

# **SBWL Tourismusanalyse und Freizeitmarketing, Vertiefungskurs 2**

## **1. Teil: Zerlegungsmodelle und naive Prognosemethoden für Zeitreihen**

Regina Tüchler

# Inhalt

---

## Einleitung

### 1. Einführung in das Modellieren von Zeitreihen

1.1. Zeitreihendaten und ihre grafische Darstellung

1.2. Das stochastische Modell

### 2. Zeitreihendatensätze für diese LV

### 3. Einführung in R

## UE. Übungen 1

# Inhalt

---

## 4. Empirische Momente von Zeitreihendaten

4.1. Autokorrelation, Autokorrelogramm

4.2. Stationarität

## 5. Empirische Momente in $\mathbb{R}$

# Inhalt

---

6. Das klassische Komponentenmodell

7. Zeitreihen ohne Trend und ohne Saisonalität

7.1. Gleitender Durchschnitt

7.2. Exponentielles Glätten

8. Einfache Filterverfahren in R

UE. Übungen 2

# Inhalt

---

## 9. Zeitreihen mit Trend

9.1 Lineare und nichtlineare Trends

9.2. Transformation der Daten

9.3. Holt Verfahren

## 10. Holt Verfahren in R

## UE. Übungen 3

# Inhalt

---

## 11. Zeitreihen mit Trend und Saisonalität

11.1. Methode der kleinen Trends

11.2. Saison-Trend Zerlegung mittels gleitendem Durchschnitt

11.3. Holt-Winters Verfahren

## 12. Saison-Trend Modellierung in R

UE. Übungen 4

# Einleitung

---

Im Mittelpunkt des ersten Teils dieser LV steht die Analyse und Prognose von Zeitreihen (ZR) mit Hilfe von sogenannten naiven Methoden. Dabei wird nur die Information aus der einen ZR im Verfahren verwendet. Es werden verschiedene Strukturen dieser ZR wie z.B. Trend oder Saison modelliert, um die zukünftige Entwicklung vorherzusagen.

Beispiel: Anhand von Tourismusdaten über monatliche Nächtigungszahlen in Österreich für die Jahre 1995-2003 sollen die Werte für 2004 vorhergesagt werden.

# Einleitung

---

Auf die Anwendung und Demonstration der Methoden anhand von Beispielen wird in dieser LV großer Wert gelegt. Damit die Studierenden, die Methoden selber anwenden können, wird die Open-Source Software R verwendet.

Zu jedem Thema gibt es einen theoretischen Teil, der immer anhand von Beispielen erklärt wird. Es folgt dann ein Kapitel, in dem gezeigt wird, wie die Verfahren mit R durchgeführt werden können. Am Schluss von jedem Thema gibt es ein Kapitel mit Übungsbeispielen, die den Studierenden helfen sollen, den Stoff selbständig zu üben.

# 1.1. Zeitreihendaten und ihre grafische Darstellung

## **Daten**

Entstehen durch Beobachtung von Variablen (Merkmalen)

## **Zeitreihendaten**

Beobachtung eines Merkmals zu verschiedenen Zeitpunkten  
z.B. tägliche Temperaturmessung, monatliche Nächtigungszahlen in einem Hotel

## **Querschnittsdaten**

Beobachtung eines Merkmals an verschiedenen Untersuchungseinheiten aus einer Grundgesamtheit  
z.B. Bewertung eines Produkts durch Konsumenten in einer Marktforschungsstudie

# 1.1. Zeitreihendaten und ihre grafische Darstellung

## Eigenschaften von Zeitreihendaten

- Messung eines Merkmals zu verschiedenen Zeitpunkten
- feste Ordnung der Beobachtungen nach der Zeit
- Beobachtungen sind in der Regel nicht unabhängig
- zu jedem Zeitpunkt liegt nur eine Beobachtung vor

# 1.1. Zeitreihendaten und ihre grafische Darstellung

## Zeitreihenplot

Er ist der Ausgangspunkt einer statistischen Analyse. Mit Hilfe dieser grafischen Darstellung kann man einen Eindruck über die Struktur der Daten gewinnen. Daher ist der Zeitreihenplot der wichtigste Ausgangspunkt, wenn man ein geeignetes Modell finden möchte.

Beispiele:

In Abb. 1 sehen wir einen Zufallsprozess. Die Beobachtungen sind hier unabhängig.<sup>1</sup>

---

<sup>1</sup>Sie kommen aus einer Normalverteilung mit Mittelwert 0, Varianz 4.

## 1.1. Zeitreihendaten und ihre grafische Darstellung

In der ZR aus Abb. 2 ist deutlich ein linearer Trend erkennbar. In Abb. 3 sehen wir eine ZR mit einem quadratischen Trend, in Abb. 4 weist die ZR eine deutliche Saisonalität auf und in Abb. 5 ist sowohl ein Trend als auch eine Saisonalität enthalten. Wenn man die ZR aus Abb. 2-5 modellieren will, müssen die jeweils speziellen Strukturen der ZR im Modell berücksichtigt werden. Wir werden daher Modelle kennen lernen, die Trend und/oder Saisonalität enthalten.

## 1.1. Zeitreihendaten und ihre grafische Darstellung

In Abb. 6 sehen wir eine ZR, die ihre Struktur ab dem Jahr 1983 deutlich ändert. Wenn man so eine ZR analysiert, ist es daher angebracht zusätzliche Hintergrundinformation, die den Bruch erklären kann, zu sammeln. In diesem Beispiel ist 1983 die Gurtenpflicht eingeführt worden. Mit solchen Reihen muß man geeignet umgehen. Je nach Anwendung kann man die Daten in 2 Datensätze unterteilen oder ein Regressionsmodell verwenden, das auch den Parameter über die Gurtenpflicht enthält (2. Teil dieser LV) oder eigene Strukturbruchmodelle verwenden.

## 1.1. Zeitreihendaten und ihre grafische Darstellung

In Abb. 7 sind einige untypische Datenpunkte zu beobachten, sogenannte Ausreißer. Die Ausreißer könnten die Analyse verzerren und es ist daher auch hier sinnvoll, zusätzliche Informationen einzuholen. Sind diese Ausreißer aus Anwendersicht erklärbar? Kann es sich um Messfehler handeln? Dann wären Korrekturmöglichkeiten je nach Anwendung z.B. Ersetzen durch Schätzwert oder Weglassen der Ausreißer (bei Querschnittdaten).

# 1.1. Zeitreihendaten und ihre grafische Darstellung

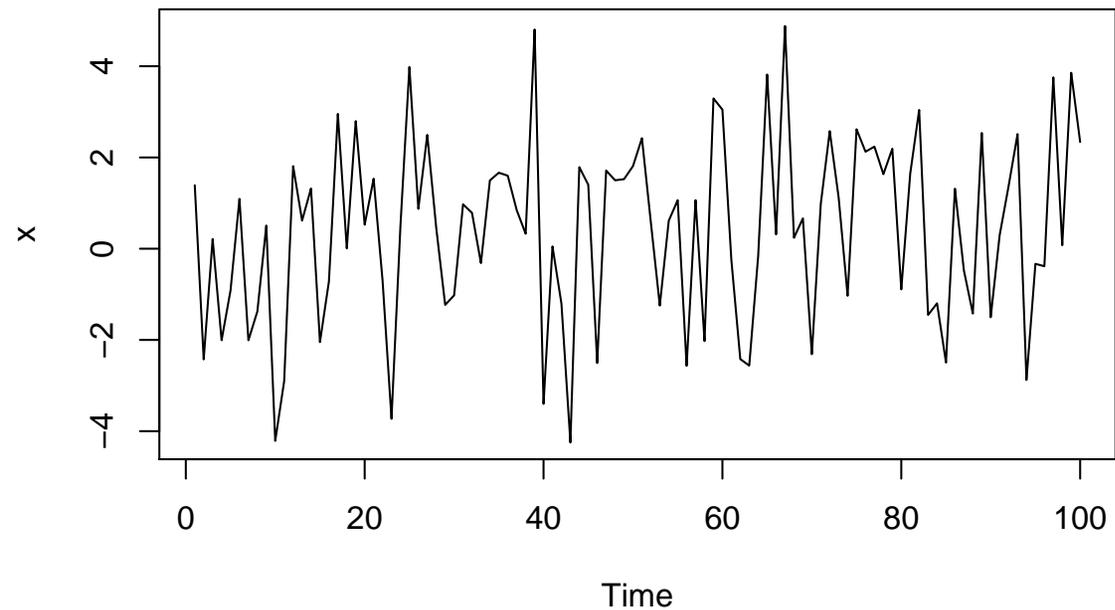


Abbildung 1: Zufallsprozess.

# 1.1. Zeitreihendaten und ihre grafische Darstellung

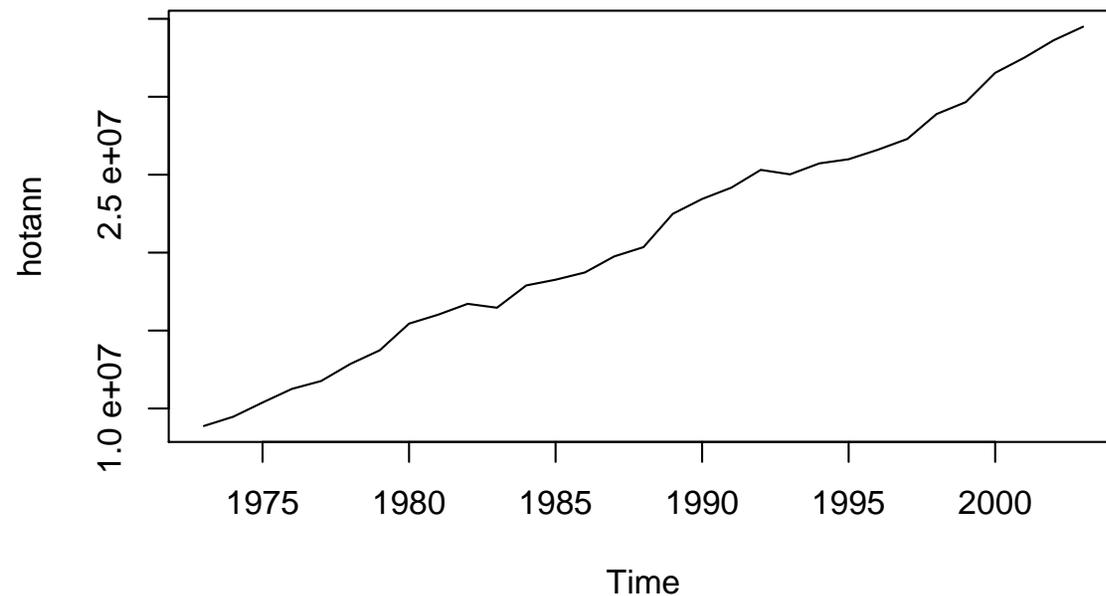


Abbildung 2: Jährliche Nächtigungen in 4/5 Sterne Hotels (A).

# 1.1. Zeitreihendaten und ihre grafische Darstellung

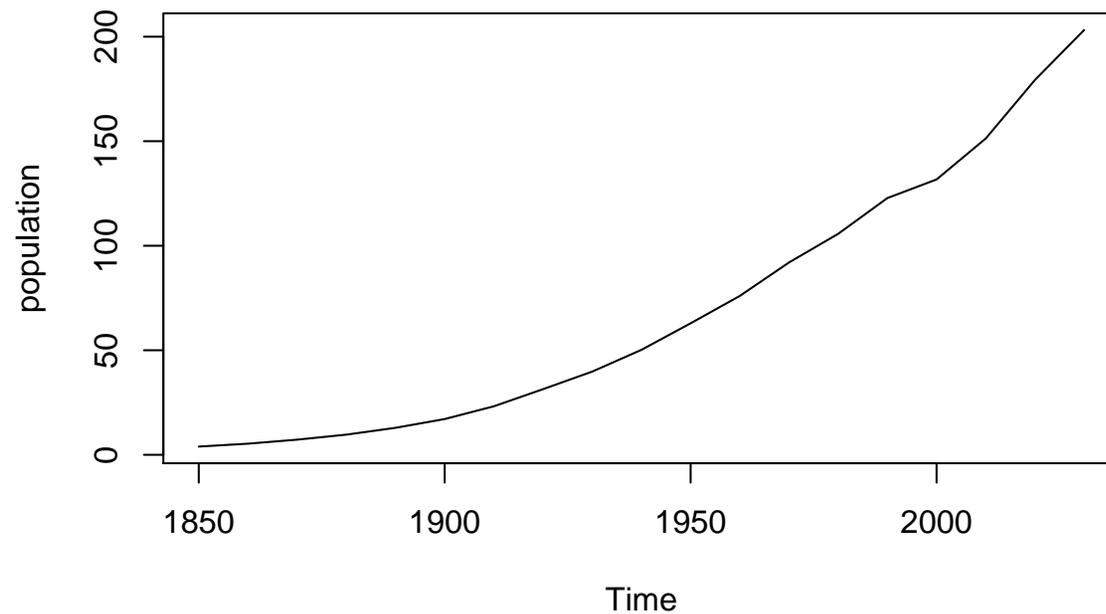


Abbildung 3: US Bevölkerungszahlen.

