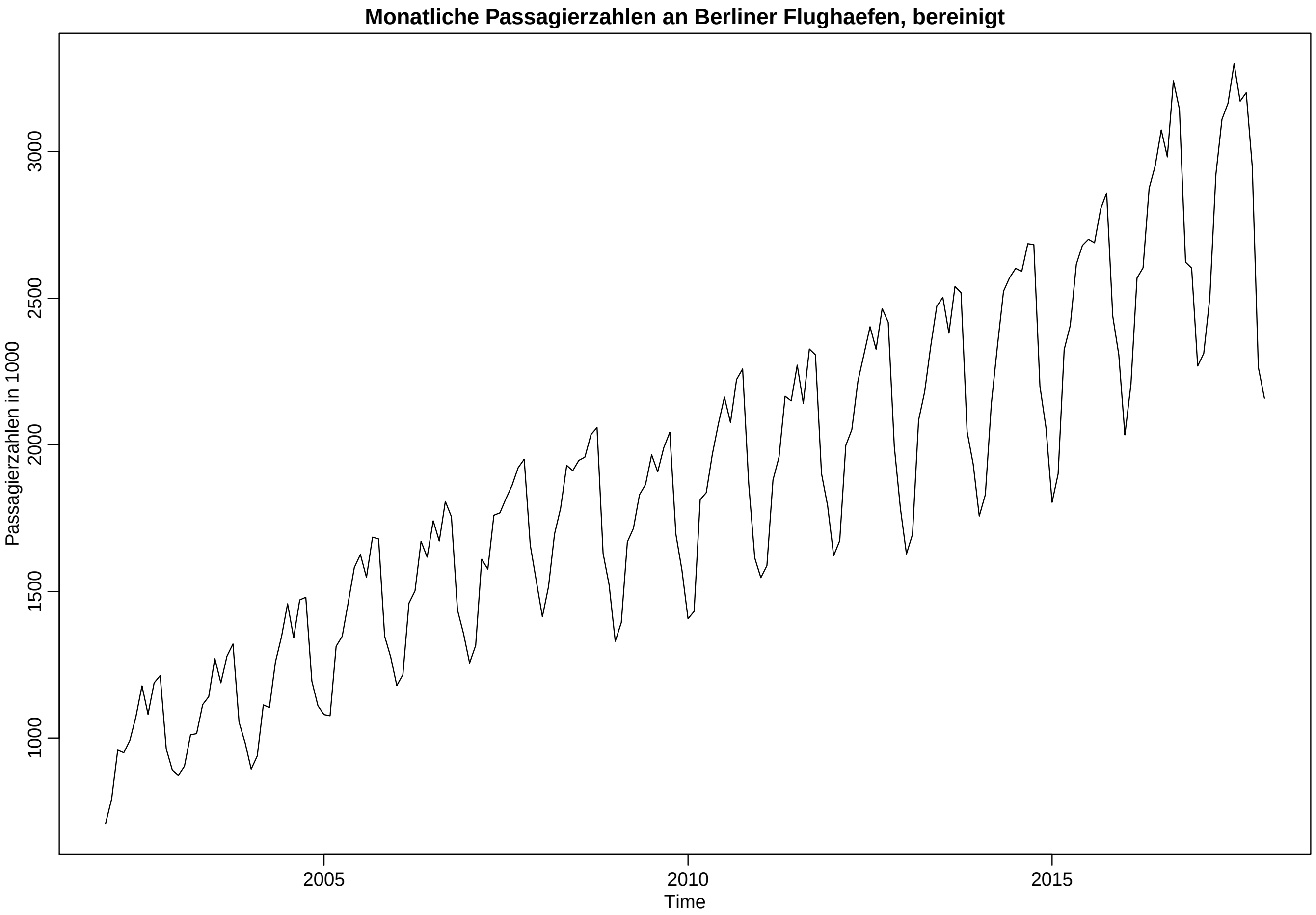
→ Ausreißer! April 2010, der Vulkan... Wir ersetzen den Wert durch den Mittelwert aus April 2009 und April 2011.

Wir schauen uns an:

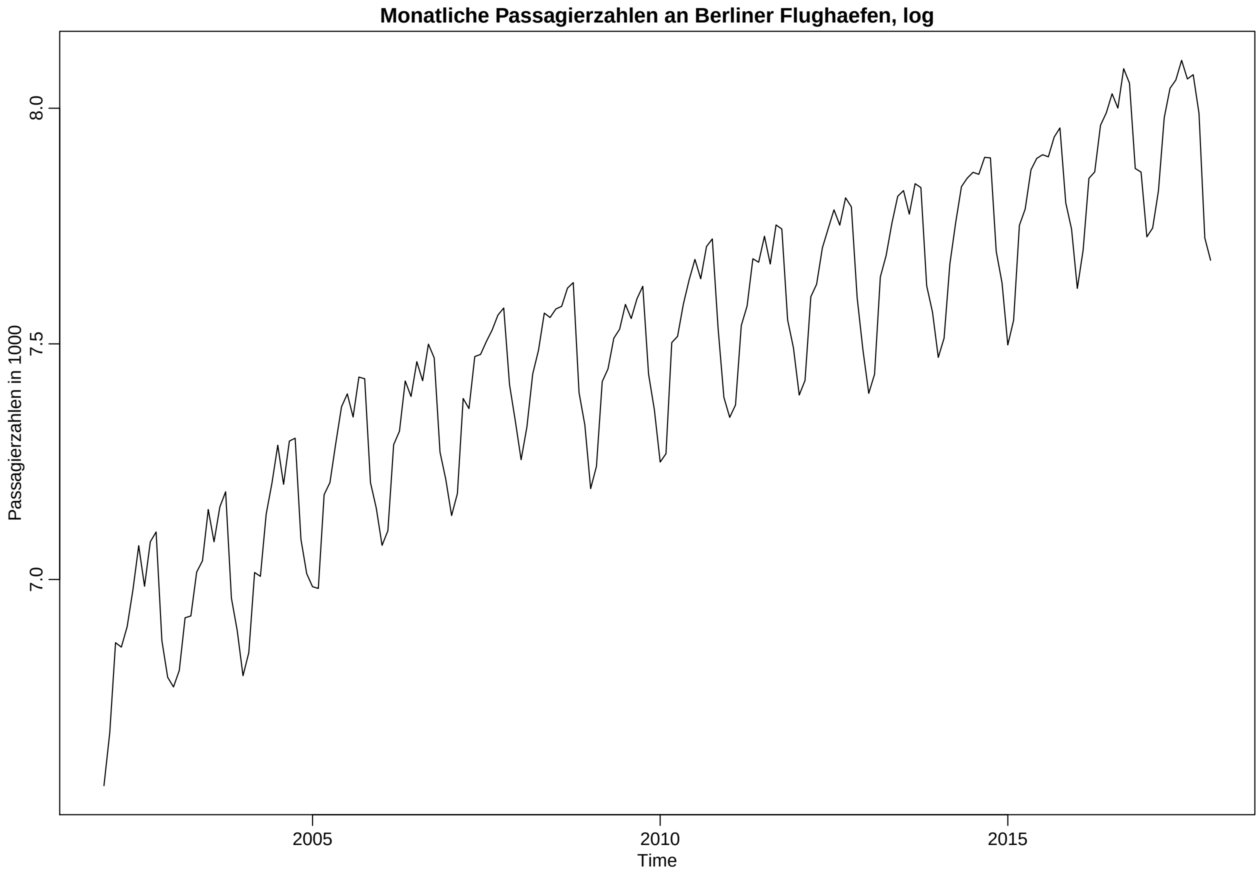


→ Ausreißer! April 2010, der Vulkan... Wir ersetzen den Wert durch den Mittelwert aus April 2009 und April 2011.

Wir schauen uns an:



Wir schauen uns an:

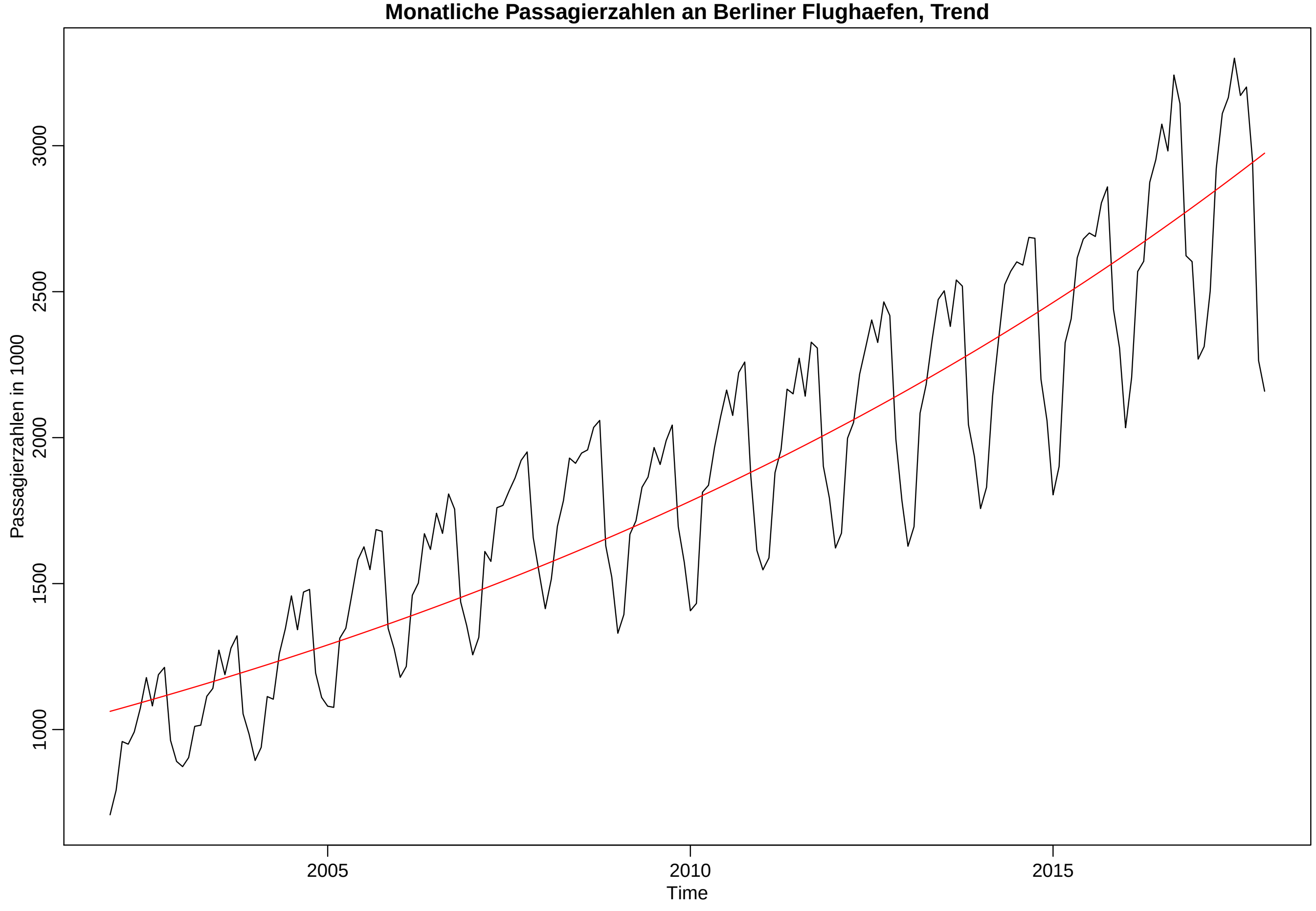


$\rightarrow$ Schwankungen wachsen mit der Zeit $\rightarrow$ Log nehmen!
Dann sind wir bei einem additiven Modell.

Saisonbereinigung: Mehrere Möglichkeiten

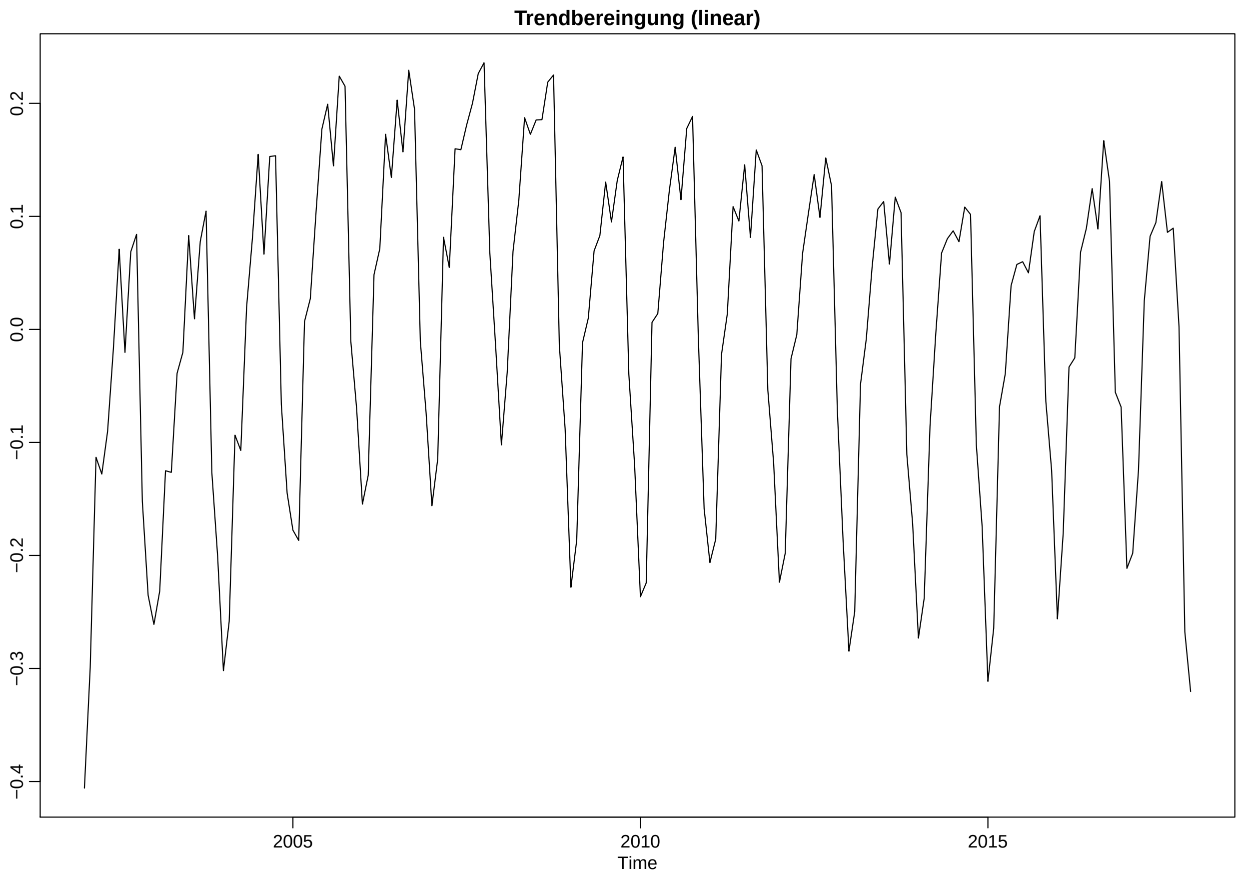
- Phasendurchschnittsverfahren
  1) Trend bestimmen und bereinigen (z.B. mit polynomialer Regression oder gleitenden Durchschnitten)
  2) Saisonkomponente als Durchschnitt der Phase über die Jahre
  3) Das Gesamtmodell aus Trend und Saison wieder zusammensetzten
- Regression mittels Saison-Dummies (eine Regression für beides, Trend und Saison)
- Regression mit trigonometrischen Polynomen
- Differenzenbildung (haben wir bereits gesehen)
- Viele etablierte Verfahren als Mischung aus diesen Methoden (und evt. noch anderen).