# 1.1. Zeitreihendaten und ihre grafische Darstellung

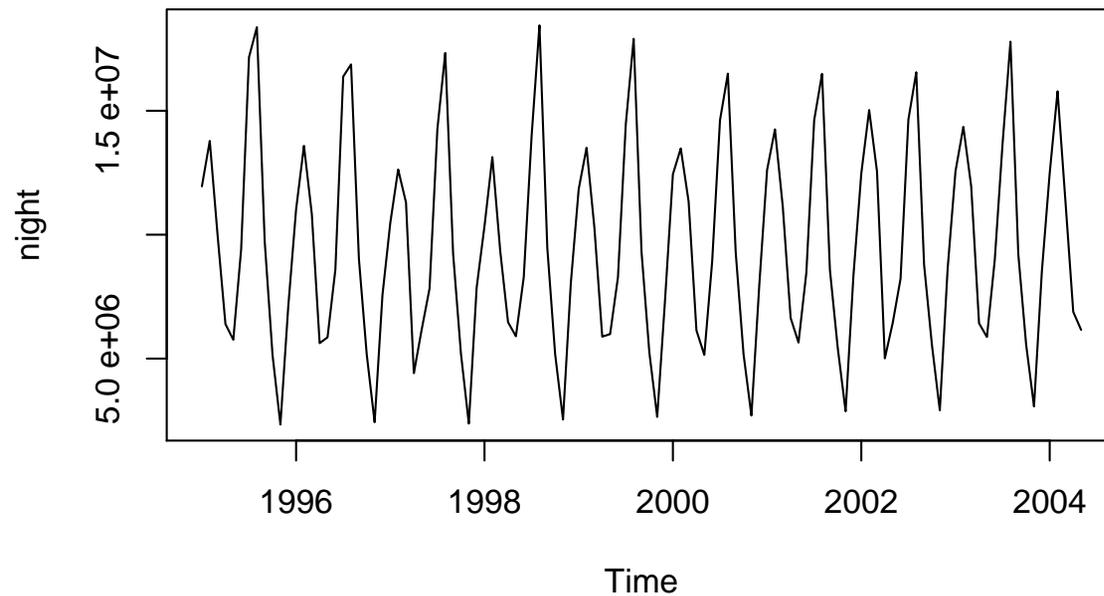


Abbildung 4: Monatliche Nächtigungen in Österreich.

# 1.1. Zeitreihendaten und ihre grafische Darstellung

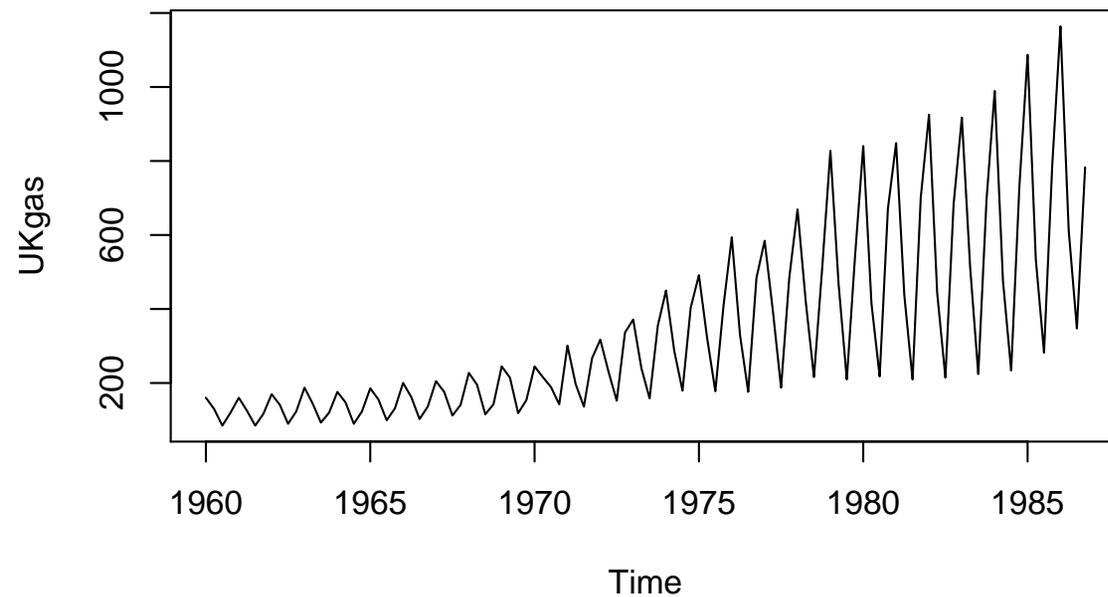


Abbildung 5: Erdgasverbrauch in GB.

# 1.1. Zeitreihendaten und ihre grafische Darstellung

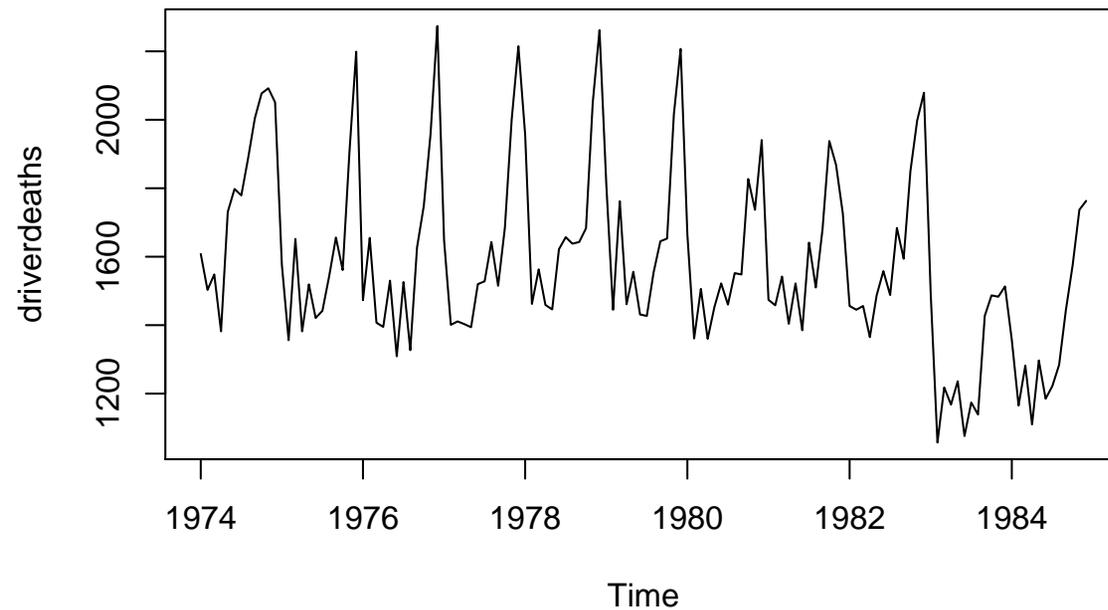


Abbildung 6: Verkehrstote bzw. Schwerverletzte in GB.

# 1.1. Zeitreihendaten und ihre grafische Darstellung

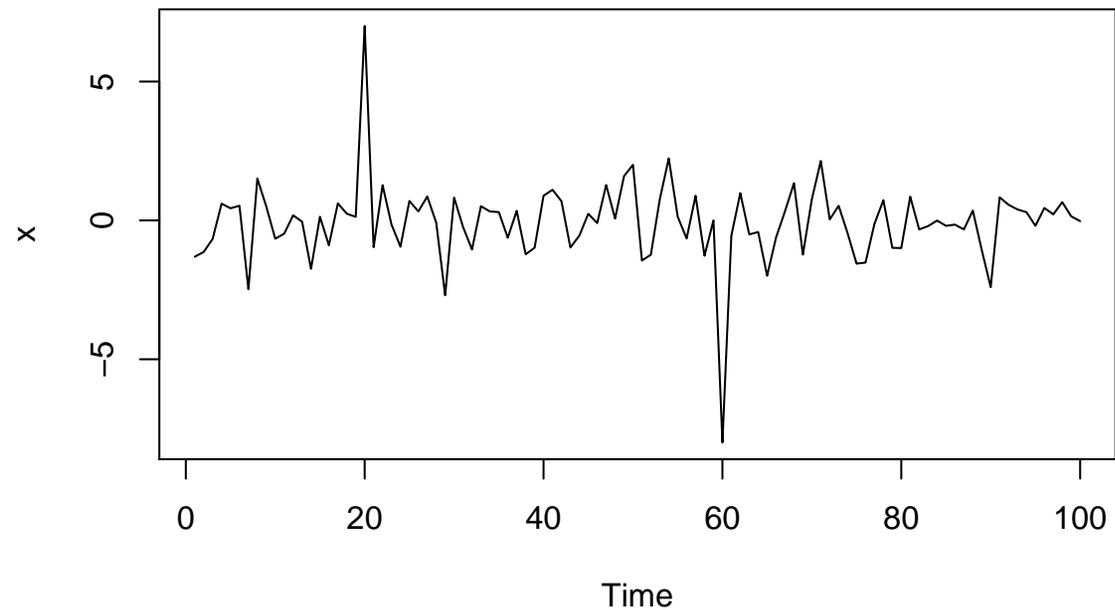


Abbildung 7: ZR mit Ausreißern.

## 1.2. Das stochastische Modell

---

### Das deterministische Modell

Exakte quantitative Beziehung zwischen Variablen (Merkmalen)

z.B. deterministische Nachfrage/Preis Relation:

$$N = \beta_1 + \beta_2 p \text{ oder } N = \beta_1 p^{\beta_2}$$

### Das stochastische Modell

Durch zufallsabhängigen Fehlerterm  $\varepsilon$  verschleierte quantitative Beziehung zwischen den Variablen

z.B. stochastische Nachfrage/Preis Relation:

$$N = \beta_1 + \beta_2 p + \varepsilon \text{ oder } N = \beta_1 p^{\beta_2} \varepsilon$$

## 2. Zeitreihendatensätze für diese LV

---

Überblick über die Datensätze, die im ersten Teil dieser LV verwendet werden:

**Nil-Daten:** LV-Seite: "nil.rda"

Zeitreihe mit dem Wasserstand des Nils bei Assuan für die Jahre 1901 bis 1970.

**Approve-Daten:** LV-Seite: "approve.rda"

Ausschnitt aus dem R-Datensatz `presidents`, über die Zustimmung der US Bürger zu ihrem Präsidenten für die Jahre 1953-1971 (Quartalsdaten).

## 2. Zeitreihendatensätze für diese LV

---

**Hotel-Daten:** LV-Seite: "hotelts-annual.rda"

Daten über die Anzahl an Übernachtungen in österreichischen 4/5-Sterne-Hotels pro Jahr für den Zeitraum 1973-2003.

**US Bevölkerungszahlen:** In R: data(uspop)

Einwohnerzahlen der USA 1790-1970 (Zehnjahresdaten).

**Austres:** In R: data(austres)

Einwohnerzahlen Australiens für 2. Quartal 1971 - 2. Quartal 1993.

**Nächtigungsdaten:** LV-Seite: "nightts-monthly.rda"

Die ZR enthält Tourismusdaten über die Anzahl an Übernachtungen in allen österreichischen Unterbringungsformen pro Monat (Zeitraum Jän. 1995 - Mai 2004).

## 2. Zeitreihendatensätze für diese LV

---

**UKgas:** In R: `data(UKgas)`

Erdgaskonsum in UK für 1960-1986 (Quartalsdaten).

**CO<sub>2</sub>-Daten:** In R: `data(co2)`

Datensatz mit monatlichen CO<sub>2</sub>-Messungen für 1959-1997.

**USAccDeaths:** In R: `data(USAccDeaths)`

Daten über monatliche tödliche Unfälle in den USA in den Jahren 1973-1978.

### 3. Einführung in R

---

Um die Methoden dieser LV auf dem Computer zu rechnen, werden wir das Programmpaket R verwenden. Die Homepage zu dieser open source software ist:

<http://www.R-project.org/>

In "EinfirR.pdf" von der LV-Seite stehen Anleitungen für die ersten Schritte in R, wie wir sie in dieser LV brauchen werden.

# Übungen 1

---

1. "EinfirR.pdf": Installieren Sie R auf ihrem PC und probieren Sie die dortigen Befehle aus.
2. Laden Sie die Nil-Daten von der LV-Seite. (Befehl: `load("nil.rda")`). Die Daten müssen dafür im Working Directory abgespeichert worden sein.) Zeichnen Sie einen Zeitreihenplot.

## 4. Empirische Momente von ZR

---

Im Allgemeinen wird sich das Verhalten einer ZR über die Zeit ändern, z.B. Trend, saisonales Verhalten, sich verändernde Variabilität.

Eine ZR wird als **stationär** bezeichnet, wenn sich ihr Verhalten über die Zeit nicht systematisch ändert. Dies ist die intuitive Idee von Stationarität, eine mathematische Definition folgt später.

Beispiele: In Abb. 1 ist eine stationäre ZR gegeben. Die ZR in Abb. 2-5 ändern ihr Verhalten im Laufe der Zeit und sind daher nicht stationär.

## 4. Empirische Momente von ZR

---

In der Grundvorlesung Statistik haben wir den (empirischen) **Erwartungswert** kennen gelernt. Er wurde dort für unabhängige Realisationen  $x_1, x_2, \dots, x_T$  einer ZG  $X$  als Durchschnitt über die Datenpunkte definiert:

$$\bar{x} = \frac{1}{T} \sum_{t=1}^T x_t.$$

z.B. Körpergröße der 30 Kinder einer Schulklasse: die 30 Messungen sind Realisationen von der Zufallsvariablen "Körpergröße",  $\bar{x}$  ist die durchschnittliche KG.

## 4. Empirische Momente von ZR

---

### Erwartungswert bei ZR:

Bei ZR liegen nun keine voneinander unabhängigen Größen vor. Das Verhalten der ZR kann sich systematisch ändern.

Der Erwartungswert  $\bar{x}$  kann nur bei stationären ZR sinnvoll die Lage der Daten angeben, nicht aber bei systematischen Änderungen, wie z.B. Trend. So besitzt die durchschnittliche Zahl an Übernachtungen für die Hotel-Daten (21 121 178) keine Aussagekraft für den Verlauf dieser ZR.

## 4. Empirische Momente von ZR

---

Die (empirische) **Varianz** von unabhängige Realisationen  $x_1, x_2, \dots, x_T$  einer ZG  $X$  wurde definiert als:

$$s_x^2 = \frac{1}{T} \sum_{t=1}^T (x_t - \bar{x})^2.$$

Auch die Varianz kann nur bei stationären ZR sinnvoll als Maßzahl für die Streuung der Daten verwendet werden, nicht aber bei systematischen Änderungen.

## 4.1. Autokorrelation, Autokorrelogramm

---

Da bei einer ZR die einzelnen Datenpunkte über die Zeit im Allgemeinen abhängig sind, definieren wir im folgenden eine Maßzahl für diese Abhängigkeit. Es geht dabei darum, den Zusammenhang zwischen der ZR zum Zeitpunkt  $t$  und vergangenen Zeitpunkten  $t - 1, t - 2, \dots$  zu messen .

Die (empirische) **Autokovarianz** einer stationären ZR ist definiert als:

$$c_\tau = \frac{1}{T} \sum_{t=1}^{T-\tau} (x_t - \bar{x})(x_{t+\tau} - \bar{x}).$$

Sie ist für Verzögerungen (Lags)  $\tau = 0, 1, \dots, T - 1$  definiert.  $c_0$  ist gleich der Varianz der Daten.

## 4.1. Autokorrelation, Autokorrelogramm

---

Die **Autokovarianzfunktion** ordnet jedem Lag  $\tau = 0, 1, \dots, T - 1$  den Wert der Autokovarianz  $c_\tau$  zu.

Die Autokovarianz ist eine Maßzahl, die von der Skala abhängt. Im folgenden wird daher eine Maßzahl definiert, die skalenunabhängig ist.

Die (empirische) **Autokorrelation** kann aus den Autokovarianzen berechnet werden:

$$r_\tau = c_\tau / c_0$$

Die **Autokorrelationsfunktion** ordnet jedem Lag  $\tau = 0, \dots, T - 1$  den Wert der Autokorrelation zu.

Das **Autokorrelogramm** zeichnet die Autokorrelationsfunktion.

## 4.1. Autokorrelation, Autokorrelogramm

---

Eigenschaften der Autokorrelation<sup>2</sup>:

- $r_0 = 1$
- $-1 \leq r_\tau \leq 1$
- Der Betrag von  $r_\tau$  misst die Stärke des Zusammenhangs.
- Das Vorzeichen gibt die Richtung des Zusammenhangs an:  $r_\tau > 0$  bedeutet, dass auf große Werte von  $x_t$  auch große Werte von  $x_{t+\tau}$  folgen (analog für kleine). Bei  $r_\tau < 0$  folgen auf große  $x_t$  kleine  $x_{t+\tau}$  und umgekehrt.

---

<sup>2</sup>Vgl. Korrelationskoeffizient in der Grundvorlesung Statistik

## 4.1. Autokorrelation, Autokorrelogramm

---

### Anwendung des Autokorrelogramms:

- Neben dem Zeitreihenplot ist das Autokorrelogramm eine wichtige grafische Darstellung, um ein adäquates Modell für eine ZR zu finden. Es wird dabei für die Rohdaten (also auch für nicht stationäre ZR) eingesetzt. Am Muster kann man Hinweise für die Wahl des Modells bekommen.

## 4.1. Autokorrelation, Autokorrelogramm

---

Beispiele: In Abb. 2 hatten wir ein Beispiel für eine trendbehaftete ZR. Das dazugehörige Autokorrelogramm in Abb. 8 zeigt diesen Trend auch. Man sieht hier die positiven Autokorrelationen, die abklingen und erst ab einem Lag von 7 nicht mehr signifikant sind.

In Abb. 9 sieht man die Saisonalität der ZR aus Abb. 4 auch im Autokorrelogramm.

## 4.1. Autokorrelation, Autokorrelogramm

---

- Prüfung auf Unkorreliertheit:  
ZR, deren Zeitpunkte nicht miteinander korreliert sind, spielen bei der Modellbildung eine wichtige Rolle. Für empirische Daten wird die Autokorrelation in der Regel von 0 verschieden sein, auch wenn tatsächlich keine Korrelation zwischen den Zeitpunkten besteht. In diesem Fall kann man mit 95% Sicherheit davon ausgehen, dass die Autokorrelationen  $r_\tau$  zwischen  $\pm 2/\sqrt{T}$  liegen. Diese Grenzen können ins Autokorrelogramm eingezeichnet werden, sodass Autokorrelationen, die diese Grenze überschreiten, leicht erkannt werden können.<sup>3</sup>

---

<sup>3</sup>Erinnern sie sich dazu an die Durchführung von statistischen Tests aus der Grundvorlesung Statistik

## 4.1. Autokorrelation, Autokorrelogramm

---

Beispiel: In Abb. 1 hatten wir einen Zufallsprozess mit voneinander unabhängigen Zeitpunkten. Im dazugehörige Autokorrelogramm, Abb. 10 liegen daher auch alle Autokorrelationen innerhalb der eingezeichneten Signifikanzgrenzen.