Phasendurchschnittsverfahren: Trend



Phasendurchschnittsverfahren: Saison

Saisonresiduen = Data - Trend



Phasendurchschnittsverfahren: Saison

$$r_k^j = \text{Saisonresiduen} = \text{Data - Trend}$$

- Über jeden Monat (Phase) $j = 1, \ldots, 12$ über $k = \frac{N}{12}$ Jahre mitteln:

$$M_j = \frac{1}{k} \sum_{i=1}^{k} r_i^j$$

- Die gemittelten Werte zentrieren:

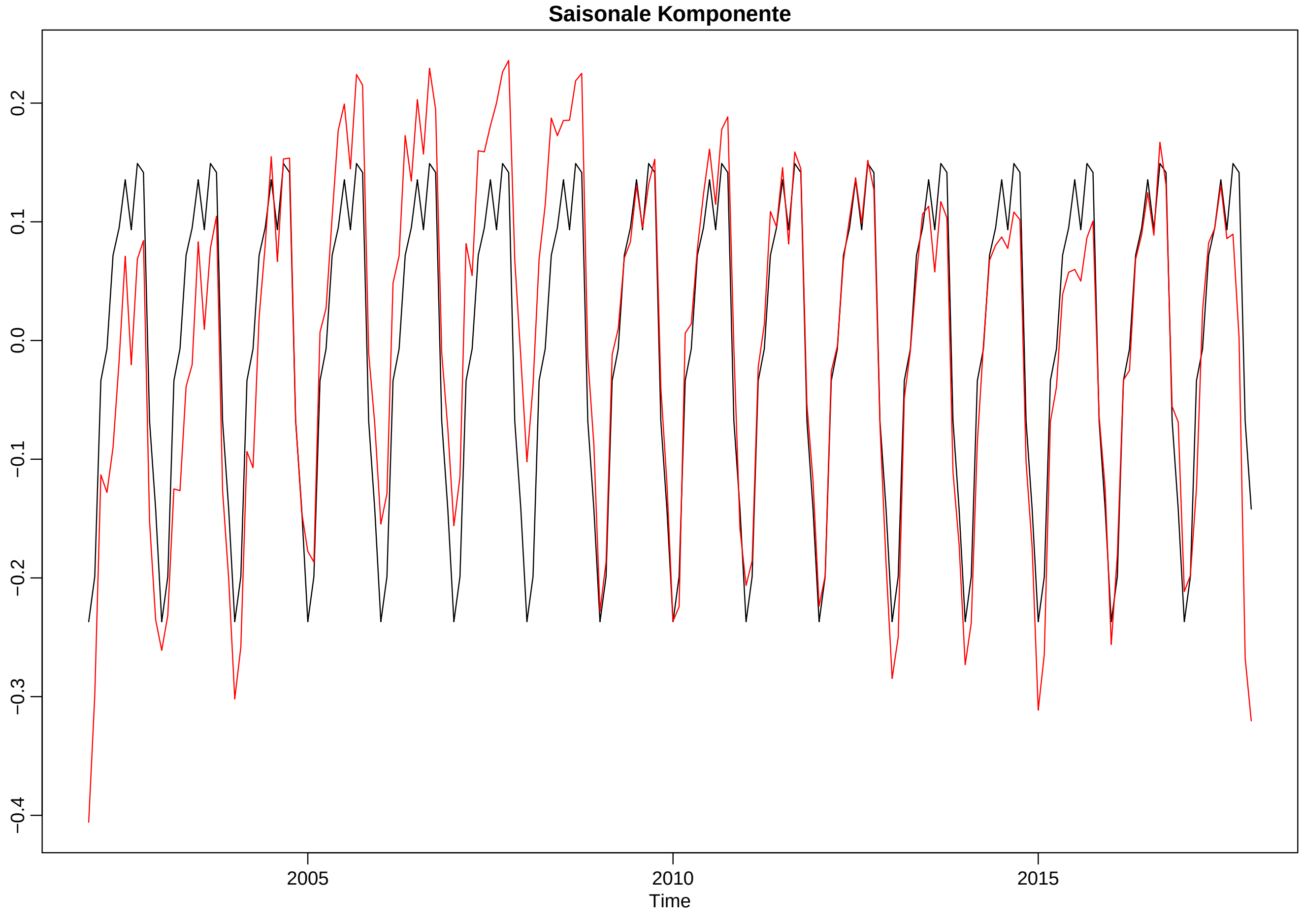
$$\bar{d} = \frac{1}{12} \left(M_1 + \cdots + M_{12} \right), \qquad S_j = M_j - \bar{d}, \qquad j = 1, \ldots, 12$$

- So erhalten wir 12 Saisonkomponenten $S_1, \ldots, S_{12}$ mit

$$S_1 + \cdots + S_{12} = 0,$$

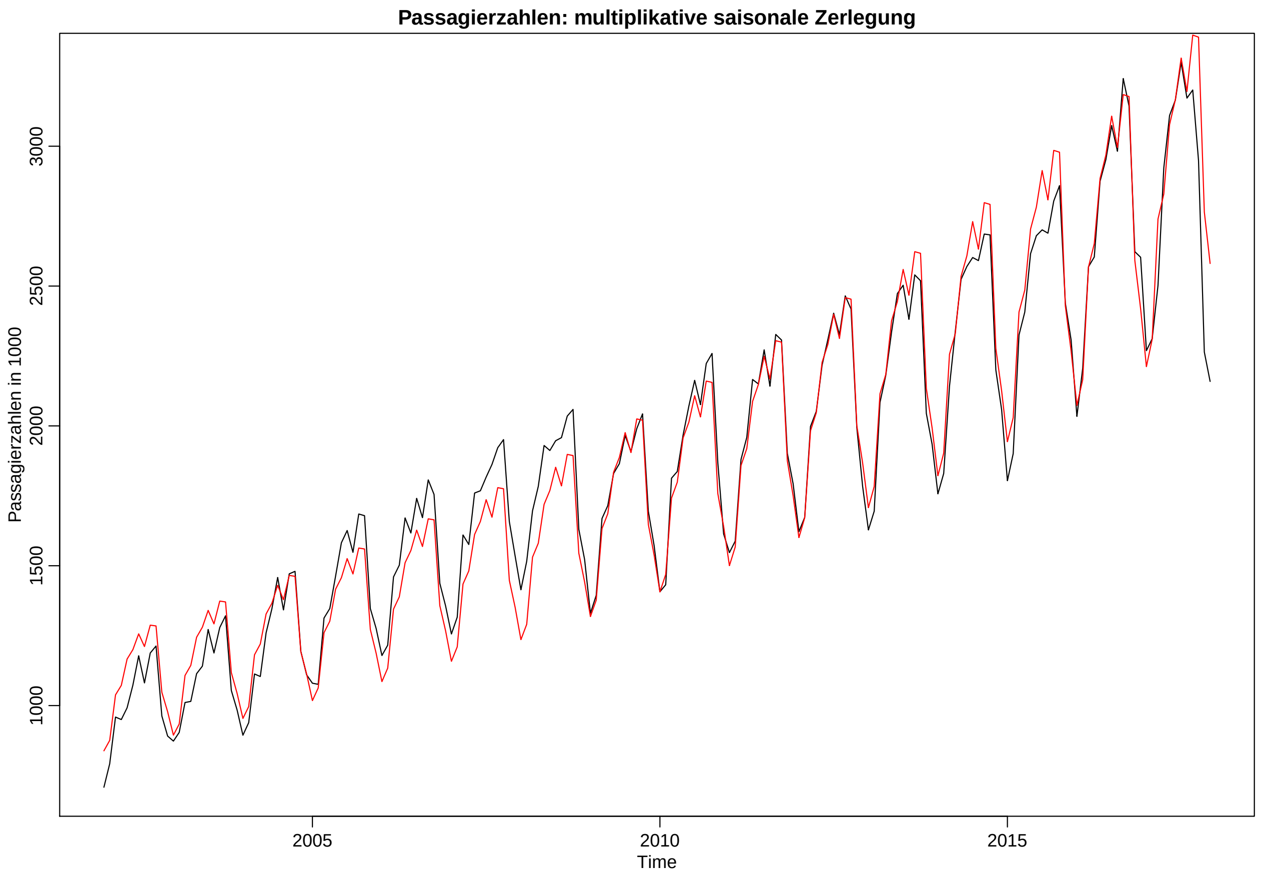
diese werden zu einer Reihe aus k Jahren fortgesetzt.

Phasendurchschnittsverfahren: Saison

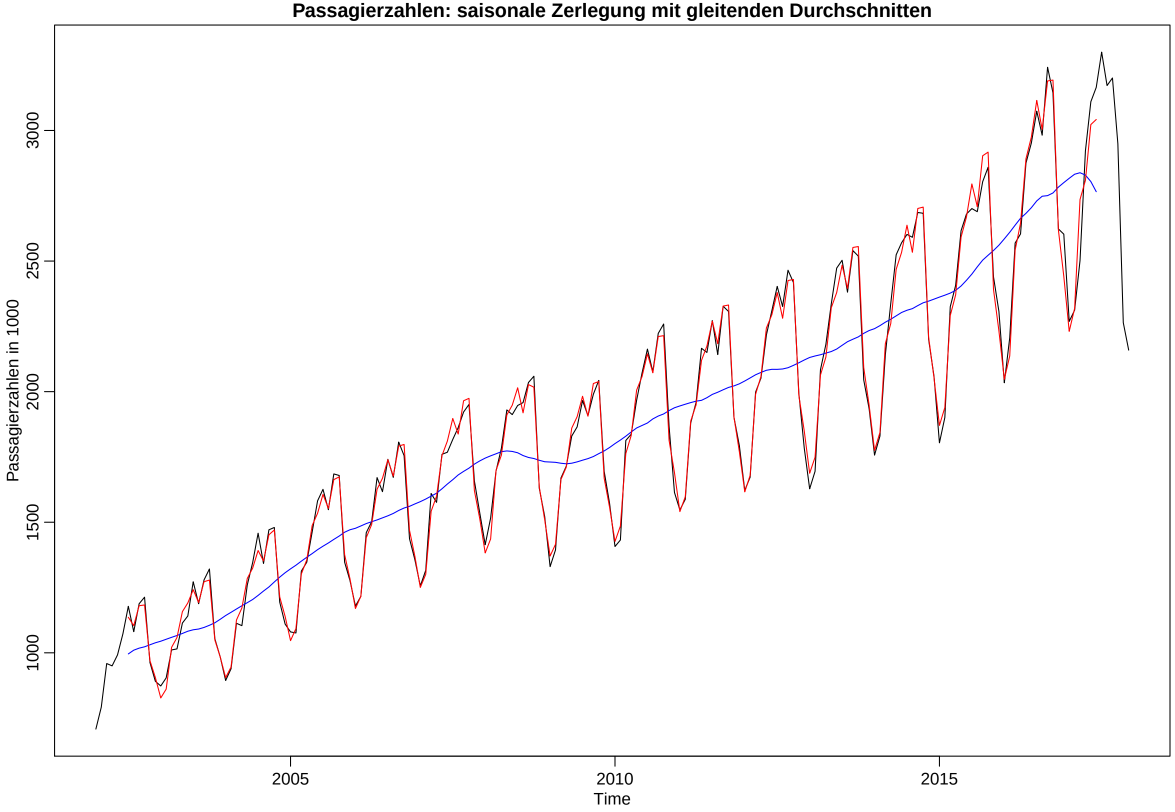


Phasendurchschnittsverfahren: Modell

$$\text{Modell} = \text{Trend} + \text{Saison}$$



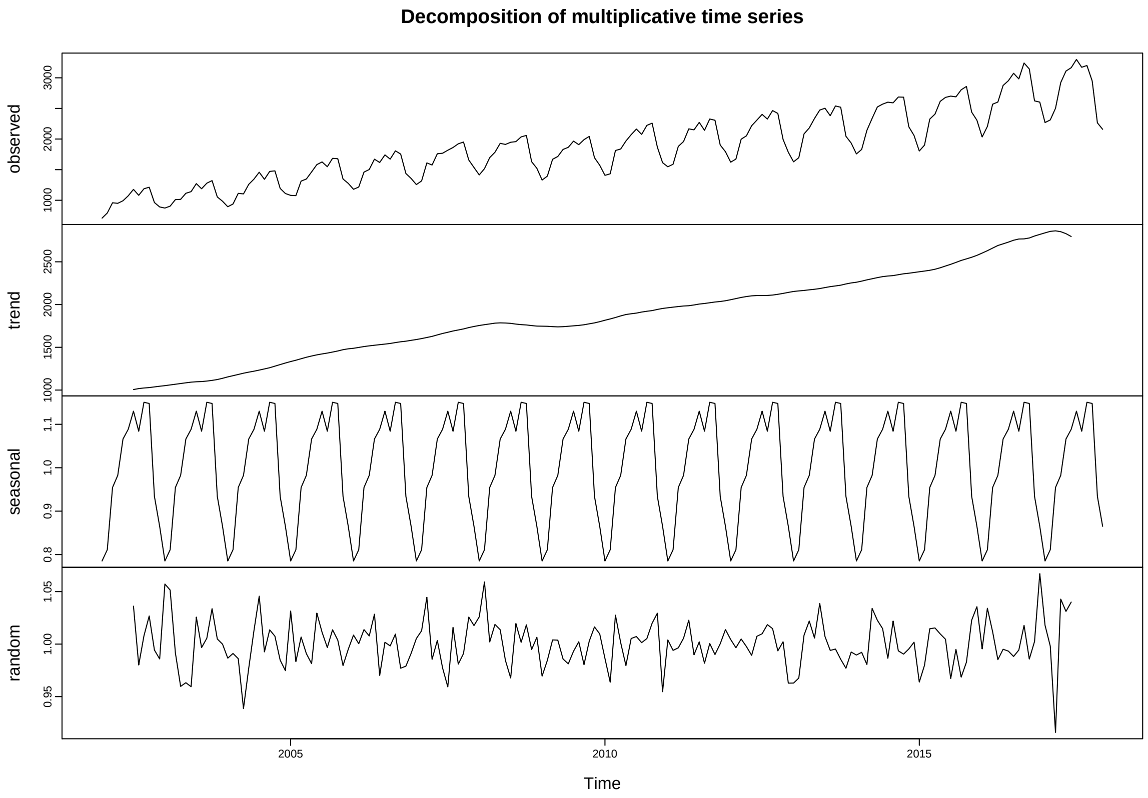
Phasendurchschnittsverfahren: Trend über gleitende Durchschnitte



→ Bessere Anpassung, aber fehlende Werte am Rand: Prognose?