## 4.1. Autokorrelation, Autokorrelogramm

---

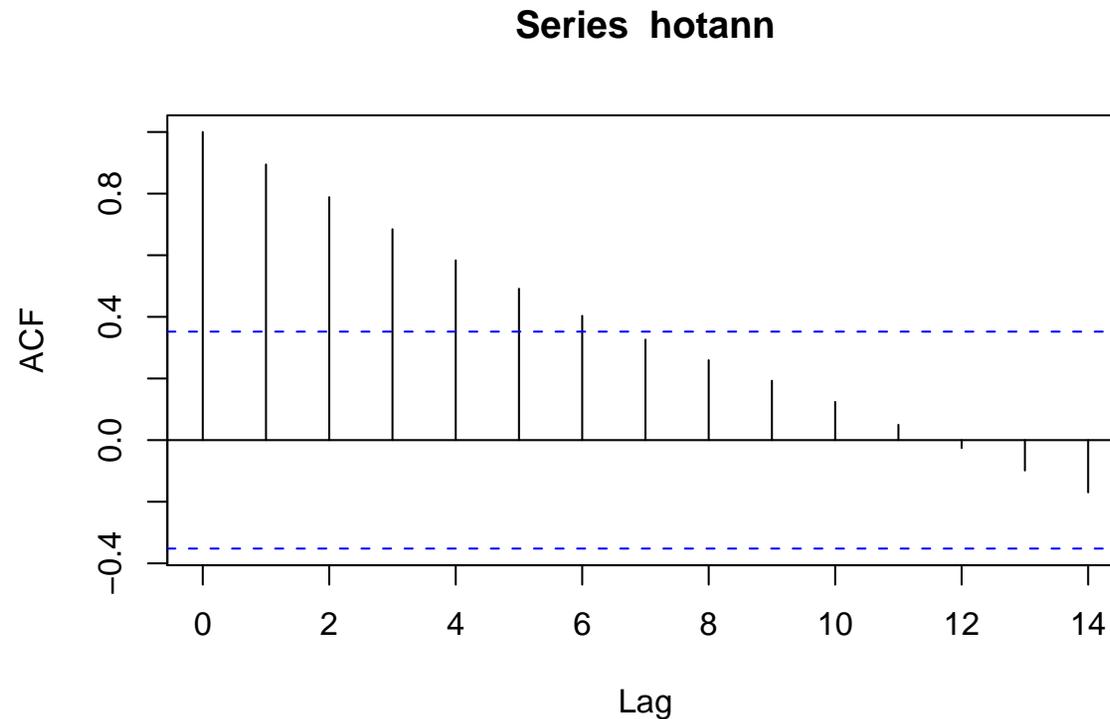


Abbildung 8: Autokorrelogramm der Hotel-Daten

## 4.1. Autokorrelation, Autokorrelogramm

---

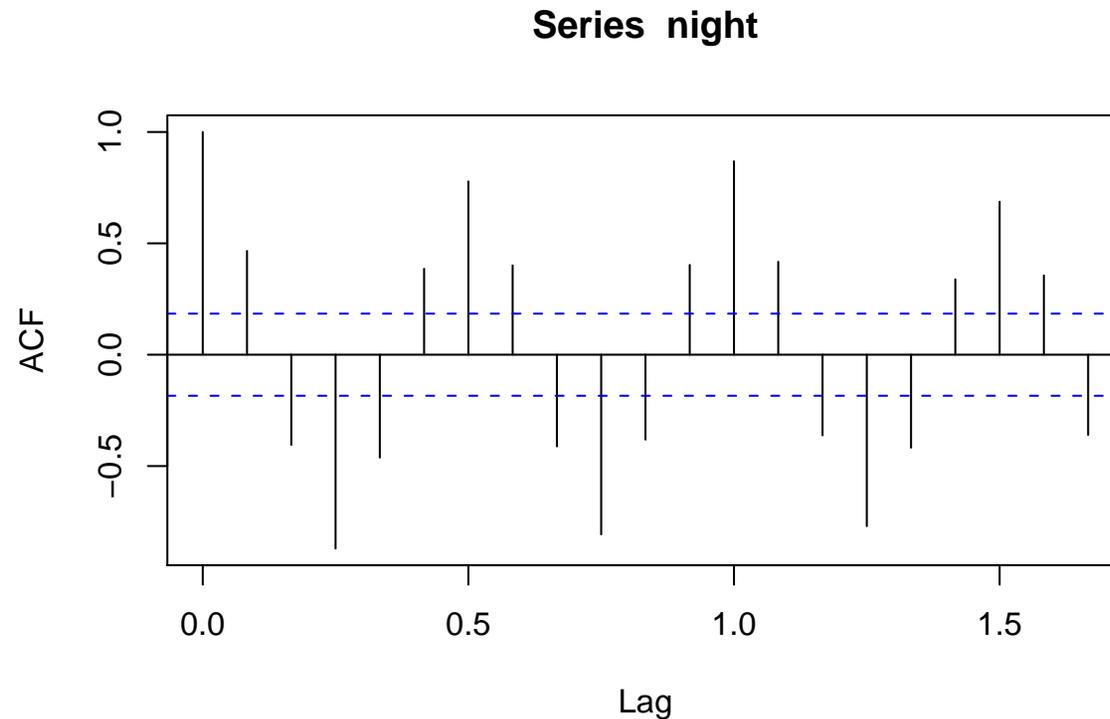


Abbildung 9: Autokorrelogramm der Nächtigungs-Daten

## 4.1. Autokorrelation, Autokorrelogramm

---

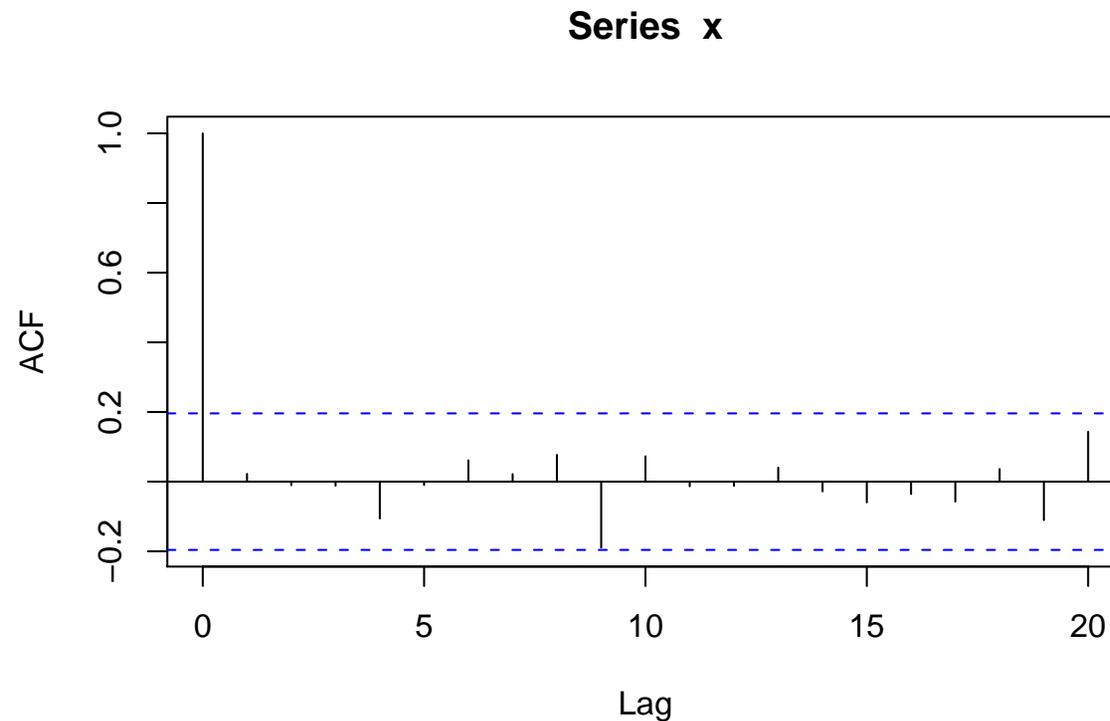


Abbildung 10: Autokorrelogramm d. Zufallsprozesses (Abb.1).

## 4.2. Stationarität

---

**Ein stochastischer Prozess** ist eine Folge von Zufallsvariablen  $X_t$  für Zeitpunkte  $t = 1, 2, \dots$ . Die konkreten ZR sind endlich viele Realisationen von diesem Zufallsprozess:  $x_1, \dots, x_T$ .

### **Mathematische Definition von Stationarität:**

Ein stochastischer Prozess  $X_t$  ist (schwach) stationär, wenn

- $E(X_t) = \mu$  von  $t$  unabhängig ist
- $Cov(X_t, X_{t+s}) = \gamma_X(s)$  von  $t$  unabhängig,  $s = 0, \pm 1, \pm 2, \dots$

Erwartungswert und Varianz sind also endlich und konstant über die Zeit.<sup>4</sup>

---

<sup>4</sup>streng stationäre Prozesse haben identische Verteilung unabhängig von  $t$

## 4.2. Stationarität

---

Bei der Modellbildung werden wir oft einen stationären Zufallsprozess  $\varepsilon_t$  verwenden, für den gilt:

- $E(\varepsilon_t) = 0$
- 

$$\gamma_\varepsilon(\tau) = \text{Cov}(\varepsilon_t, \varepsilon_{t+\tau}) = \begin{cases} \sigma^2 & \tau = 0 \\ 0 & \tau = \pm 1, \pm 2, \dots \end{cases}$$

D.h. die Autokorrelationen sind 0.

Ein Beispiel für so einen stationärer Zufallsprozess hatten wir in Abb. 1 und 10.

## 5. Empirische Momente in R

---

In "EmpirMomenteinR.pdf" wird beschrieben, wie man die empirischen Momente in R berechnet und wie man ein Autokorrelogramm zeichnet.

## 6. Das klassische Komponentenmodell

---

Gegeben sei eine ZR  $x_t$  für die Zeitpunkte  $t = 1, \dots, T$ . Im **additiven klassischen Komponentenmodell** wird sie folgendermaßen zerlegt:

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

$\hat{m}_t$  ist die Trendkomponente für die langfristige systematische Änderung des Niveaus.

$\hat{s}_t$  ist die Saisonkomponente. Man nimmt hier eine Periodenlänge  $p$  an, z.B.  $p = 12$  bei Monatsdaten.

Weiters gilt, dass sich die einzelnen Saisonkomponenten über die Jahre nicht verändern:

$$\hat{s}_{t+p} = \hat{s}_t$$

## 6. Das klassische Komponentenmodell

---

und dass die Summe über alle Saisonkomponenten 0 ist:

$$\sum_{k=1}^p \hat{S}_k = 0.$$

In den Saisonkomponenten ist daher keine Information über das Niveau der ZR enthalten, sondern nur über die saisonale Schwankung um dieses Niveau.

## 6. Das klassische Komponentenmodell

---

$\varepsilon_t$  ist die Restkomponente (Residuen). Diese ist eine zufällige Störkomponente, die übrig bleibt, wenn man von der ursprünglichen ZR die Trend- und die Saisonkomponente abzieht. Sie entspricht daher der um Trend und Saison bereinigten ZR:

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

Die Störung  $\varepsilon_t$  ist stationär und nicht autokorreliert. Wir nehmen im folgenden an, dass das Niveau der ZR in der Trendkomponente  $m_t$  enthalten ist. Der Erwartungswert von  $\varepsilon_t$  ist dann gleich 0, die Varianz ist konstant. D.h. in der Restkomponente sind keine systematischen Einflüsse mehr enthalten.

## 6. Das klassische Komponentenmodell

---

In den nächsten Abschnitten werden wir Methoden zur Bestimmung der Trend- und die Saisonkomponenten kennen lernen. Die Methode soll so gewählt werden, dass die Restkomponente keine systematischen Einflüsse enthält.

Wir werden bei diesen Verfahren die ZR  $x_t, t = 1, \dots, T$  glätten. D.h. irreguläre Schwankungen werden weggefiltert.

Ziel der Verfahren ist es, **Prognosen** über das zukünftige Verhalten der ZR zu erstellen. Wir werden in den folgenden Abschnitten die Prognosefähigkeit der Verfahren für verschiedenen Szenarien untersuchen.

## 6. Das klassische Komponentenmodell

---

### Vorgangsweise bei der Analyse von ZR:

1. **Schritt:** Zeitreihenplot der Daten
2. **Schritt:** Welche Muster sind erkennbar? Gibt es Trend, Saison? Sind Ausreißer oder Brüche erkennbar? Sollen die Daten transformiert werden? Welche Methoden können gewählt werden?
3. **Schritt:** Modellschätzung, Ziel sind stationäre Residuen
4. **Schritt:** Modellwahl, Modellprüfung und Residuenanalyse

Falls 4. nicht zufriedenstellend:

5. **Schritt:** andere Methode und Vergleich der Methoden

## 6. Das klassische Komponentenmodell

---

Je nachdem welche Gestalt eine ZR hat, kann das Komponentenmodell vereinfacht werden.

Abschnitt 7: Es wird nur das Niveau der ZR modelliert.

Abschnitt 9: Es wird nur die Trendkomponente modelliert.

Abschnitt 11: Es werden Trend- und Saisonkomponenten modelliert.

## 7. Zeitreihen ohne Trend und Saisonalität

---

In diesem Abschnitt wird das Niveau einer ZR modelliert. Wir gehen dabei von einem Modell aus, das die ZR in ein Niveau und eine zufällige Restkomponente aufgeteilt.

Wir werden dazu zunächst lineare Filter definieren. In Abschnitt 7.1. werden wir einen Spezialfall eines linearen Filters, die Methode des gleitenden Durchschnitts, kennen lernen und ihre Glättungseigenschaften untersuchen.

Das Verfahren des exponentiellen Glättens ist ebenfalls ein spezieller linearer Filter und ist für die Prognose von ZR, die keinen Trend und keine Saisonalität enthalten, geeignet (Abschnitt 7.2.).

## 7. Zeitreihen ohne Trend und Saisonalität

---

### Die Nil-Daten:

In Abb. 11 ist ein Zeitreihenplot über den Wasserstand des Nils bei Assuan für die Jahre 1901 bis 1970 gezeichnet. Diese Daten weisen keinen Trend und keine Saisonalität auf. Ein passendes Prognosemodell muss daher das zukünftige Niveau vorhersagen. Daher ist hier ein Modell geeignet, das zufällige Schwankungen um ein Niveau enthält.

## 7. Zeitreihen ohne Trend und Saisonalität

---

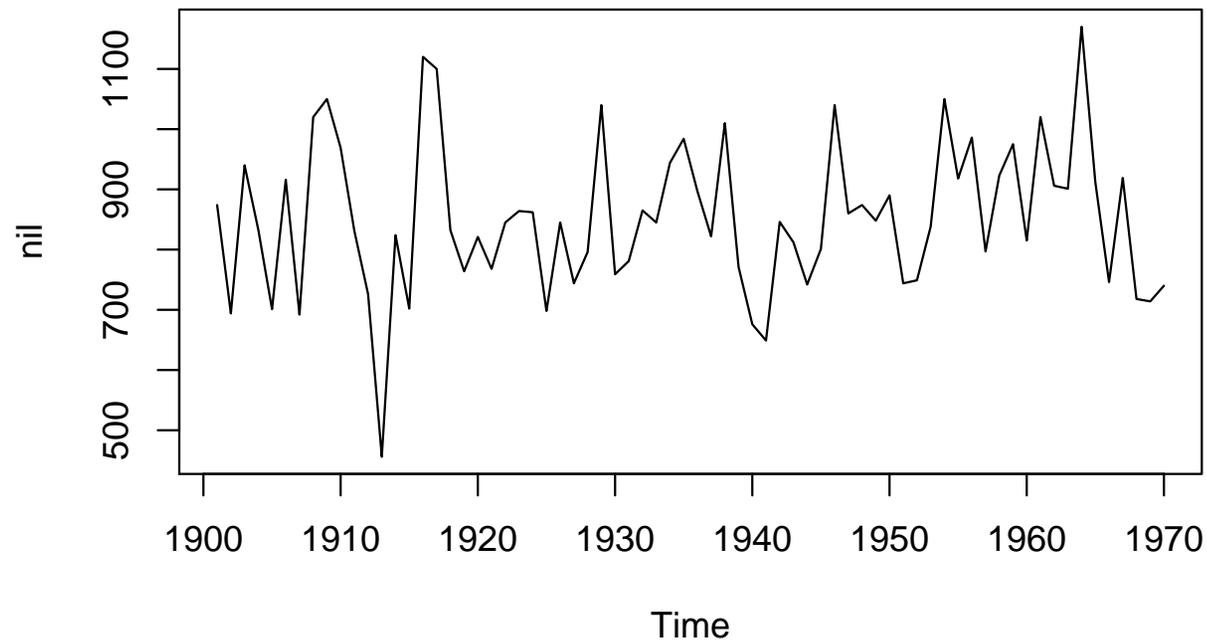


Abbildung 11: Die Nil-Daten.