Phasendurchschnittsverfahren: Trend über gleitende Durchschnitte

Berechnung wie bei dem polynomialen Trend oder direkt mit "decompose"



Regression mittels Saison-Dummies

Wir führen 12 Saisonvariablen

$$Q_i := \begin{cases} 1 & \text{für Monat } i \\ 0 & \text{sonst} \end{cases} \qquad i = 1, \ldots, 12$$

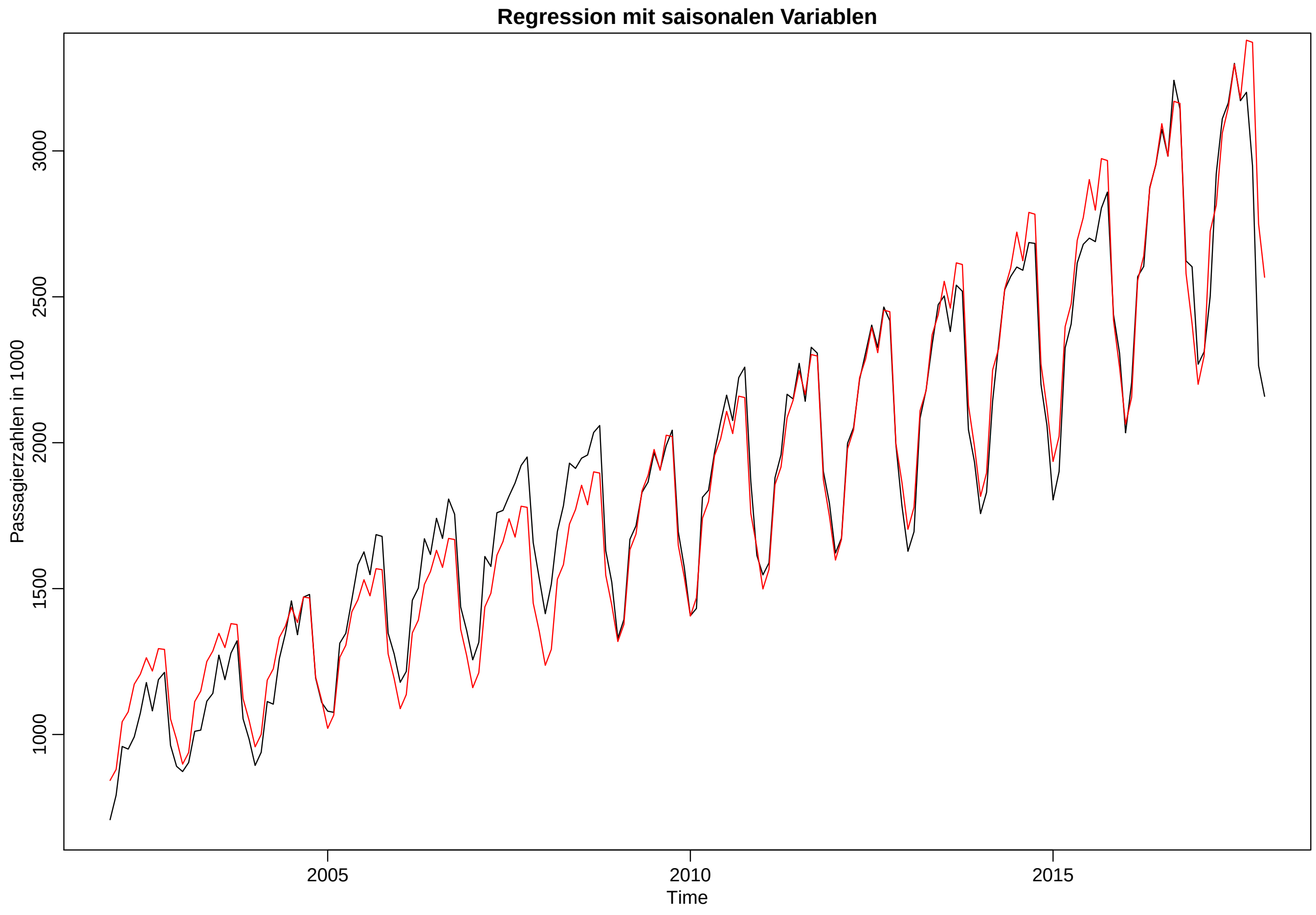
ein und machen einen (homogenen) Regressionsansatz

$$x_t = \beta_1 t + \beta_2 t^2 + \cdots + \beta_n t^n + \alpha_1 Q_1 + \cdots + \alpha_{12} Q_{12} + u_t.$$

(Inhomogen auch möglich, dann andere Interpretation der Koeffizienten.)

Vorteil: Trend und Saison werden gleichzeitig behandelt.

Regression mittels Saison-Dummies



Regression mit trigonometrischen Polynomen

Statt Saison-Dummies verwenden wir die Terme

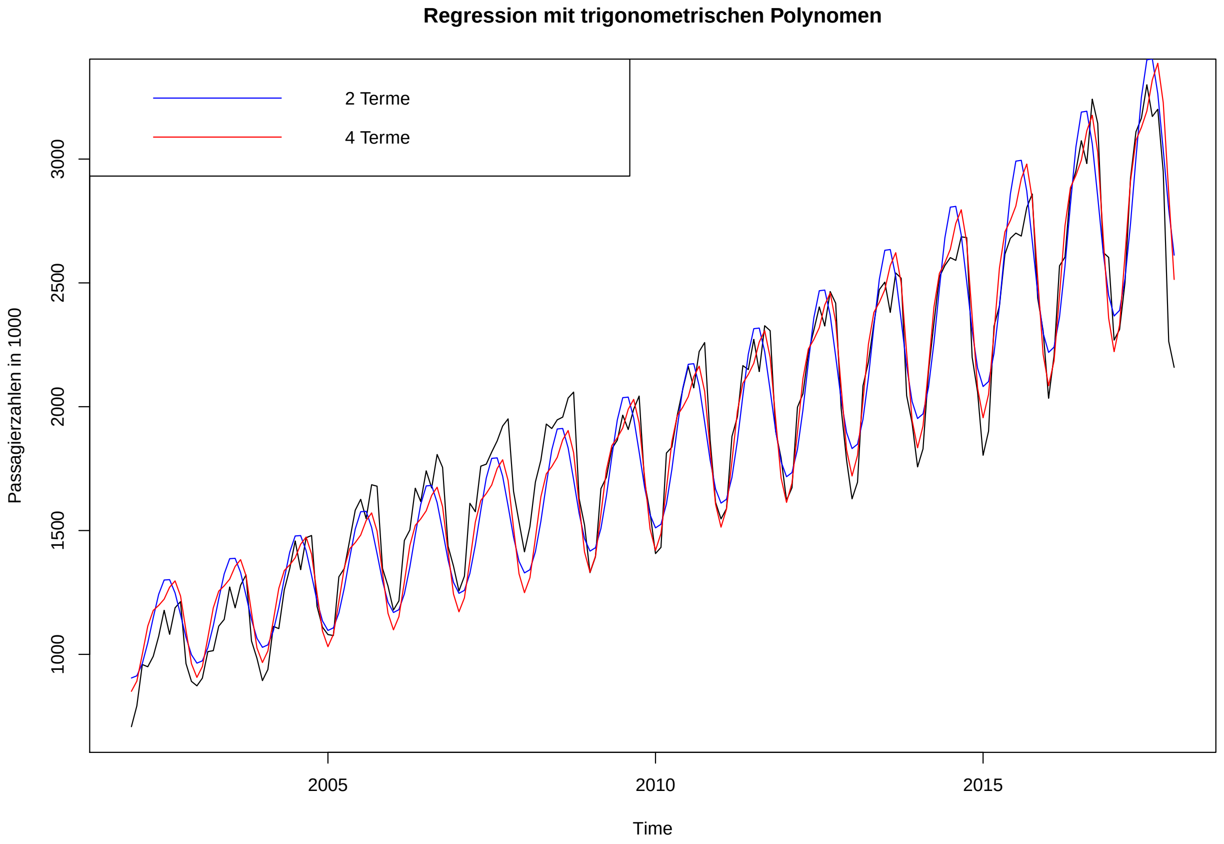
$$\cos\left(\frac{2\pi}{12}it\right), \quad \sin\left(\frac{2\pi}{12}it\right), \qquad i = 1, \ldots, 6.$$

Meistens reicht schon $i = 1$ (2 Terme), kompliziertere Saisonfiguren brauchen evt. mehr. Diese Terme sind periodisch mit Periode 12, für andere Perioden (z.B. Quartalszahlen) müssen sie entsprechend angepasst werden.

Wieder machen wir den Regressionsansatz

$$x_t = \beta_0 + \beta_1 t + \cdots + \beta_n t^n + \sum_{i=1}^{6} \left(\alpha_i \cos\left(\frac{2\pi}{12}it\right) + \phi_i \sin\left(\frac{2\pi}{12}it\right)\right).$$

Regression mit trigonometrischen Polynomen

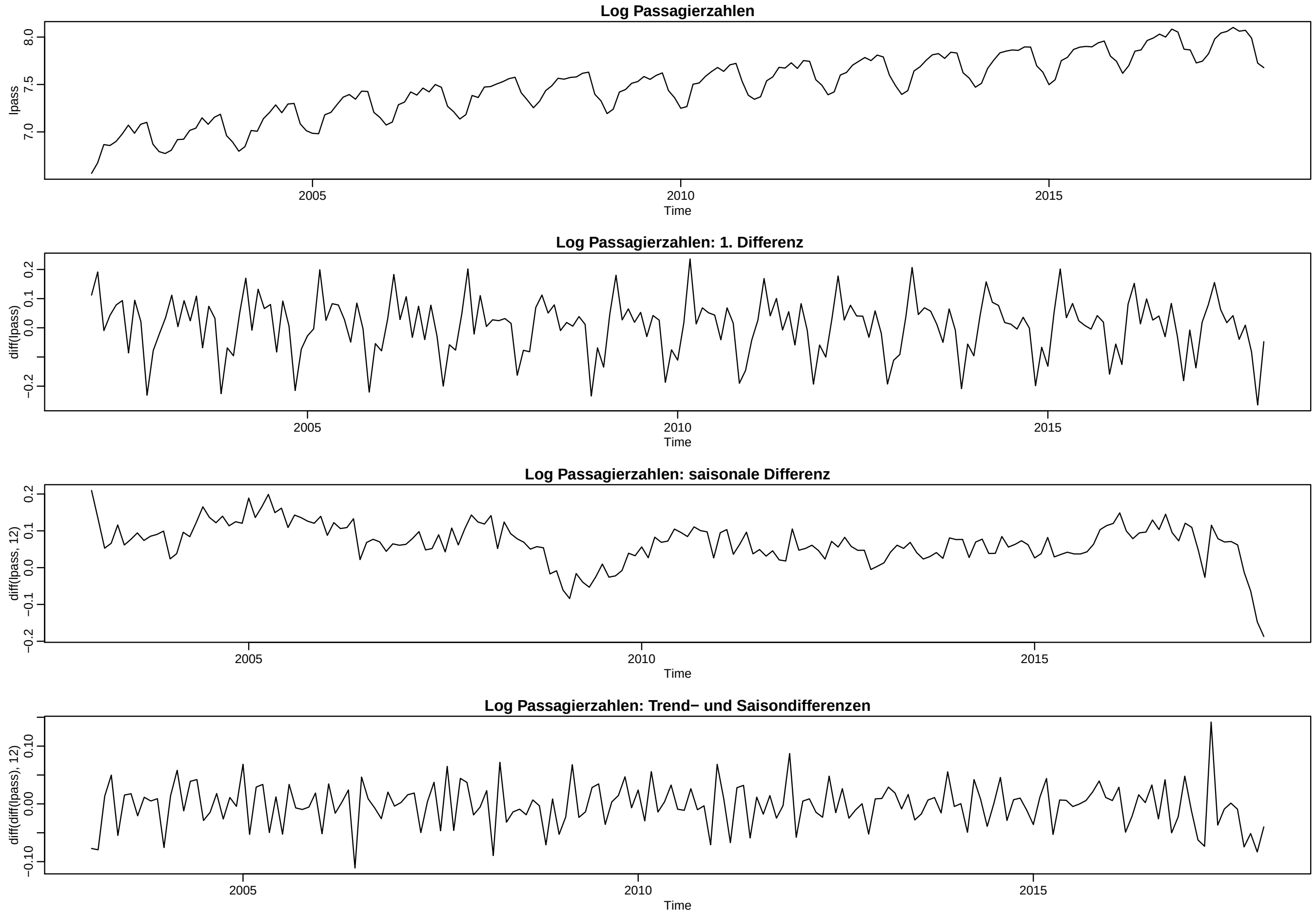


Hier mit 2 Termen.