## 7.1. Gleitender Durchschnitt

---

### Linearer Filter:

Ein linearer Filter transformiert eine ZR  $(x_t)$ ,  $t = 1, \dots, T$  in eine gefilterte ZR  $(y_t)$ :

$$y_t = a_{-q}x_{t+q} + a_{-q+1}x_{t+q-1} + \dots + a_{s-1}x_{t-s+1} + a_sx_{t-s}$$
$$t = s + 1, \dots, T - q$$

$a_{-q}, \dots, a_s$  sind Konstante.

## 7.1. Gleitender Durchschnitt

---

### Einfacher gleitender Durchschnitt:

Er ist ein linearer Filter mit symmetrischem Zeitfenster  $[t - q, \dots, t + q]$ :

$$y_t = \sum_{u=-q}^q a_u x_{t-u}, \quad t = q + 1, \dots, T - q$$

mit den Gewichten

$$a_u = \frac{1}{2q + 1} \quad u = -q, \dots, q$$

## 7.1. Gleitender Durchschnitt

---

Ein einfaches Beispiel ist der Mittelwert aus allen Zeitpunkten  $x_t$  ( $1, \dots, T$ )<sup>5</sup>:

$$\bar{x} = \sum_{t=1}^T \frac{1}{T} x_t$$

Der einfache gleitende Durchschnitt ist also der Mittelwert über ein bestimmtes Zeitfenster.

---

<sup>5</sup>Für ungerades  $T$

## 7.1. Gleitender Durchschnitt

---

### **Einfacher gleitender Durchschnitt der Nil-Daten:**

In Abb. 12 wird der einfache gleitende Durchschnitt für die Nil-Daten berechnet. Es werden dafür verschieden lange Zeitfenster verwendet. Man sieht deutlich, dass bei  $q = 5$  die Glättung am stärksten ist, während bei  $q = 1$  die Fluktuationen auch in der geglätteten Reihe noch viel stärker erkennbar sind.

## 7.1. Gleitender Durchschnitt

---

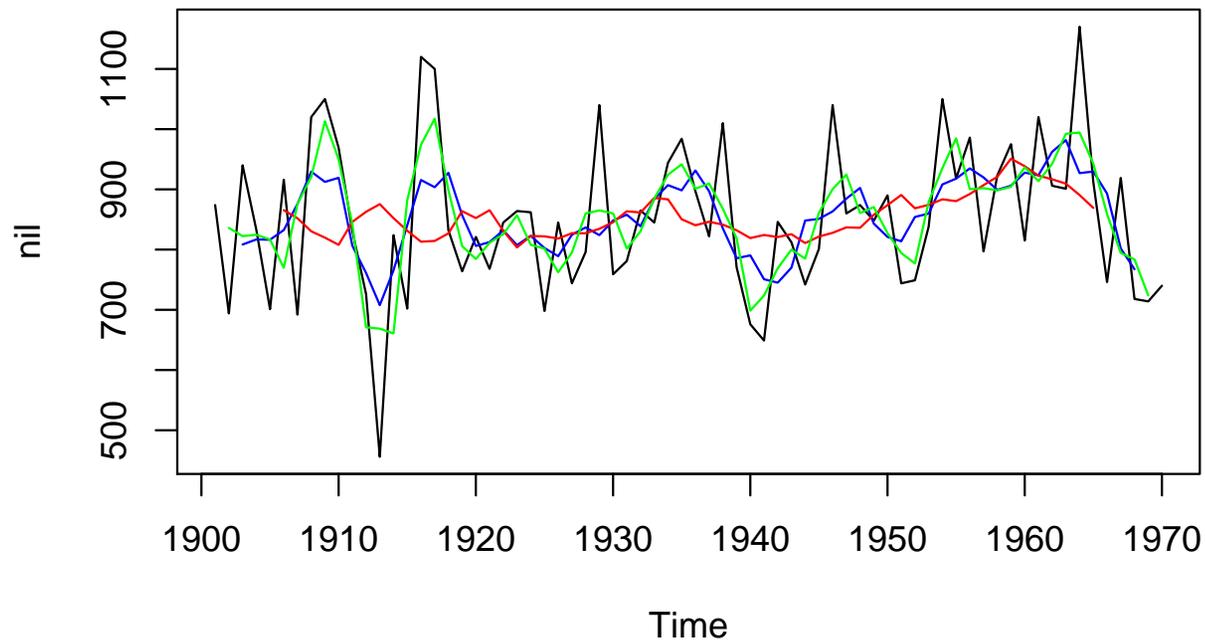


Abbildung 12: Einfacher gleitender Durchschn. für  $q = 1, 2, 5$ .

## 7.1. Gleitender Durchschnitt

---

### Zur Anwendung des einfachen gleitenden Durchschnitts:

- Das Verfahren schätzt ein lokales Niveau der Zeitreihe indem der Durchschnitt über die Werte in  $[t - q, t + q]$  gebildet wird.
- Die Glättung ist umso stärker, je mehr Terme einbezogen werden. Umso kürzer wird dann die gefilterte Reihe. Die Outputreihe hat um  $2 \cdot q$  Zeitpunkte weniger als die Inputreihe.
- Beim einfachen gleitenden Durchschnitt werden schnelle Fluktuationen herausgefiltert, während langfristige lineare Trendänderungen in der geglätteten ZR erhalten bleiben.

## 7.2. Exponentielles Glätten

---

Der einfache gleitende Durchschnitt ist ein symmetrisches Filterverfahren, bei dem um den Zeitpunkt  $t$  ein symmetrisches Zeitfenster  $[t - q, t + q]$  beim Glätten verwendet wird. Die Outputreihe wird daher an beiden Enden um  $q$  Werte kürzer.

Beim Verfahren des **exponentiellen Glättens** handelt es sich um ein rekursives Filterverfahren. Wir erhalten Outputwerte bis zum Endzeitpunkt  $T$  und können dieses Verfahren daher für eine Prognose über den Zeitpunkt  $T$  hinaus verwenden.

## 7.2. Exponentielles Glätten

---

Beim exponentiellen Glätten wird an die ZR ein lokales Niveau angepasst.

Es wird ein konstantes Niveau für zukünftige Zeitpunkte prognostiziert.

Das exponentielle Glätten ist daher zur Prognose von ZR geeignet, die keinen langfristigen Trend und keine saisonalen Schwankungen aufweisen.

## 7.2. Exponentielles Glätten

---

Gegeben sei die ZR  $(x_t)$  für die Zeitpunkte  $t = 1, \dots, T$ . Der geglättete Wert der ZR ist eine Einschnitt-Vorhersage und wird mit  $\hat{x}_t(1)$  bezeichnet.

Notation:  $\hat{x}_t(h)$  heißt, dass zum Zeitpunkt  $t$  eine Vorhersage des Wertes der ZR für  $h$  Zeitpunkte in die Zukunft gemacht wird, also für den Zeitpunkt  $t + h$ .

## 7.2. Exponentielles Glätten

---

Die Einschritt-Vorhersage zum Zeitpunkt  $t$  (Bez.:  $\hat{x}_t(1)$ ) ist also jener Wert, der für den Zeitpunkt  $t + 1$  vorhergesagt wird. Er entspricht dem gewichteten Mittel aus der aktuellen Beobachtung  $x_t$  und aus der Einschritt-Vorhersage für den Zeitpunkt  $t$ :  $\hat{x}_{t-1}(1)$ .

Rekursion:

$$\hat{x}_t(1) = \alpha x_t + (1 - \alpha)\hat{x}_{t-1}(1), \quad t = 1, \dots, T.$$

Mit Startwert  $\hat{x}_0(1) = x_1$ . Der Glättungsparameter  $\alpha$  liegt zwischen 0 und 1.

## 7.2. Exponentielles Glätten

---

Wir bezeichnen das für den Zeitpunkt  $t$  geschätzte, lokale Niveau der ZR mit  $\hat{a}_t$ . Dann gilt:

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

$\hat{a}_{T+1}$  ist dann das über den Zeithorizont der ZR hinaus geschätzte Niveau, und entspricht der Prognose  $\hat{x}_T(h)$ ,  $h = 1, 2, \dots$

## 7.2. Exponentielles Glätten

---

Das rekursive Verfahren:

$$\begin{array}{llll} \text{Startwert:} & \hat{x}_0(1) & = \hat{a}_1 & = x_1 \\ \text{Rekursion:} & \hat{x}_1(1) & = \hat{a}_2 & = \alpha x_1 + (1 - \alpha)\hat{x}_0(1) = x_1 \\ & \hat{x}_2(1) & = \hat{a}_3 & = \alpha x_2 + (1 - \alpha)\hat{x}_1(1) \\ & \vdots & & \\ & \hat{x}_{T-1}(1) & = \hat{a}_T & = \alpha x_{T-1} + (1 - \alpha)\hat{x}_{T-2}(1) \\ \text{Prognose f. } T+1: & \hat{x}_T(1) & = \hat{a}_{T+1} & = \alpha x_T + (1 - \alpha)\hat{x}_{T-1}(1) \\ \text{Prognose f. } T+h: & \hat{x}_T(h) & = \hat{a}_{T+1} & \end{array}$$

## 7.2. Exponentielles Glätten

---

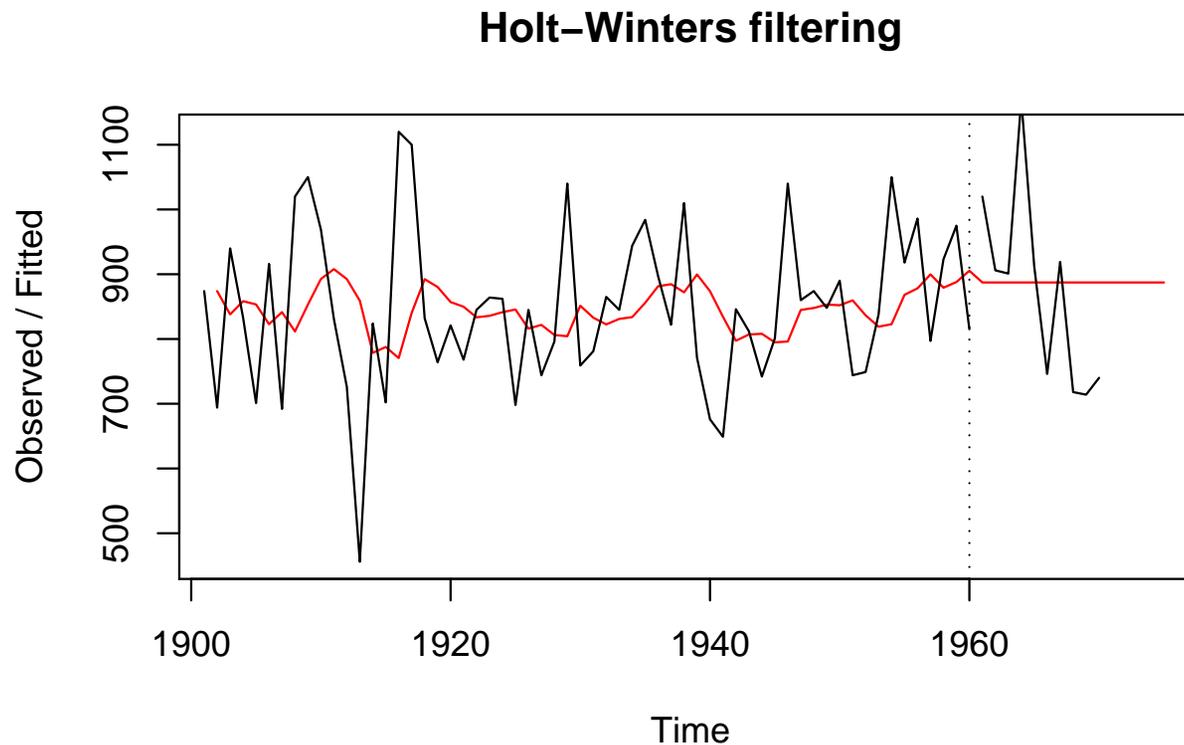


Abbildung 13: Exponentielles Glätten der Nil-Daten,  $\alpha = 0.2$ .