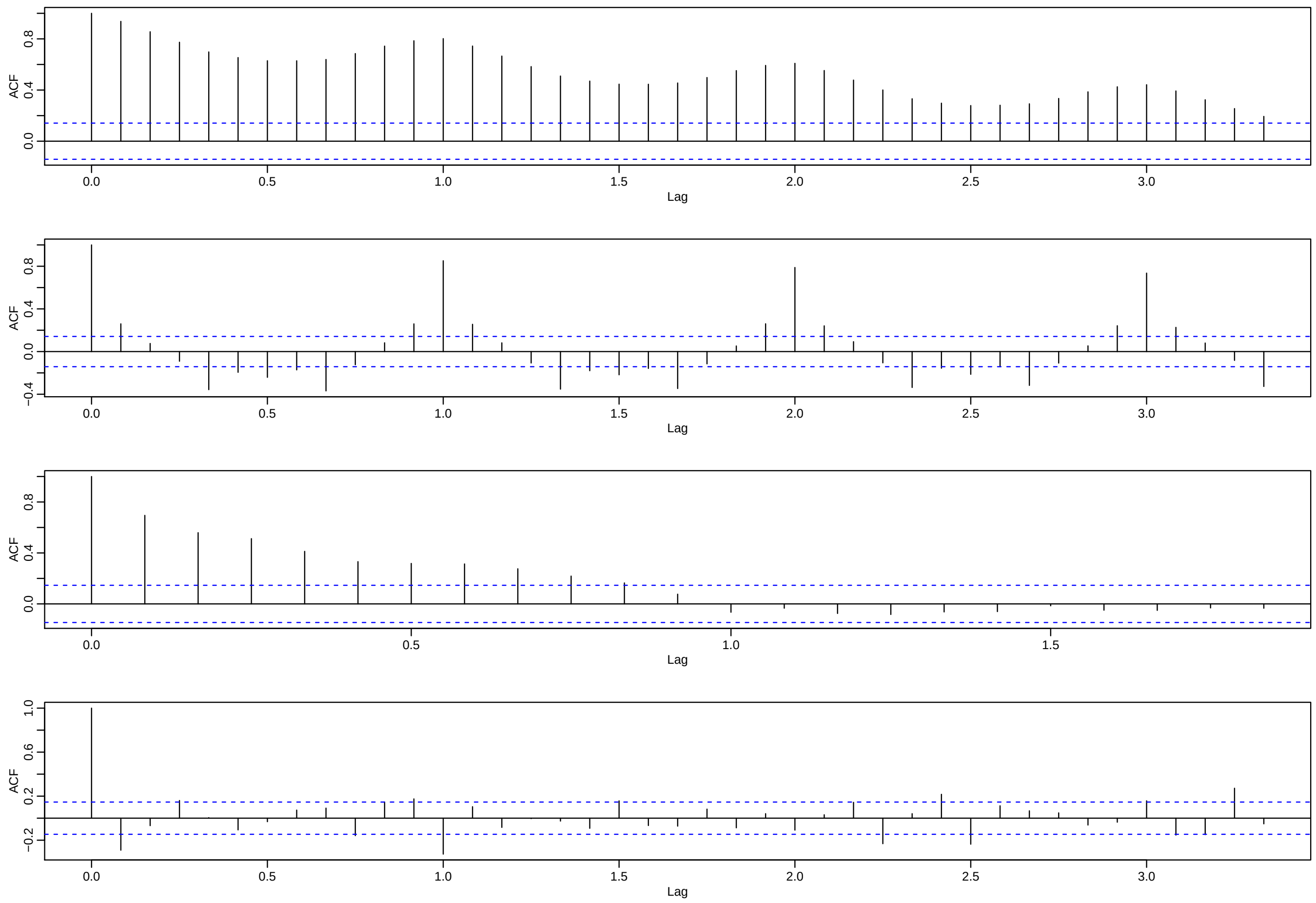
Vergleich der Modelle

```
> Rquadrat(mm, passb)
[1] 0.966658
> Rquadrat(mmg, passb)
[1] 0.9946772
> Rquadrat(exp(fitted.values(reg0)), passb)
[1] 0.9679645
> Rquadrat(exp(fitted.values(tfit)), passb)
[1] 0.940457
> Rquadrat(exp(fitted.values(tfit3a)), passb)
[1] 0.9604411
> AIC(reg0,tfit,tfit3a)
       df       AIC
reg0   14 -491.7374
tfit    5 -406.0068
tfit3a  7 -468.1257
> BIC(reg0,tfit,tfit3a)
       df       BIC
reg0   14 -446.1325
tfit    5 -389.7193
tfit3a  7 -445.3233
```

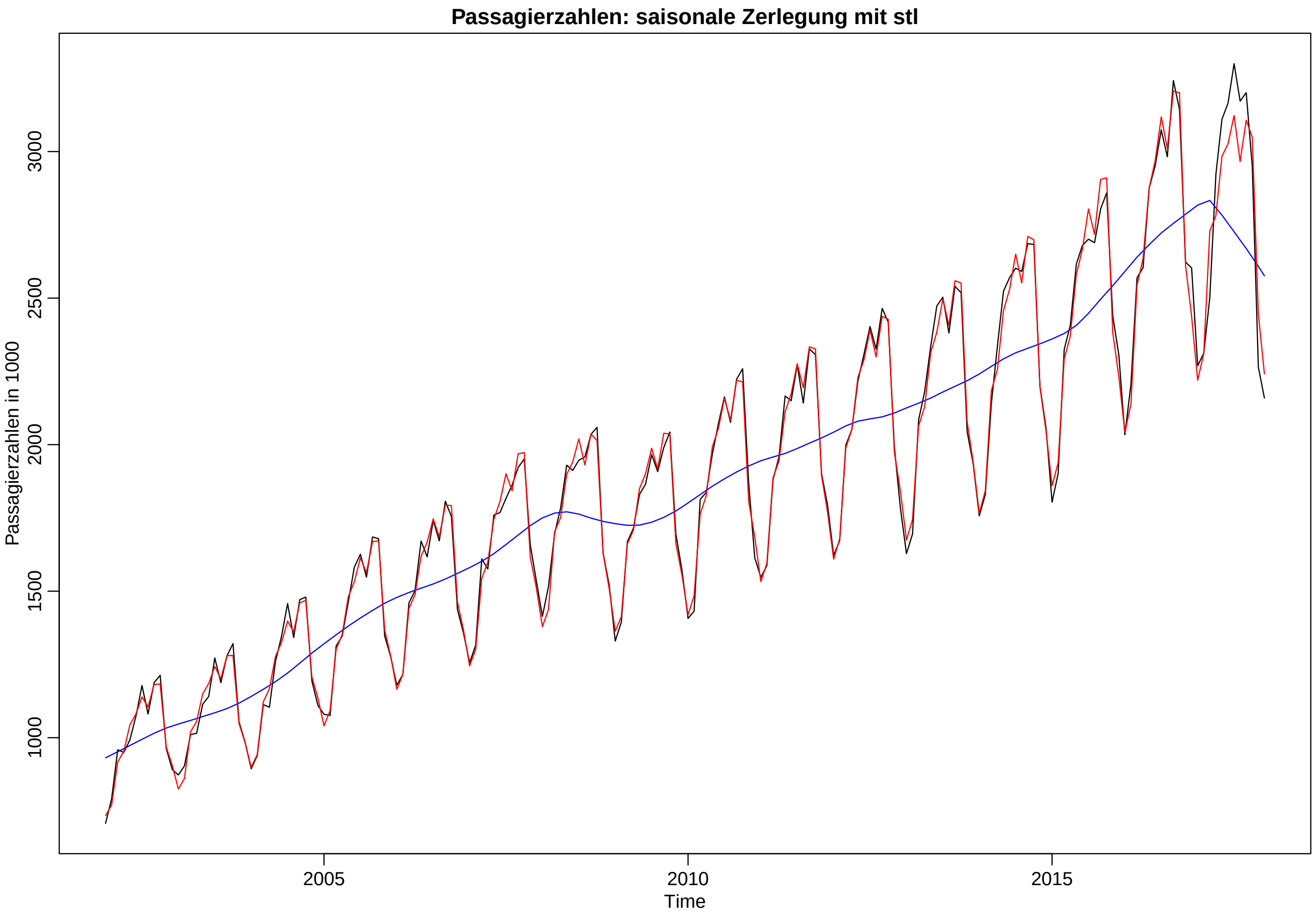
Saisonbereinigung durch Differenzenbildung



Saisonbereinigung durch Differenzenbildung



Zerlegung mit “stl”



Saisonbereinigung: Etablierte Verfahren

Berliner Verfahren BV4.1.

- Wird seit 2004 vom Statistischem Bundesamt verwendet
- Kostenloser Download unter `https://www.destatis.de/DE/Methoden/Zeitreihen/SoftwareZeitreihenanalyse.html`
- Zuerst (nach Wunsch) Kalender- bzw. Ausreißerbereinigung
- Schätzung des Trends lokal mit Polynomen 3. Grades, am Rand mit hohem linearen Anteil (“Randfilter”)
- Schätzung der Saisonkomponente lokal mit trigonometrischen Polynomen

Saisonbereinigung: Etablierte Verfahren

Census X-12-ARIMA-Verfahren, mittlerweile X-13-ARIMA-SEATS

- Vom U.S. Bureau of the Census entwickelt
- Kostenloser Download unter
  https://www.census.gov/srd/www/x13as/
- Kalender- und Ausreißerbereinigung
- Modellierung der Reihe durch Verbindung von Regressionsanalyse und ARIMA-Verfahren
- Schätzung der Saisonkomponente durch iterative Anwendung verschiedener gleitender Durchschnitte
- Diagnostika zur Überprüfung der Güte

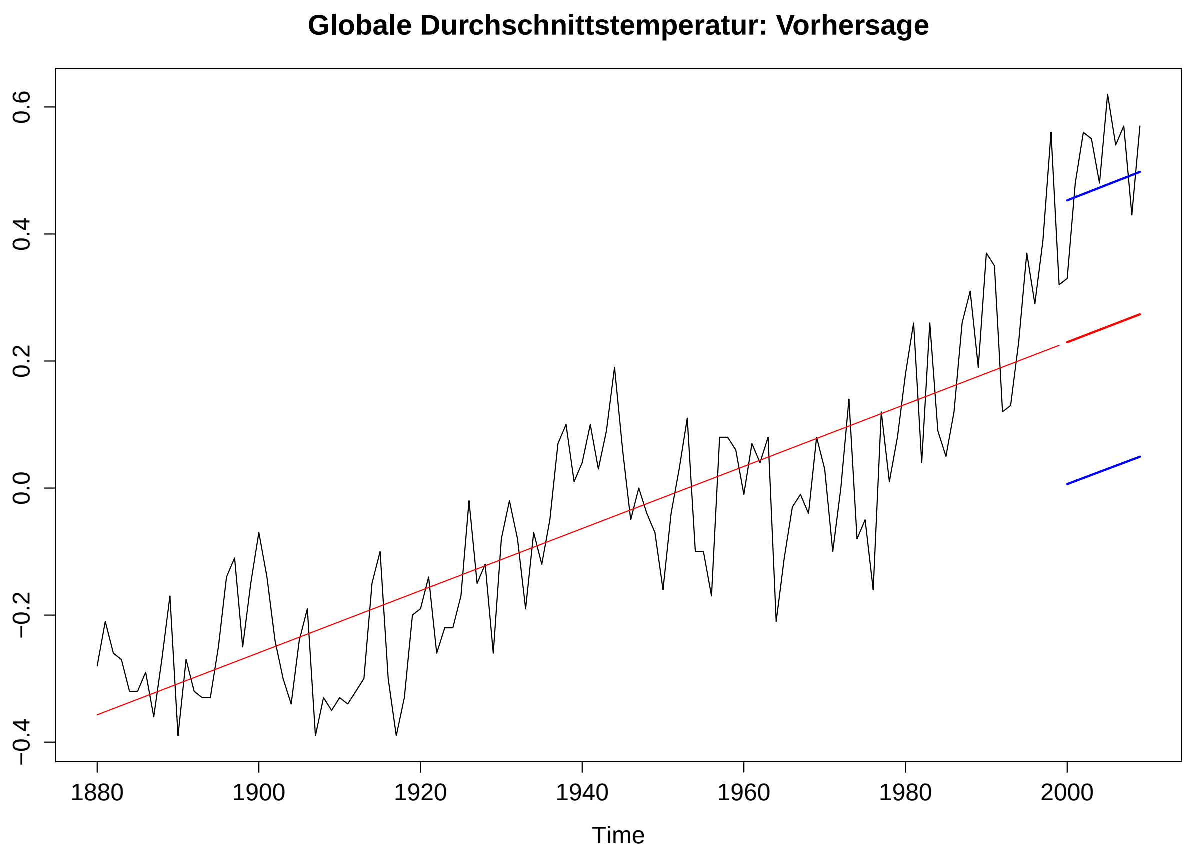
Gliederung der Vorlesung

1) Einführung und Grundlagen
2) Klassische Zeitreihenanalyse
   a) Trendbestimmung
   b) Glättung von Zeitreihen durch Filter (lokale Trendbestimmung)
   c) Saisonbereinigung
   d) Prognosen, exponentielles Glätten
3) Modellierung von Reihen durch stochastische Prozesse
   a) Autoregressive (AR) Modelle
   b) Moving Average (MA) Modelle
   c) ARMA Modelle
   d) ARIMA Modelle