## 7.2. Exponentielles Glätten

---

### **Exponentielles Glätten der Nil-Daten:**

Wir haben Daten über den Wasserstand des Nils bei Assuan für die Jahre 1901 bis 1970.

Die Nil-Daten der Jahre 1901-1960 wurden mit exponentiellem Glätten modelliert. Die Jahre 1961-1975 wurden vorhergesagt. Da hier Daten vorliegen, die keinen systematischen Trend und keine saisonalen Schwankungen aufweisen, liefert dieses Verfahren eine sinnvolle Prognose.

## 7.2. Exponentielles Glätten

---

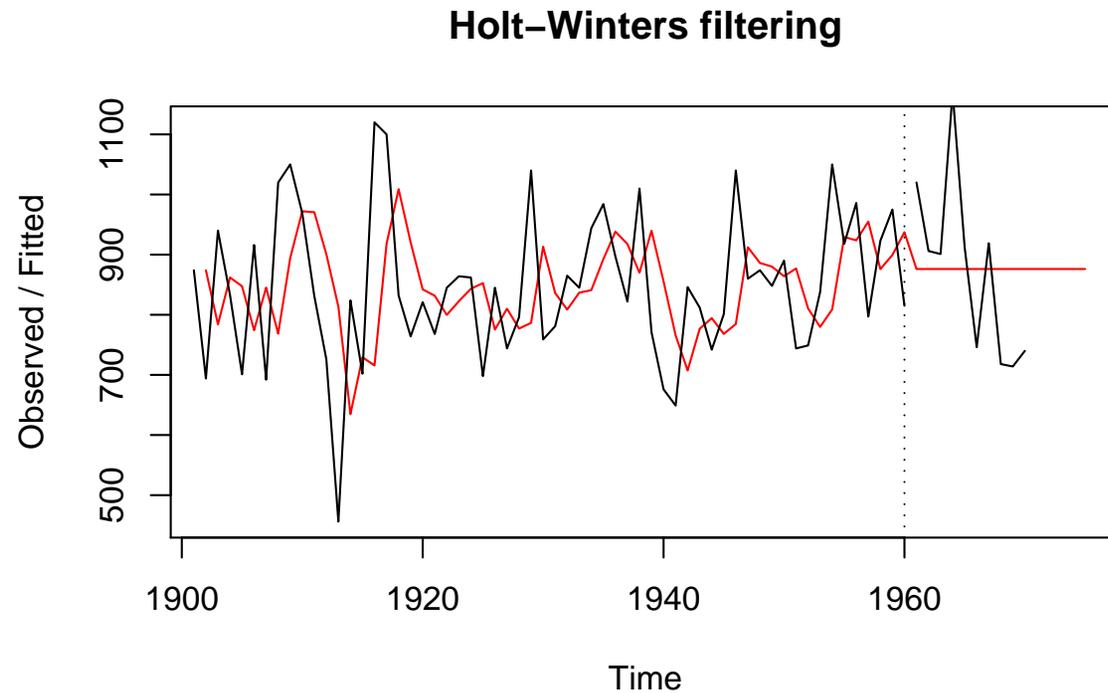


Abbildung 14: Exponentielles Glätten der Nil-Daten,  $\alpha = 0.5$ .

## 7.2. Exponentielles Glätten

---

### Wahl des Glättungsparameters $\alpha$ beim exponentiellen Glätten der Nil-Daten:

In Abb. 13 wurde  $\alpha = 0.2$  und in Abb. 14 wurde  $\alpha = 0.5$  gewählt. Man sieht, dass die Glättung für den kleineren Wert von  $\alpha$  viel stärker ist, denn da werden auch länger zurückliegende Werte stärker berücksichtigt.

## 7.2. Exponentielles Glätten

---

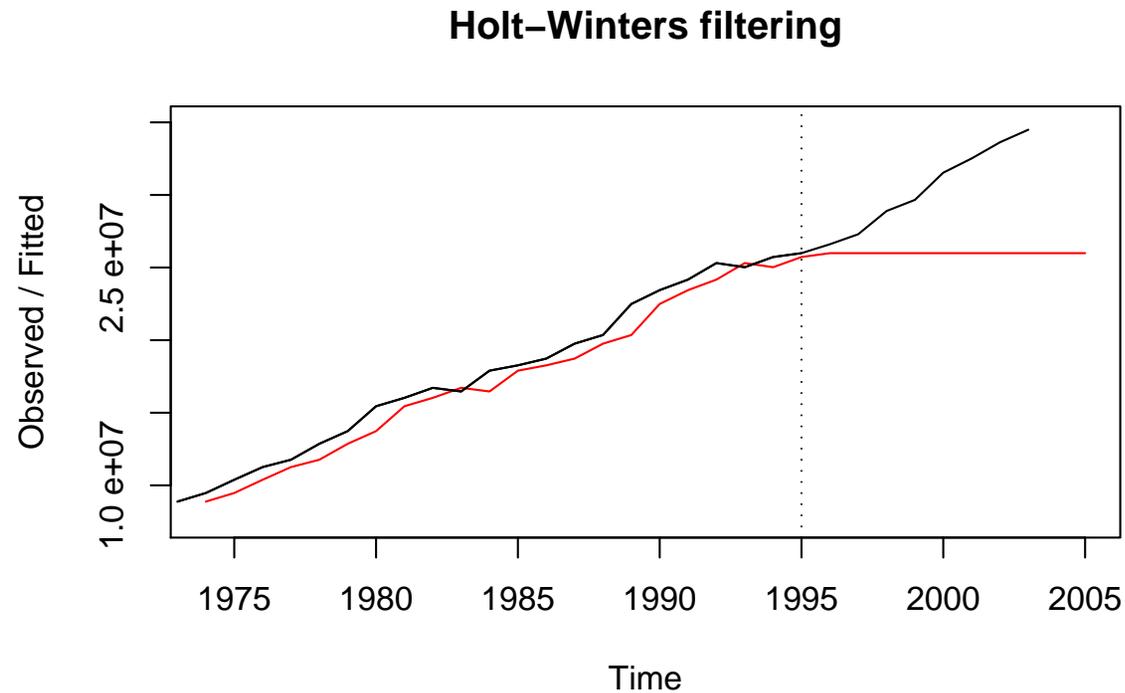


Abbildung 15: Exponentielles Glätten der Hotel-Daten.

## 7.2. Exponentielles Glätten

---

### Exponentielles Glätten der Hotel-Daten:

Wir haben Daten über die Anzahl an Übernachtungen in österreichischen 4/5-Sterne-Hotels pro Jahr für den Zeitraum 1973-2003.

Die Hotel-Daten in Abb. 15 weisen einen deutlichen Trend auf. Die Jahre 1973-1995 wurden zur Prognose der Jahre 1996-2005 verwendet. Man sieht deutlich, dass die Prognose eines konstanten Niveaus für Daten mit einem langfristigen Trend unpassend ist.

## 7.2. Exponentielles Glätten

---

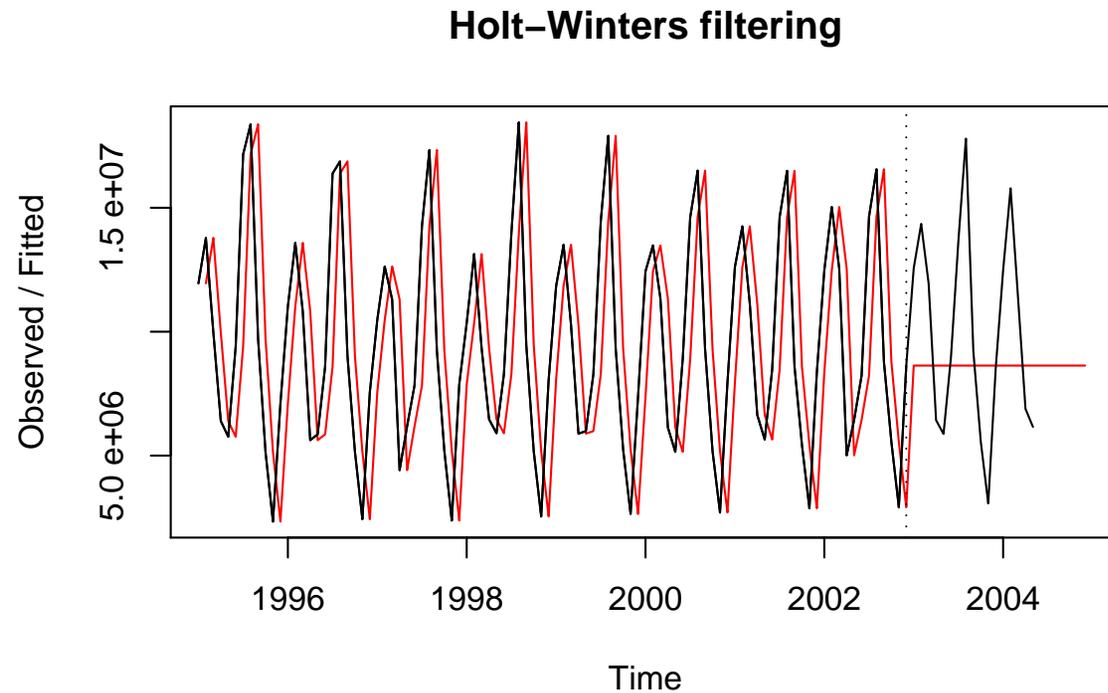


Abbildung 16: Exponentielles Glätten der Nächtigungs-Daten.

## 7.2. Exponentielles Glätten

---

### **Exponentielles Glätten der Nächtigungs-Daten:**

Die ZR enthält Tourismusdaten über die Anzahl an Übernachtungen in allen österreichischen Unterbringungsformen pro Monat (Zeitraum Jän. 1995 - Mai 2004).

Die Daten in Abb. 16 weisen eine deutliche Saisonalität auf. Die Jahre 1995-2002 wurden zur Prognose der Jahre 2003-2004 verwendet. Man sieht deutlich, dass die Prognose eines konstanten Niveaus für Daten mit einer Saisonalität unpassend ist.

## 7.2. Exponentielles Glätten

---

### Zur Anwendung des einfachen exponentiellen Glättens:

- Der Glättungsparameter  $\alpha$  liegt zwischen 0 und 1.
- Die Glättung ist umso stärker, je näher der Glättungsparameter an 0 liegt. In diesem Fall werden auch länger zurückliegende Beobachtungen noch stärker berücksichtigt.
- Die letzte Beobachtung erhält ein umso größeres Gewicht, je näher der Glättungsparameter an 1 liegt. In diesem Fall spielen Beobachtungen, die schon länger zurück liegen eine geringe Rolle.