Prognosen

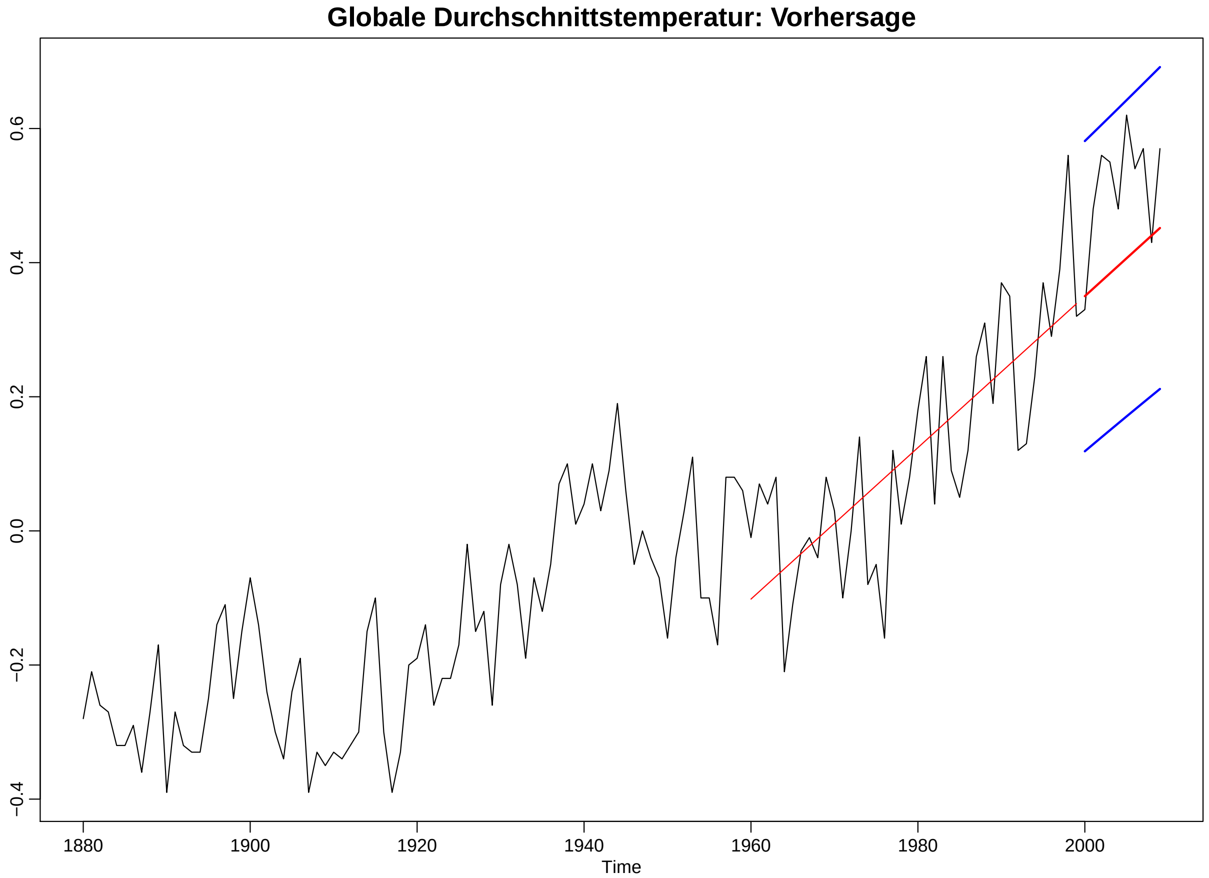
- Trendextrapolation: bei Polynomen kann außerhalb des Stützbereichs schief gehen $\rightarrow$ genau schauen! Konfidenzintervalle verfügbar, allerdings nur unter der Annahme von einem Gaußschen White Noise für Residuen.
- Exponentielles Glätten nach Holt-Winters
- ARIMA-Prognosen $\rightarrow$ später

Prognosen: Trendextrapolation



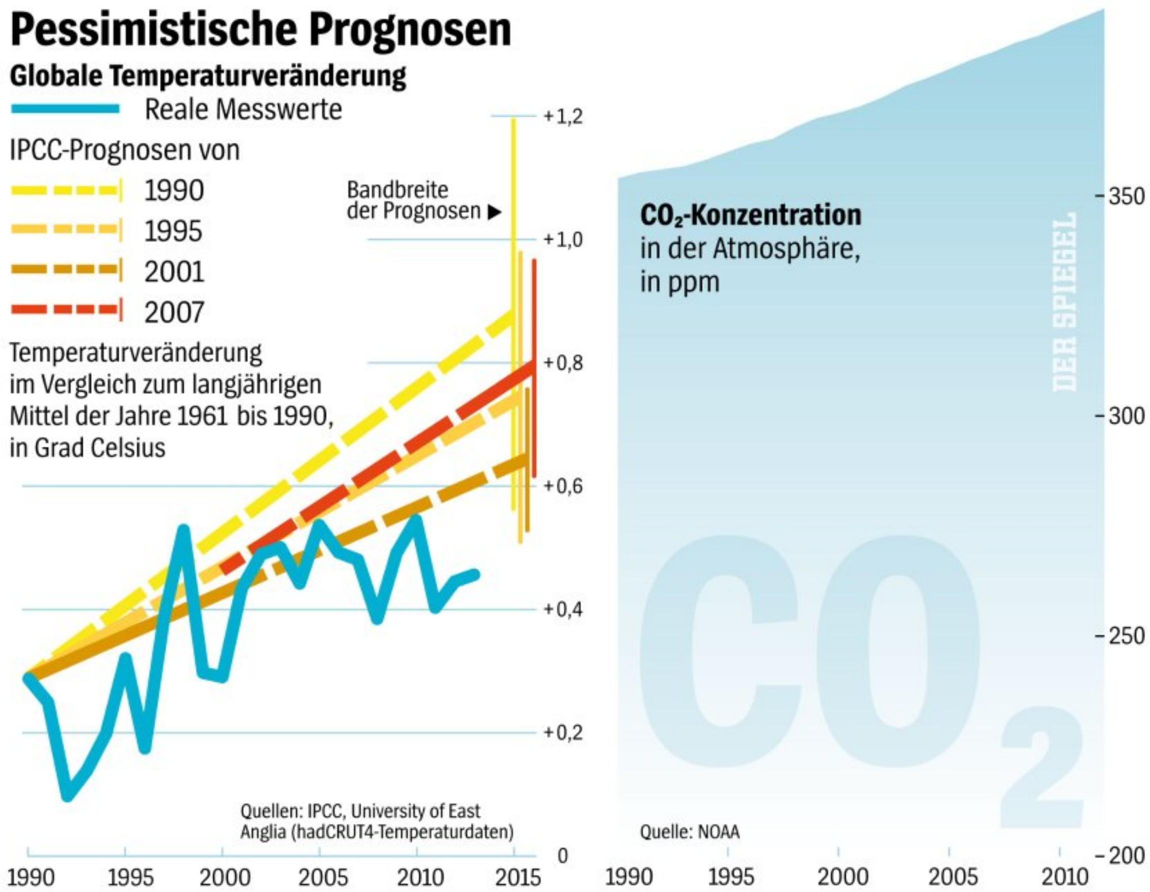
Linearer Trend zu niedrig...

Prognosen: Trendextrapolation

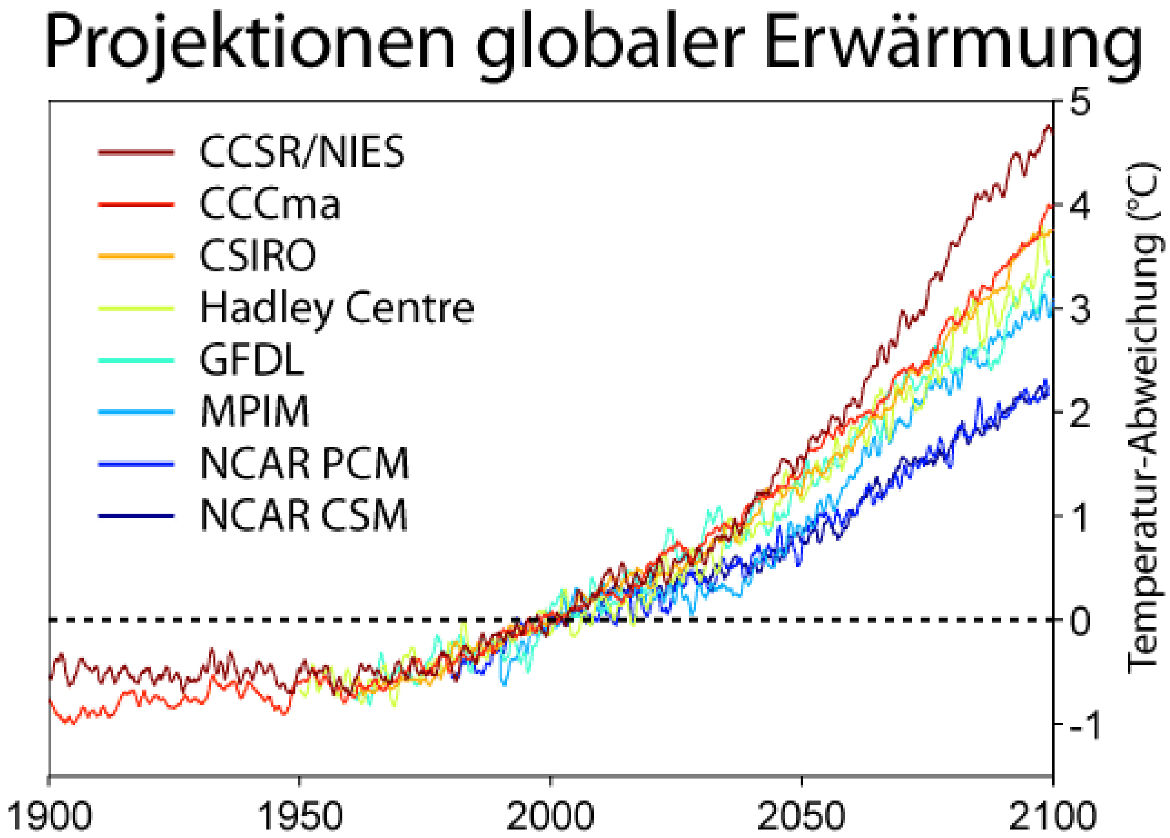


...außer man nimmt nur die letzten 50 Jahre.

Prognosen: Trendextrapolation

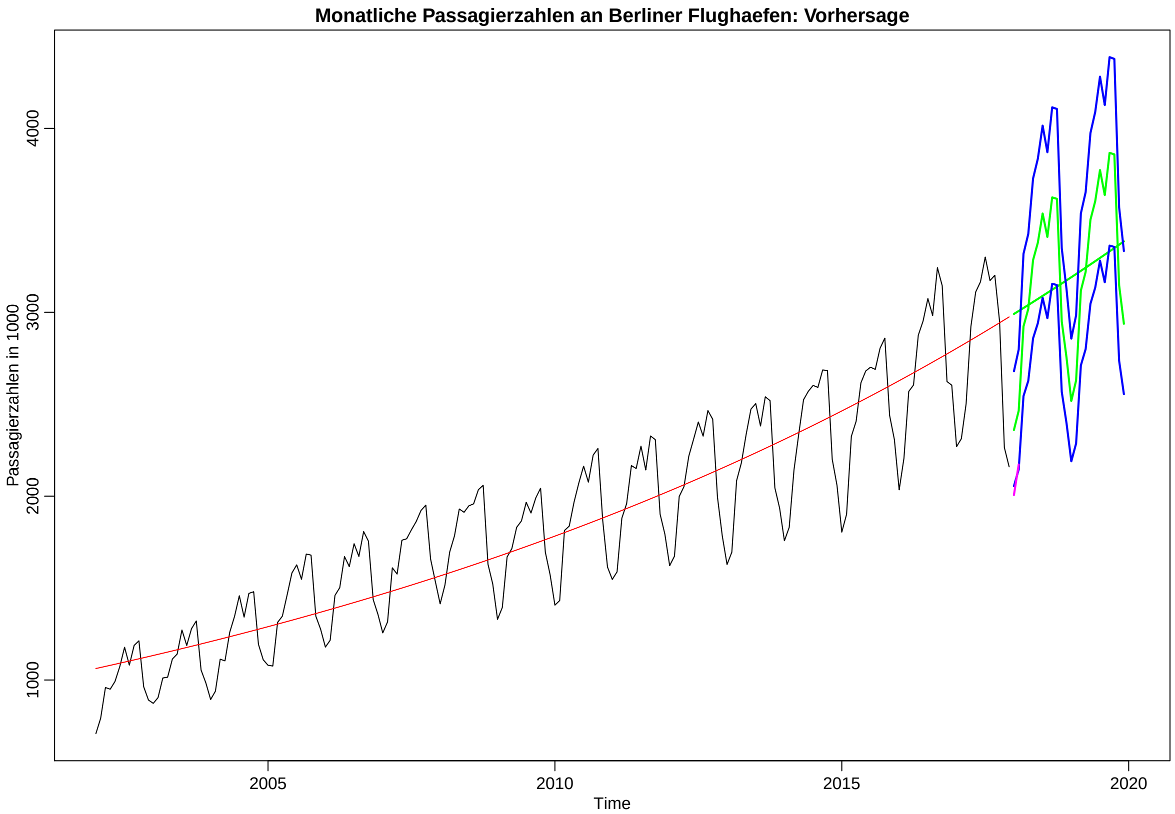


Prognosen: Trendextrapolation



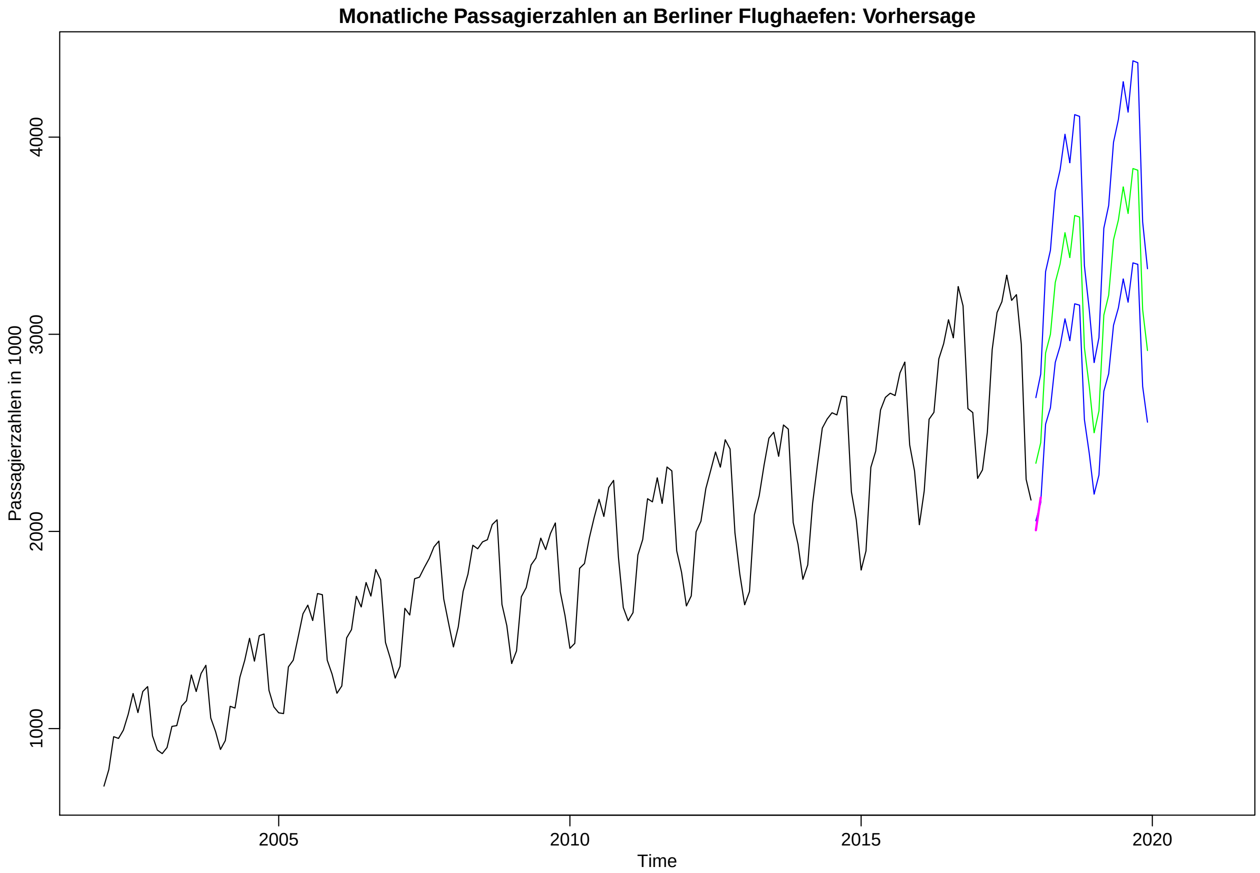
https://de.wikipedia.org/wiki/Globale_Erw%C3%A4rmung

Prognosen: Beispiel Passagierzahlen



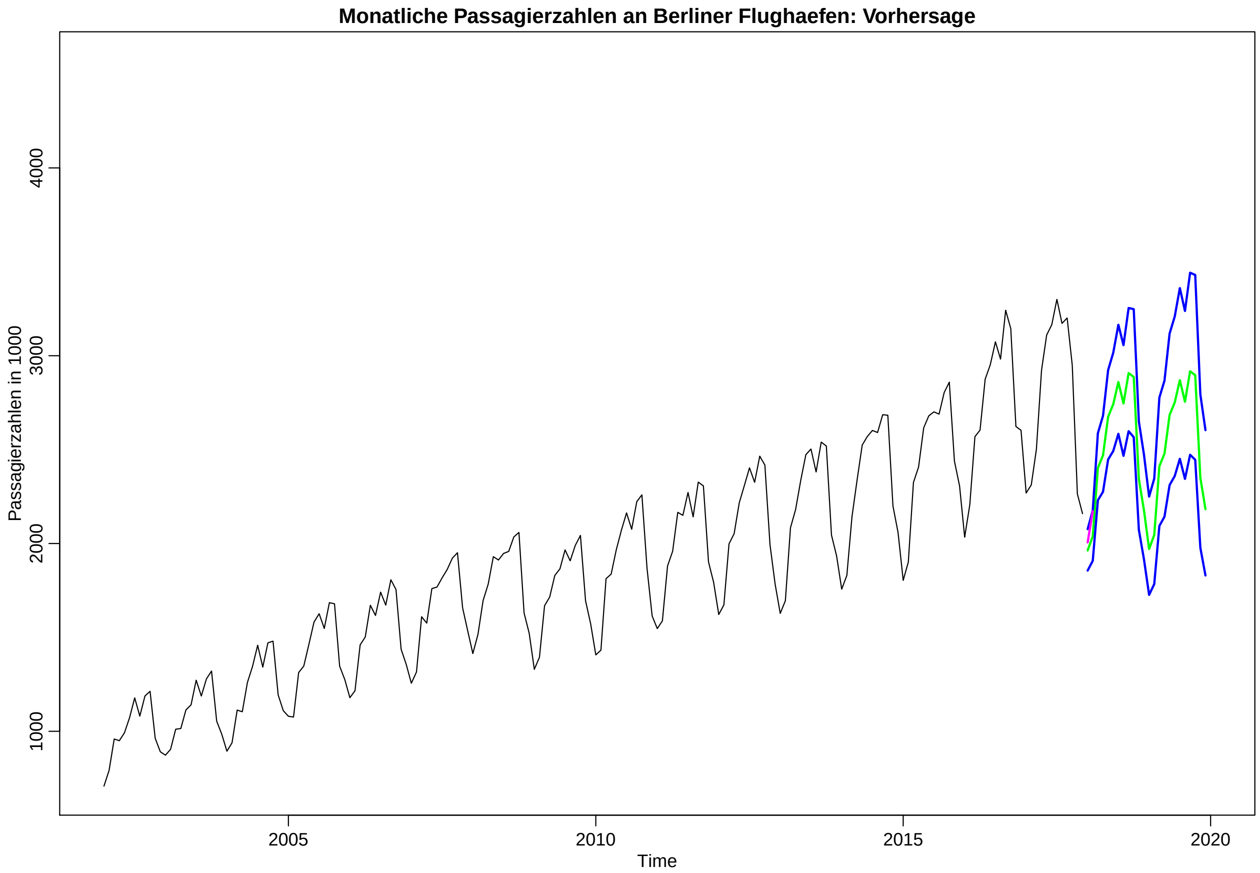
Genaue Werte: http://www.berlin-airport.de/de/presse/basisinformationen/verkehrsstatistik/.

Prognosen: Beispiel Passagierzahlen



Lineare Regression mit Saison-Dummies

Prognosen: Beispiel Passagierzahlen



Mit "stl"

Das einfache exponentielle Glätten

Ausgangspunkt ist ein einfaches Modell ohne Trend und Saison

$$X_t = \mu + \varepsilon_t$$

mit $\mu \in \mathbb{R}$ und White Noise (ε_k). Die zugehörige Reihe ist $(x_t)_{t=1,\ldots,N}$, und wir haben eine (Ein-Schritt-) Prognose $\hat{x}^N_{N+1}$ für x_{N+1}.

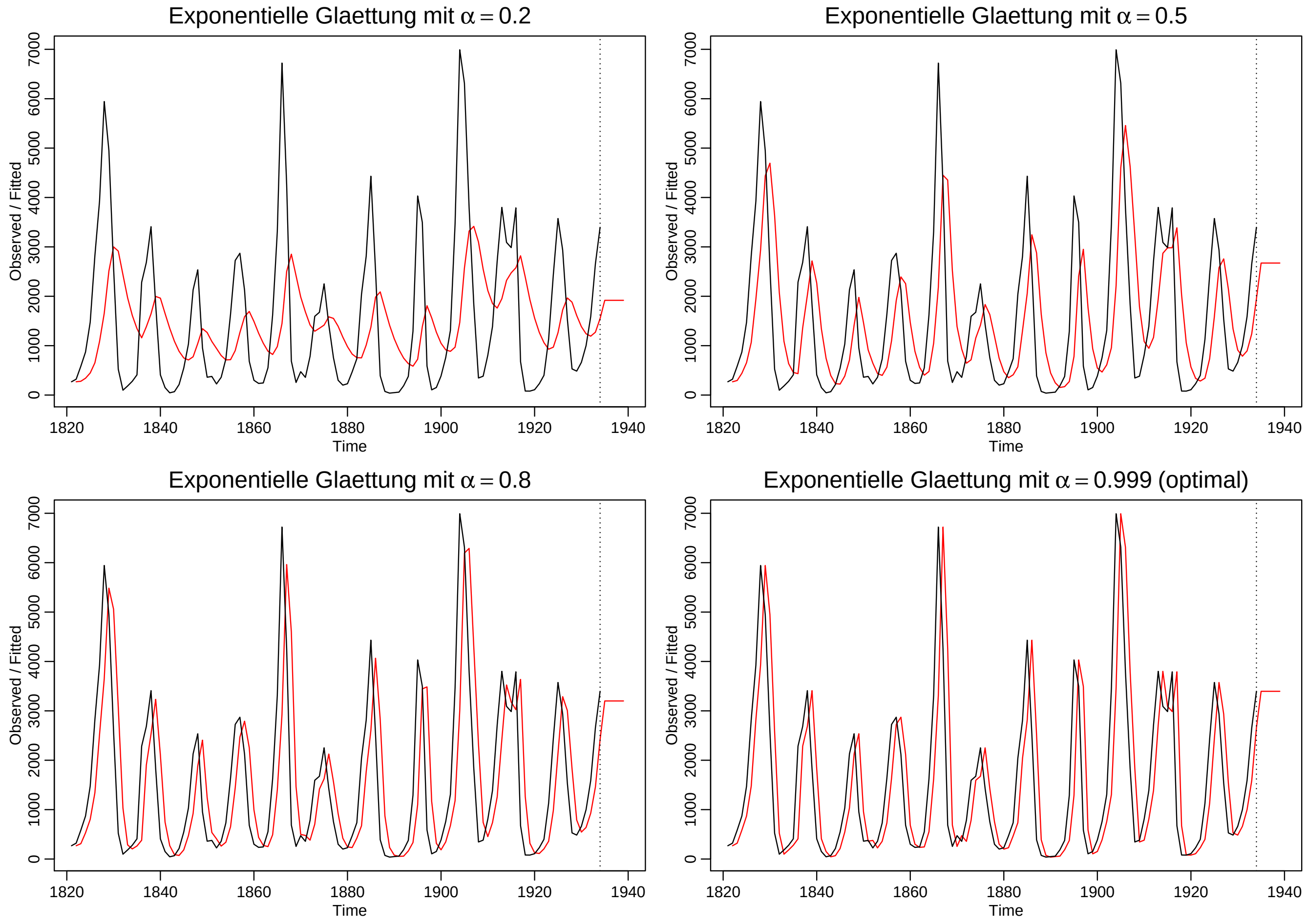
Nun kommt eine neue Beobachtung x_{N+1} hinzu, und wir möchten die Prognose "updaten" $\rightarrow$ Ansatz:

$$\hat{x}^{N+1}_{N+2} = (1-\alpha)\hat{x}^N_{N+1} + \alpha x_{N+1}$$

– ein rekursiver Filter aus der alten Prognose und dem neuen Wert.
Meist wird als Startwert $\hat{x}^1_1 := x_1$ gewählt.
$\alpha \in (0,1)$ heißt Glättungsparameter.

Das einfache exponentielle Glätten

$\alpha \in (0,1)$ heißt Glättungsparameter: Je kleiner α, desto stärker wird geglättet.



Das einfache exponentielle Glätten

$\alpha \in (0,1)$ heißt Glättungsparameter und wird so gewählt, dass

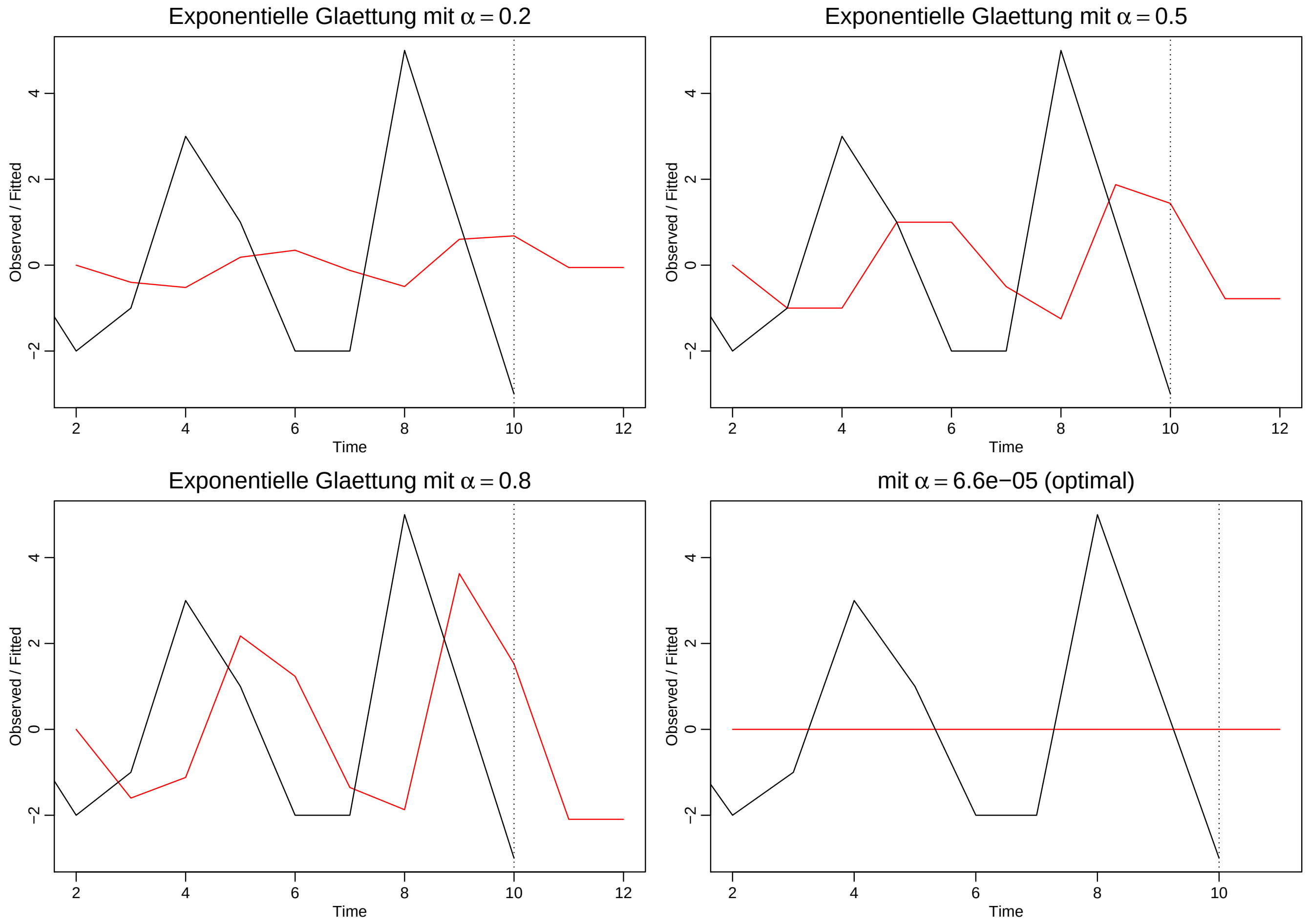
$$\sum_{t=m}^{N-1} \left(x_{t+1} - \hat{x}_1^t\right)^2 \stackrel{!}{\to} \min_{\alpha}.$$

(Die untere Summationsgrenze m soll den Effekt des Startwertes vernachlässigbar machen.) Meist reicht es die Werte $0.1, 0.2, \ldots, 0.9$ für α auszuprobieren, R macht es automatisch.