## 7.2. Exponentielles Glätten

---

- Der Glättungsparameter  $\alpha$  soll so gewählt werden, dass die Fehlerquadratsumme zwischen Einschnitt-Vorhersage und wahren Wert der ZR möglichst klein wird:

$$SSE(\alpha) = \sum_{t=m}^{T-1} (x_{t+1} - \hat{x}_t(1))^2 \rightarrow \min$$

( $m$  wird hier so groß gewählt, dass der Effekt des Startwertes vernachlässigbar ist.)

Oft wird aus  $\alpha = 0.1, 0.2, \dots, 0.9$  jenes  $\alpha$  gewählt, das die kleinste  $SSE(\alpha)$  aufweist.

## 7.2. Exponentielles Glätten

---

- Für die Prognose ist wichtig, dass diese nur bei ZR ohne langfristigem Trend und ohne saisonale Schwankungen mit Hilfe des exponentiellen Glättens durchgeführt werden. Der lineare Trend bleibt zwar bei der geglätteten ZR bis zum Zeitpunkt  $T$  erhalten. Prognostiziert wird bei diesem Verfahren aber nur ein konstantes Niveau. (Vgl. Abb. 15)

## 8. Einfache Filterverfahren in R

---

Um in R die Methoden für das Glätten von ZR ohne Trend und Saisonalität durchzuführen, werden die Funktionen `filter` und `HoltWinters` verwendet.

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

# Übungen 2

---

Lösen Sie die folgenden Beispiele mit Hilfe von R:

1. Folgen Sie der Anleitung in "ExpGlaettinR.pdf" und rechnen Sie das Beispiel "Einfacher gleitender Durchschnitt der Nil-Daten" nach.
2. Laden Sie die Quartalsdaten "approve.rda" in das Arbeitsverzeichnis. Glätten Sie die ZR mit dem einfachen gleitenden Durchschnitt und wählen Sie einmal ein Fenster mit Breite 5 und dann ein Fenster mit Breite 9.
3. Folgen Sie der Anleitung in "ExpGlaettinR.pdf" und rechnen Sie das Beispiel "Exponentielles Glätten der Nil-Daten" nach.

## Übungen 2

---

4. Laden Sie die Daten "approve.rda" in das Arbeitsverzeichnis. Gehen Sie wie in Bsp. 3. vor. Die Daten der Jahre 1953-1971 sollen verwendet werden, um Werte für 1972 vorherzusagen. (Wir haben hier Quartalsdaten, daher müssen 4 Werte für 1972 vorhergesagt werden.) Beim exponentiellen Glätten soll  $\alpha = 0.2$  gewählt werden. Vergleichen Sie mit einer alternativen Wahl  $\alpha = 0.4$ . Wie lautet Ihre Prognose für das 4. Quartal 1972?
5. Lassen Sie für die Daten aus Bsp. 4. den Glättungsparameter von R optimieren. Wie lautet die Prognose jetzt? Wie hat sich die Fehlerquadratsumme verändert?

## 9. Zeitreihen mit Trend

---

In diesem Abschnitt beschäftigen wir uns mit ZR, die einen langfristigen linearen Trend, aber keine saisonalen Schwankungen, enthalten. Das zugrunde liegende Modell ist daher ein Modell das die ZR in eine Niveau-, eine lineare Trendkomponente und eine zufällige Restkomponente zerlegt.

## 9.1. Lineare und nichtlineare Trends

---

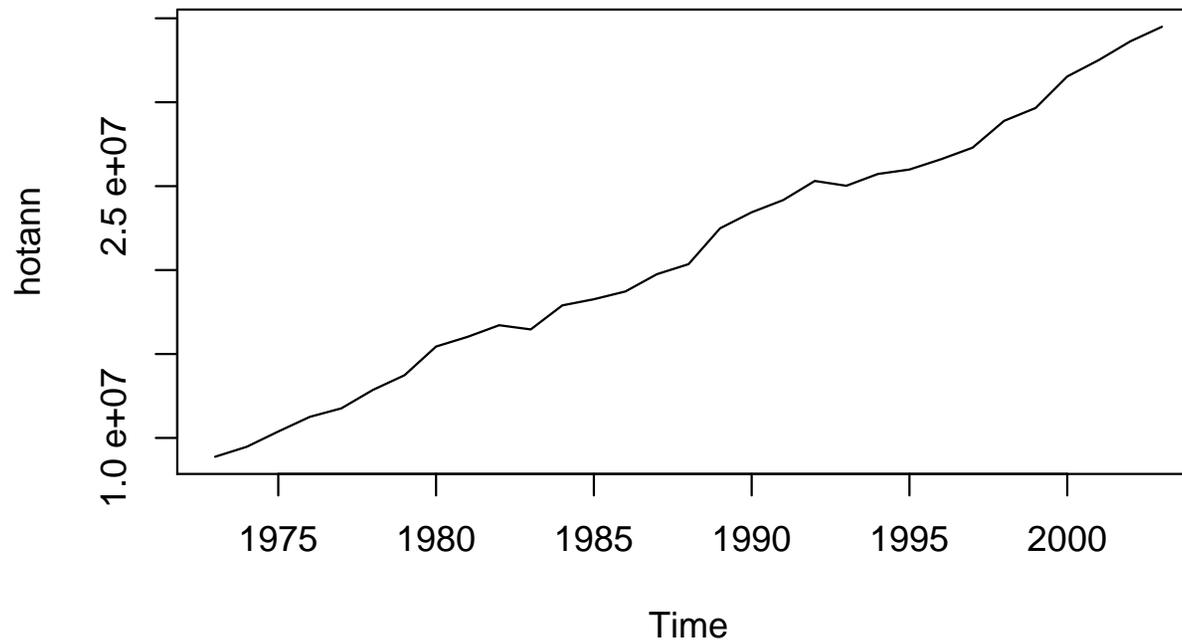


Abbildung 17: Zeitreihenplot Hotel-Daten.

## 9.1. Lineare und nichtlineare Trends

---

Die Hotel-Daten in Abb. 17 zeigen einen deutlichen linearen Trend. Man kann gut eine Gerade als Näherung an die Daten legen.

In diesem Abschnitt wird das Holt Verfahren hergeleitet. Es ist ein Verfahren, das Prognosen unter der Annahme eines linearen Trends trifft.

## 9.2. Transformation der Daten

---

Bei Daten, die einen Trend aufweisen, der aber nicht linear ist, kann eine Transformation dieser Daten sinnvoll sein.

Wir geben 2 Beispiele:

Die Daten in Abb. 18 zeigen einen nahezu quadratischen Trend. In Abb. 19 sehen wir die ZR, die sich als Wurzel aus den ursprünglichen Daten ergibt.

Die Daten in Abb. 20 zeigen einen nahezu exponentiellen Trend. In Abb. 21 wurden die Daten logarithmiert.

## 9.2. Transformation der Daten

---

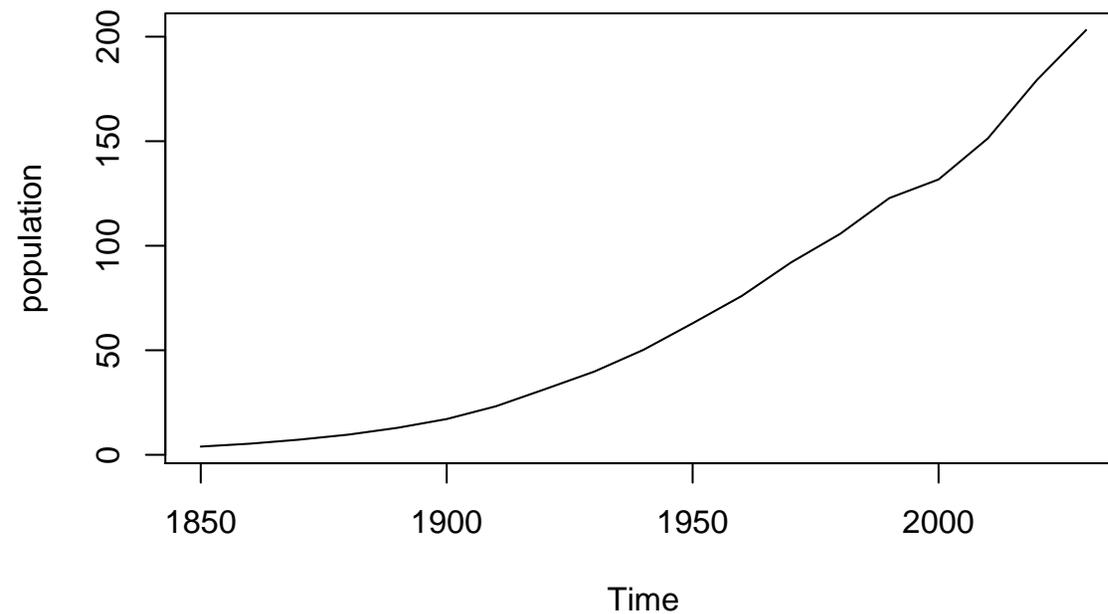


Abbildung 18: US population.

## 9.2. Transformation der Daten

---

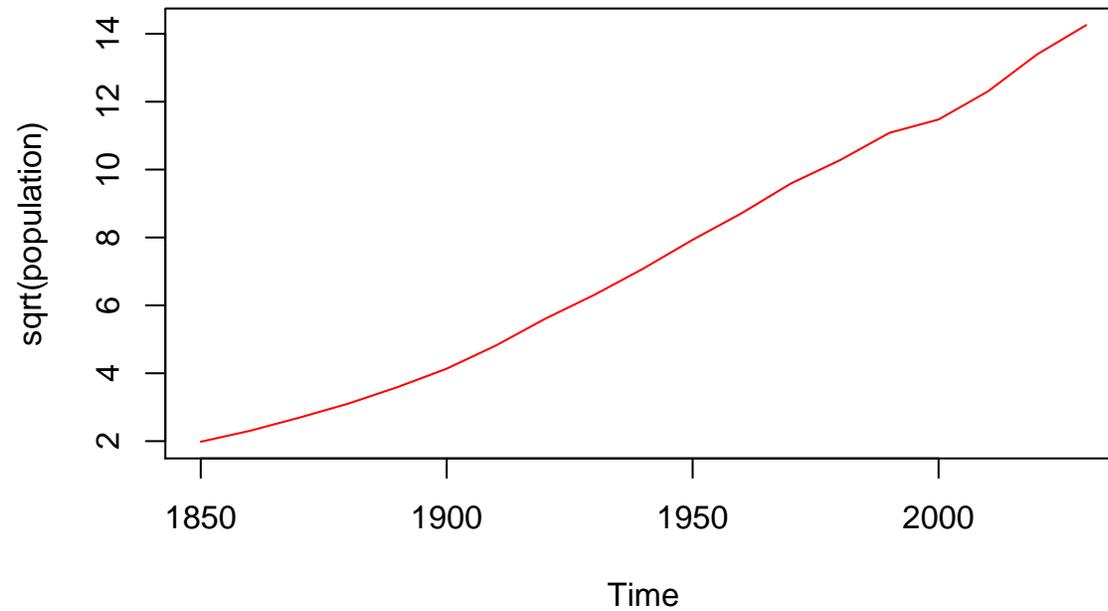


Abbildung 19: Transformation der US population.