Da das zugrundeliegende Modell eines mit einem konstantem Niveau ist, wird für h-Schritt Prognosen ($h > 1$) einfach die 1-Schritt-Prognose fortgeschrieben:

$$\hat{x}_{N+h}^N = \hat{x}_{N+1}^N.$$

Das einfache exponentielle Glätten



Das einfache exponentielle Glätten

```
#Annual Canadian Lynx trappings 1821-1934
#einfach
plot(lynx, main="Anzahl gefangener Luchse in Kanada 1821-1934")
(hwl2 = HoltWinters(lynx, alpha=.2, beta=F, gamma=F))
(hwl5 = HoltWinters(lynx, alpha=.5, beta=F, gamma=F))
(hwl8 = HoltWinters(lynx, alpha=.8, beta=F, gamma=F))
(hwl = HoltWinters(lynx, beta=F, gamma=F))
mean(lynx)
(hwl1 = HoltWinters(lynx, alpha=.1, beta=F, gamma=F))
p2=predict(hwl2, n.ahead = 5)
p5=predict(hwl5, n.ahead = 5)
p8=predict(hwl8, n.ahead = 5)
p=predict(hwl, n.ahead = 5)
par(mfrow=c(2,2), mar=c(3,2.5,2,1), cex.main=1.5)
plot(hwl2, p2, main=as.expression(bquote("Exponentielle Glaettung mit" ~ alpha == 0.2)))
plot(hwl5, p5, main=as.expression(bquote("Exponentielle Glaettung mit" ~ alpha == 0.5)))
plot(hwl8, p8, main=as.expression(bquote("Exponentielle Glaettung mit" ~ alpha == 0.8)))
plot(hwl, p, main=as.expression(bquote("Exponentielle Glaettung mit" ~ alpha == 0.999~ "(optimal)")))

#mit Trend
(hwlt = HoltWinters(lynx, gamma=F))
pt=predict(hwlt, n.ahead = 5)
plot(hwlt, pt, main=as.expression(bquote("Exponentielle Glaettung mit Trend")))
#wocher der Trend?
diff(lynx)[1]
(hwlto = HoltWinters(lynx, gamma=F, b.start=0))
pto=predict(hwlto, n.ahead = 5)
plot(hwlto, pto, main=as.expression(bquote("Exponentielle Glaettung mit Trend")))
```

Exponentielles Glätten nach Holt

Nun betrachten wir ein Modell mit linearem Trend und ohne Saison

$$X_t = m_t + \varepsilon_t = a + bt + \varepsilon_t.$$

$\rightarrow$ 2 Glättungsparameter: α zur Anpassung des Trends m_t und β zur Anpassung der Steigung b. Rekursion:

$$\hat{b}_t = (1-\beta)\hat{b}_{t-1} + \beta(\hat{m}_{t-1} - \hat{m}_{t-2}),$$

$$\hat{m}_t = (1-\alpha)(\hat{m}_{t-1} + \hat{b}_t) + \alpha x_t, \qquad t = 2, 3, \ldots, N.$$

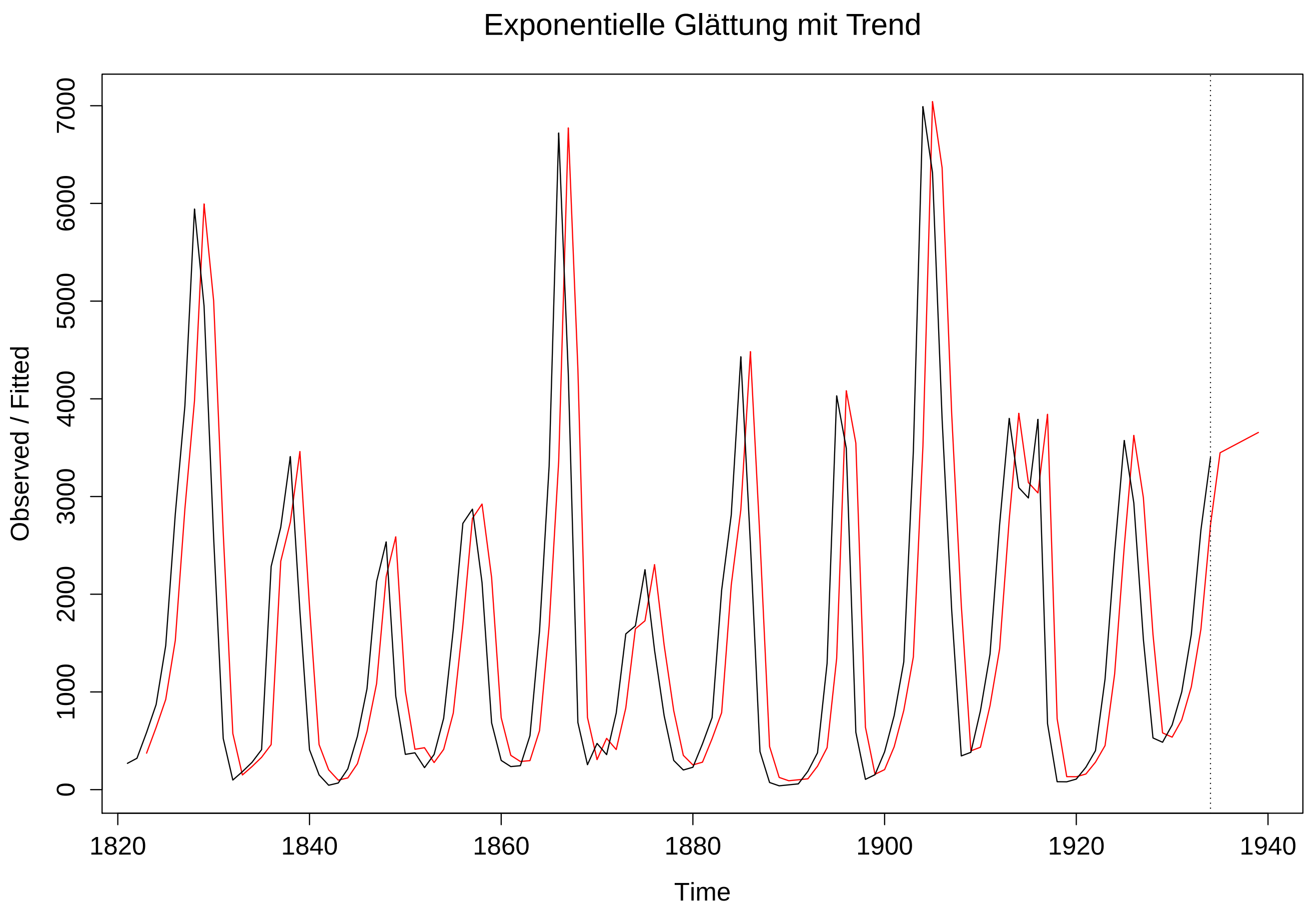
h-Schritt-Prognose:

$$\hat{x}_{N+h}^{N} = \hat{m}_N + \hat{b}_N \cdot h, \qquad h = 1, 2, \ldots$$

Startwerte: Z.B.

$$\hat{m}_1 = x_1, \quad \hat{m}_2 = x_2, \quad \hat{b}_1 = x_2 - x_1, \quad \hat{b}_2 = \hat{b}_1.$$

Exponentielles Glätten nach Holt



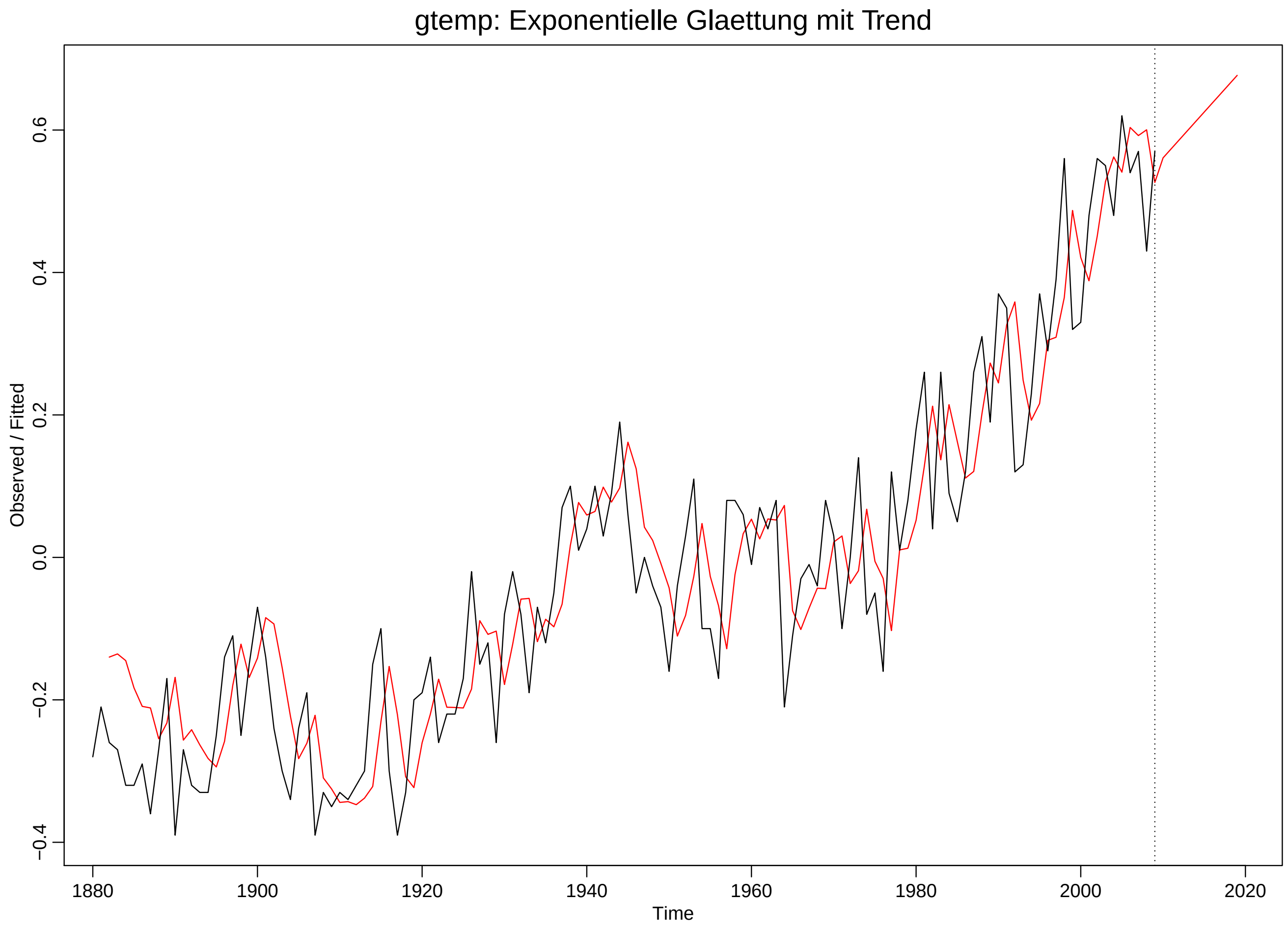
Exponentielles Glätten nach Holt

```
> (hwlt = HoltWinters(lynx, gamma=F))
Holt-Winters exponential smoothing with trend and without seasonal component.

Call:
HoltWinters(x = lynx, gamma = F)

Smoothing parameters:
 alpha: 1
 beta : 0
 gamma: FALSE

Coefficients:
  [,1]
a 3396
b   52
```

Woher der Trend?

```
> diff(lynx)[1]
[1] 52
> (hwlto = HoltWinters(lynx, gamma=F, b.start=0))
Holt-Winters exponential smoothing with trend and without seasonal component.

Call:
HoltWinters(x = lynx, gamma = F, b.start = 0)

Smoothing parameters:
 alpha: 1
 beta : 0
 gamma: FALSE

Coefficients:
  [,1]
a 3396
b    0
```

Exponentielles Glätten nach Holt



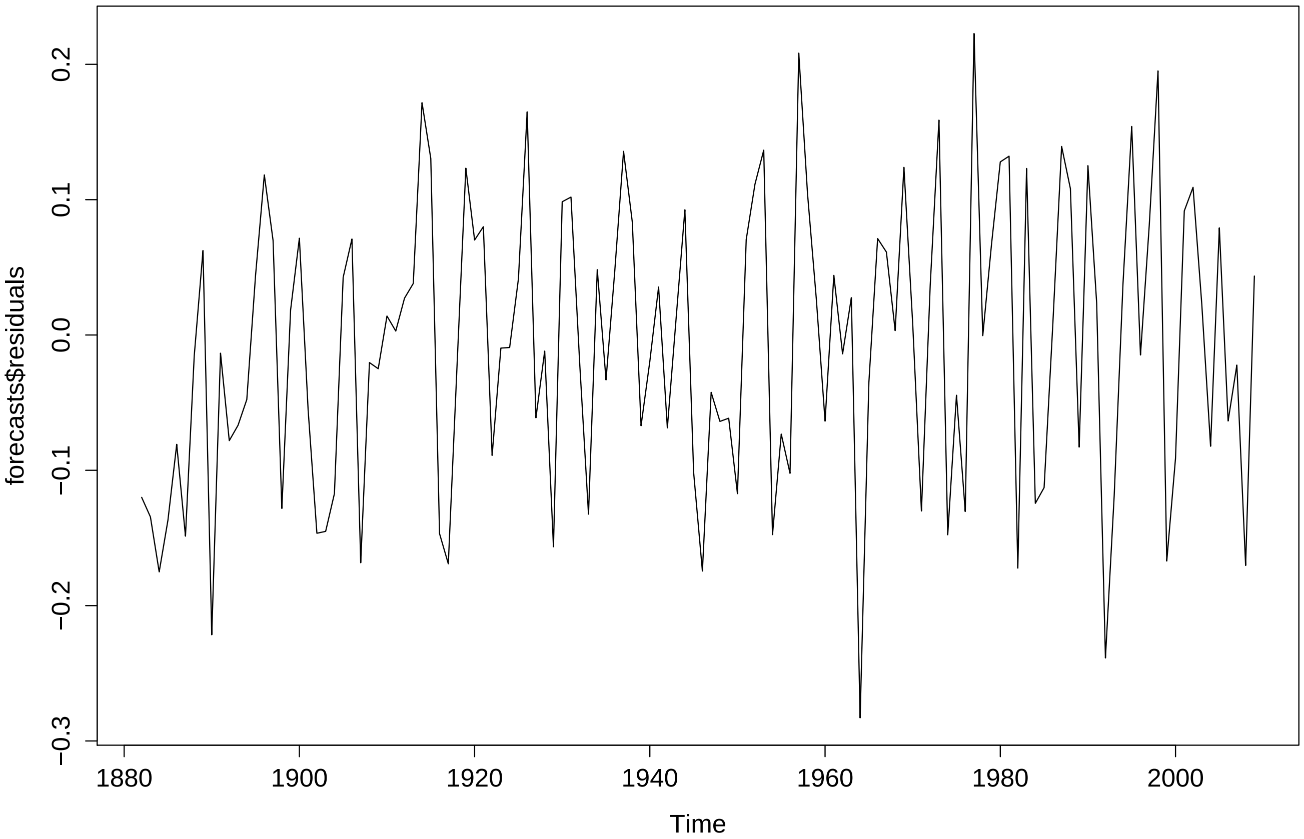
https://de.wikipedia.org/wiki/Globale_Erw%C3%A4rmung

Exponentielles Glätten nach Holt

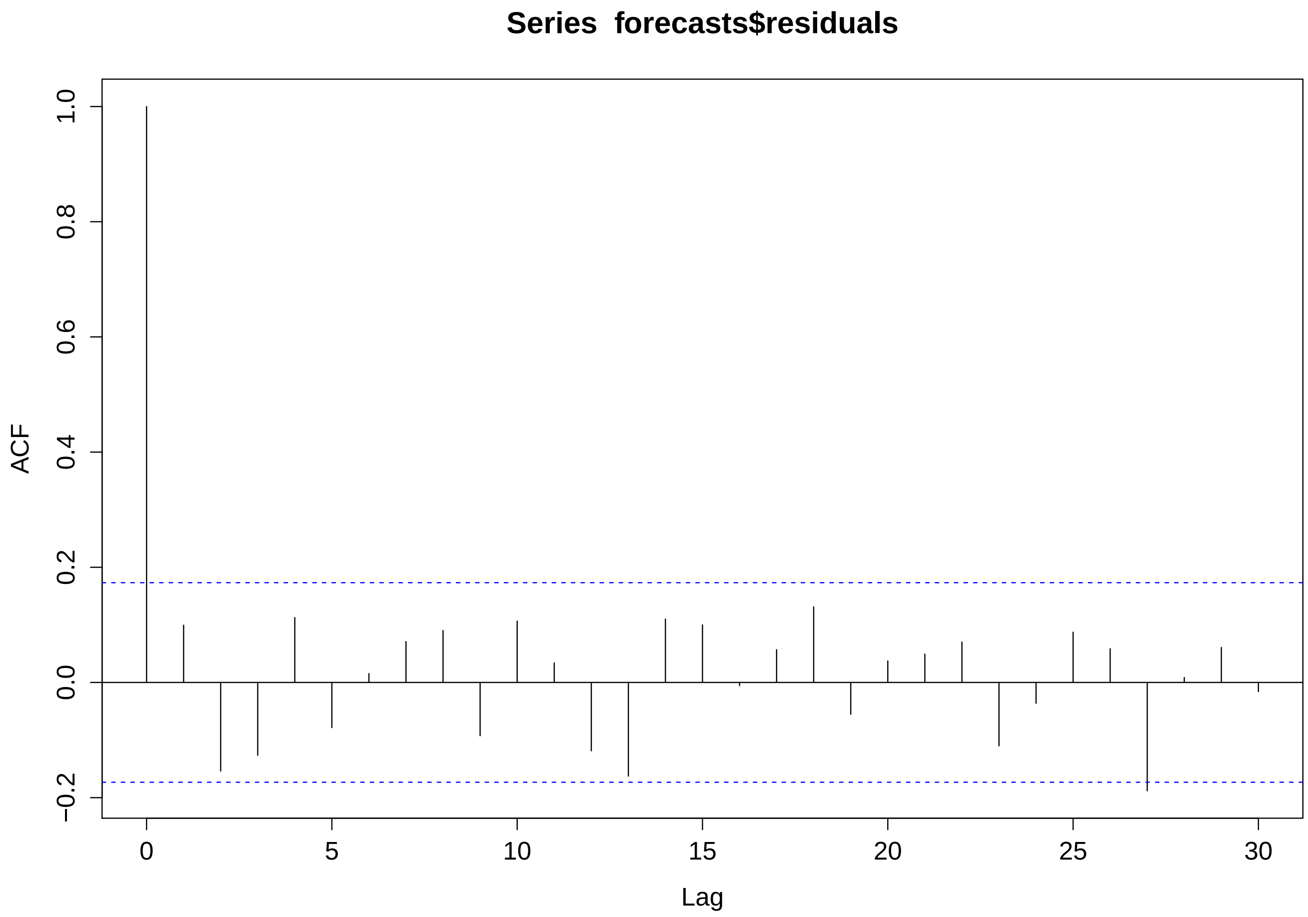
```
> (hwlt = HoltWinters(gtemp, gamma=F))
Holt-Winters exponential smoothing with trend and without seasonal component.

Call:
HoltWinters(x = gtemp, gamma = F)

Smoothing parameters:
 alpha: 0.4967435
 beta : 0.09964769
 gamma: FALSE

Coefficients:
       [,1]
a 0.5480703
b 0.0128656
```

```
#oder
(forecasts=forecast.HoltWinters(hwlt, h=10))
plot.forecast(forecasts)
res=na.omit(forecasts$residuals)
plot(res)
acf(res,30)
Box.test(res, lag=20, type="Ljung-Box", fitdf=2)
```

Exponentielles Glätten nach Holt



Exponentielles Glätten nach Holt



Exponentielles Glätten nach Holt-Winters

Nun betrachten wir ein (additives) Modell mit linearem Trend und Saison (Periode s)

$$X_t = m_t + s_t + \varepsilon_t = a + bt + s_t + \varepsilon_t.$$

$\to$ 3 Glättungsparameter: α, β wie vorher und γ zur Anpassung der Saison. Rekursion:

$$\begin{aligned}
\hat{b}_t &= (1-\beta)\hat{b}_{t-1} + \beta(\hat{m}_{t-1} - \hat{m}_{t-2}),\\
\hat{m}_t &= (1-\alpha)(\hat{m}_{t-1} + \hat{b}_t) + \alpha(x_t - \hat{s}_{t-s})\\
\hat{s}_t &= (1-\gamma)\hat{s}_{t-s} + \gamma(x_t - \hat{m}_t), \qquad t = s+1, \ldots, N.
\end{aligned}$$

h-Schritt-Prognose:

$$\hat{x}_{N+h}^N = \hat{m}_N + \hat{b}_N \cdot h + \hat{s}_{t+h-s}, \qquad h = 1, 2, \ldots$$

Exponentielles Glätten nach Holt-Winters

Rekursion

$$\hat{b}_t = (1-\beta)\hat{b}_{t-1} + \beta(\hat{m}_{t-1} - \hat{m}_{t-2}),$$
$$\hat{m}_t = (1-\alpha)(\hat{m}_{t-1} + \hat{b}_t) + \alpha(x_t - \hat{s}_{t-s})$$
$$\hat{s}_t = (1-\gamma)\hat{s}_{t-s} + \gamma(x_t - \hat{m}_t), \qquad t = s+1, \dots, N$$

kann erst ab $t = s+1$ beginnen, Startwerte Z.B.

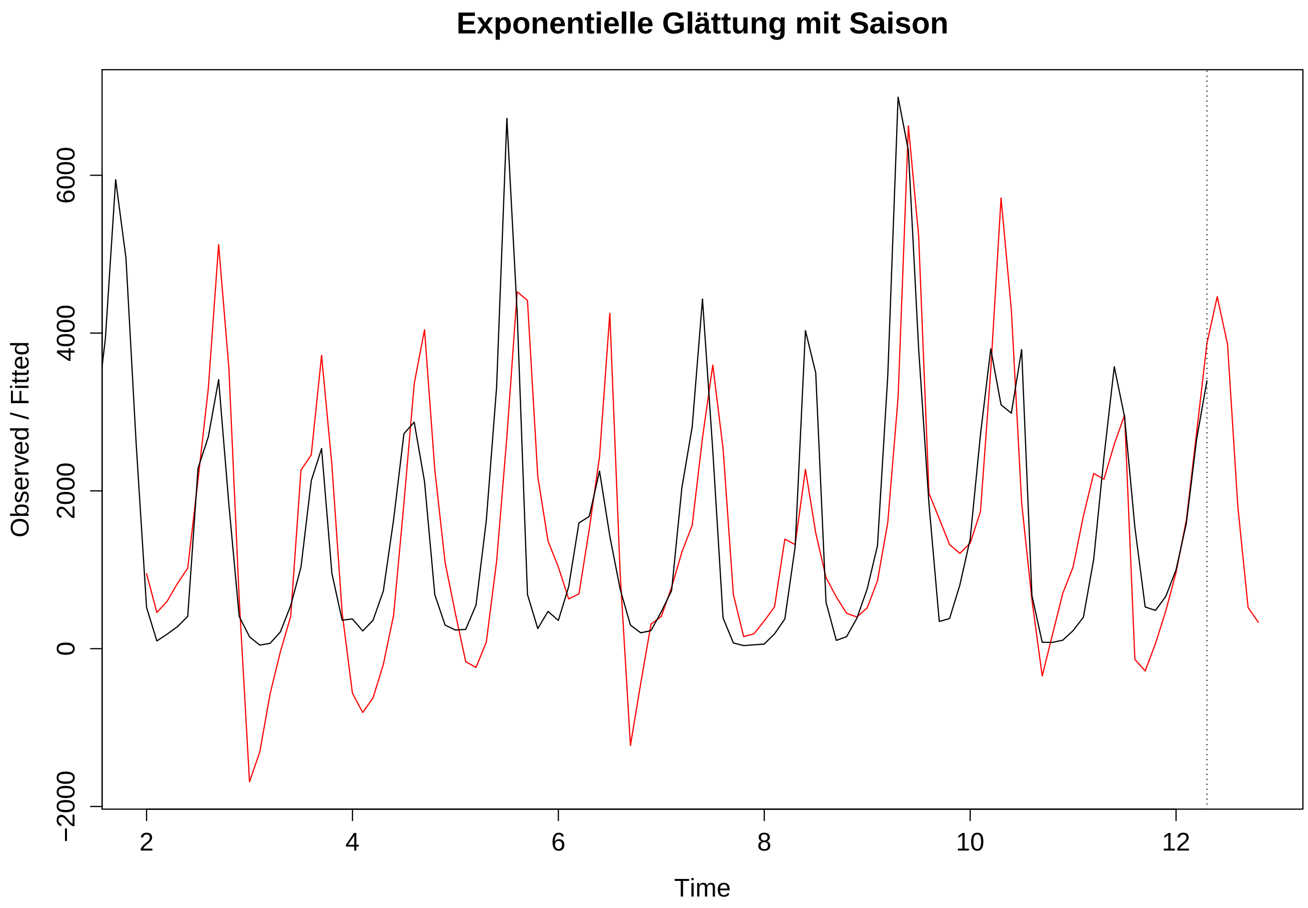
$$\hat{m}_{s-1} = \hat{m}_s = \overline{x_s} = \frac{1}{s}\sum_{k=1}^{s} x_k, \quad \hat{b}_s = 0, \quad \hat{s}_t = x_t - \overline{x_s}, \quad t = 1, \dots, s$$

Variante: Multiplikatives Modell mit linearem Trend und Saison

$$X_t = (a+bt) \cdot s_t \cdot \varepsilon_t$$

$\rightarrow$ Die Formeln entsprechend anpassen (s. Help zu “HoltWinters” in R).

Exponentielles Glätten nach Holt-Winters



Exponentielles Glätten nach Holt-Winters

```
#mit Saison
lynxn=ts(lynx, frequency = 10)
(hwls = HoltWinters(lynxn, beta=F))
ps=predict(hwls, n.ahead = 10)
plot(hwls,ps,  main="Exponentielle Glättung mit Saison")
```

Exponentielles Glätten nach Holt-Winters