## 9.2. Transformation der Daten

---

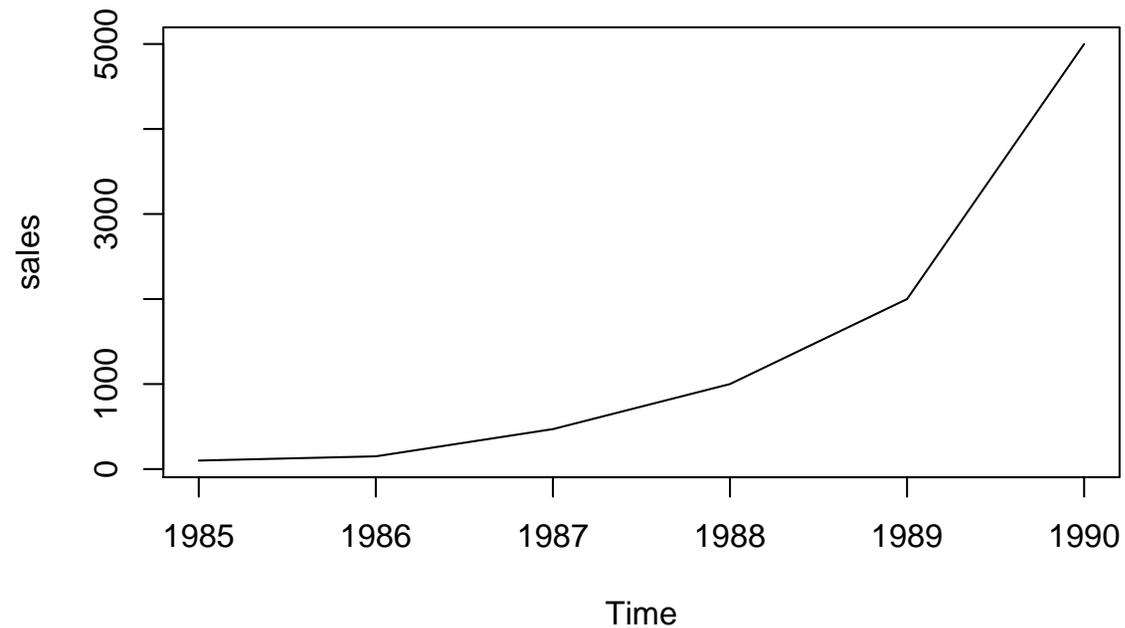


Abbildung 20: Verkaufszahlen einer Firma.

## 9.2. Transformation der Daten

---

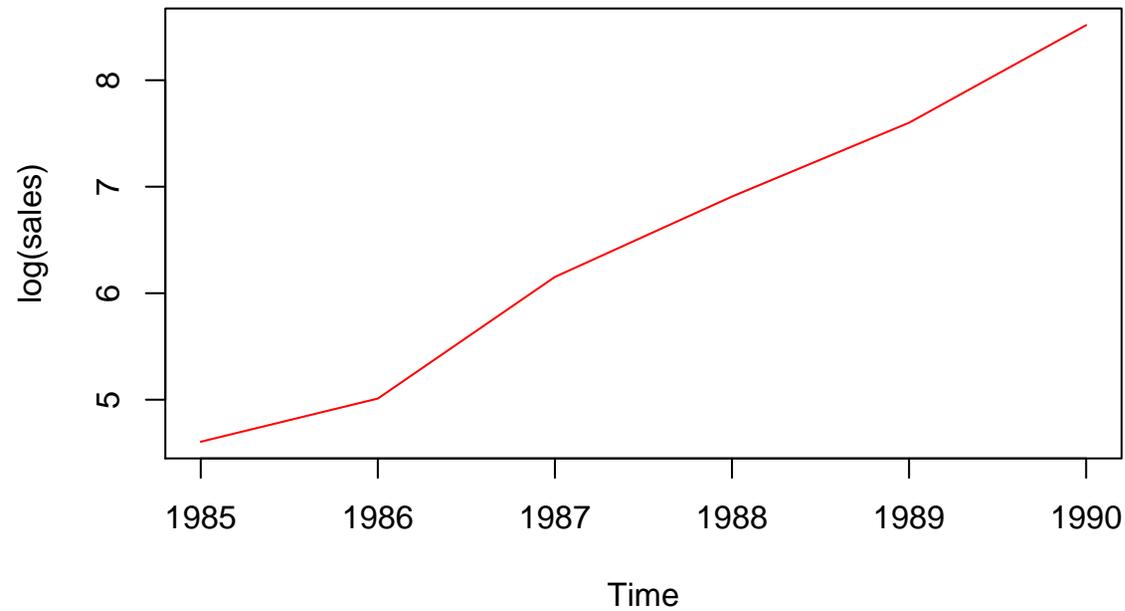


Abbildung 21: Transformation der Verkaufszahlen.

## 9.3. Holt Verfahren

---

Das Verfahren des exponentiellen Glättens wird nun so erweitert, dass auch ZR, die einen langfristigen, linearen Trend enthalten, prognostiziert werden können.

$$\hat{x}_T(h) = \hat{a}_{T+1} + \hat{b}_{T+1}h, \quad h = 1, 2, \dots$$

Die Daten setzen sich hier aus einer Niveauekomponente  $\hat{a}_{T+1}$  und einer linearen Trendkomponente  $\hat{b}_{T+1}h$  zusammen.

## 9.3. Holt Verfahren

---

Rekursion für  $t = 3, \dots, T$ :

$$\hat{a}_t = \alpha x_{t-1} + (1 - \alpha)(\hat{a}_{t-1} + \hat{b}_{t-1})$$

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

Die Einschnitt-Vorhersagen zum Zeitpunkt  $t - 1$  ergeben sich daher als Summe der Schätzwerte von Trend und Niveau:

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

## 9.3. Holt Verfahren

---

Startwerte:  $\hat{a}_2 = x_1, \hat{b}_2 = x_2 - x_1$

Rekursion:  $\hat{a}_3 = \alpha x_2 + (1 - \alpha)(\hat{a}_2 + \hat{b}_2) = x_2$   
 $\hat{b}_3 = \beta(\hat{a}_3 - \hat{a}_2) + (1 - \beta)\hat{b}_2 = x_2 - x_1$

⋮

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

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

Prognose f. T+1:  $\hat{a}_{T+1} = \alpha x_T + (1 - \alpha)(\hat{a}_T + \hat{b}_T)$   
 $\hat{b}_{T+1} = \beta(\hat{a}_{T+1} - \hat{a}_T) + (1 - \beta)\hat{b}_T$

Prognose f. T+h:  $\hat{a}_{T+h} = \hat{a}_{T+1}, \hat{b}_{T+h} = \hat{b}_{T+1} \cdot h$

## 9.3. Holt Verfahren

---

### **Holt Verfahren für die Hotel-Daten:**

Die Hotel-Daten weisen einen Trend auf, der gut durch eine lineare Trendfunktion modelliert werden kann. In Abb. 22 werden die Jahre 1973-2000 für die Prognose von 2001-2005 verwendet. Wir sehen, dass die lineare Trendprognose die wahren Daten der Jahre 2001-2003 gut approximiert.

## 9.3. Holt Verfahren

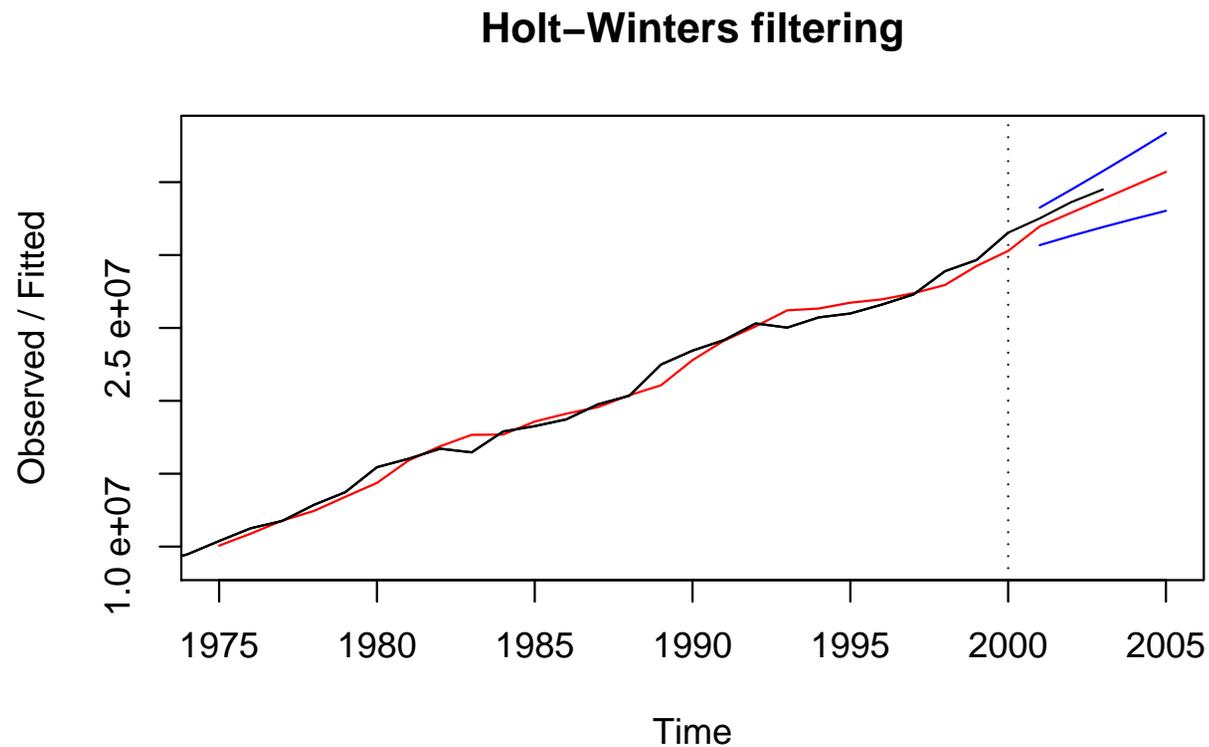


Abbildung 22: Holt-Verf. auf Hotel-Daten und Prognose für 2001-2005.

## 9.3. Holt Verfahren

---

### Zur Anwendung des Holt Verfahrens:

- Approximiert eine ZR durch eine Niveau- und eine lineare Trendkomponente.
- Da die Prognosefunktion eine lineare Funktion in der Zeit ist, können nur ZR, die einen linearen Trend aufweisen, prognostiziert werden. Weder nichtlineare Trends noch saisonale Schwankungen können vorhergesagt werden.
- Die Glättungsparameter  $\alpha$  und  $\beta$  liegen zwischen 0 und 1.

## 9.3. Holt Verfahren

---

- Für  $\alpha$  nahe bei 0 gehen auch länger zurückliegende Beobachtungen noch stärker in die Niveauschätzung ein.
- Die letzte Beobachtung erhält ein umso größeres Gewicht bei der Niveauschätzung, je näher der Glättungsparameter  $\alpha$  an 1 liegt.
- Für  $\beta$  nahe bei 0 geht die länger zurückliegende Trendentwicklung stärker in die aktuelle Trendschätzung ein.
- Die letzte Trendentwicklung erhält ein umso größeres Gewicht bei der Trendschätzung, je näher der Glättungsparameter  $\beta$  an 1 liegt.

## 9.3. Holt Verfahren

---

- Die Glättungsparameter können auch wieder durch Minimierung der Fehlerquadratsumme zwischen Einschritt-Vorhersage und wahren Wert der ZR geschätzt werden.

## 10. Holt Verfahren in R

---

Um in R das Holt-Verfahren durchzuführen, wird die Funktion `HoltWinters` verwendet.

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

# Übungen 3

---

Lösen Sie die folgenden Beispiele mit Hilfe von R:

1. Folgen Sie der Anleitung in "HoltinR.pdf" und rechnen Sie das dortige Beispiel nach.
2. Laden Sie die Daten "austres" in das Arbeitsverzeichnis. (Befehl: `data('austres')`), ist ein Datensatz, der im Package "stats" in R enthalten ist. Nähere Informationen können über `help('austres')` abgefragt werden.) Gehen Sie wie in Bsp. 1. vor. Die Daten der Jahre 1985 - 1989 sollen verwendet werden um Werte fuer 1990 -1993 vorherzusagen.

# Übungen 3

---

(Hinweis:

```
past <- window(austres, start=c(1985,1), end = c(1989,4)),  
future <- window(austres, start = c(1990,1)) )
```

Wählen Sie die Glättungsparameter  $\alpha = 0.4$  und  $\beta = 0.2$ . Wie lautet Ihre Prognose für das 2. Quartal 1993?

3. Lassen Sie für die Daten aus Bsp. 3. die Glättungsparameter von R optimieren. Welchen Wert bekommen  $\alpha$  und  $\beta$ ? Wie lautet die Prognose für das 2. Quartal 1993 jetzt? Wie hat sich die Fehlerquadratsumme für 1985-1989 verändert? Wie hat sich die mittlere Fehlerquadratsumme für 1990-2.Quartal 1993 verändert?

# 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

---

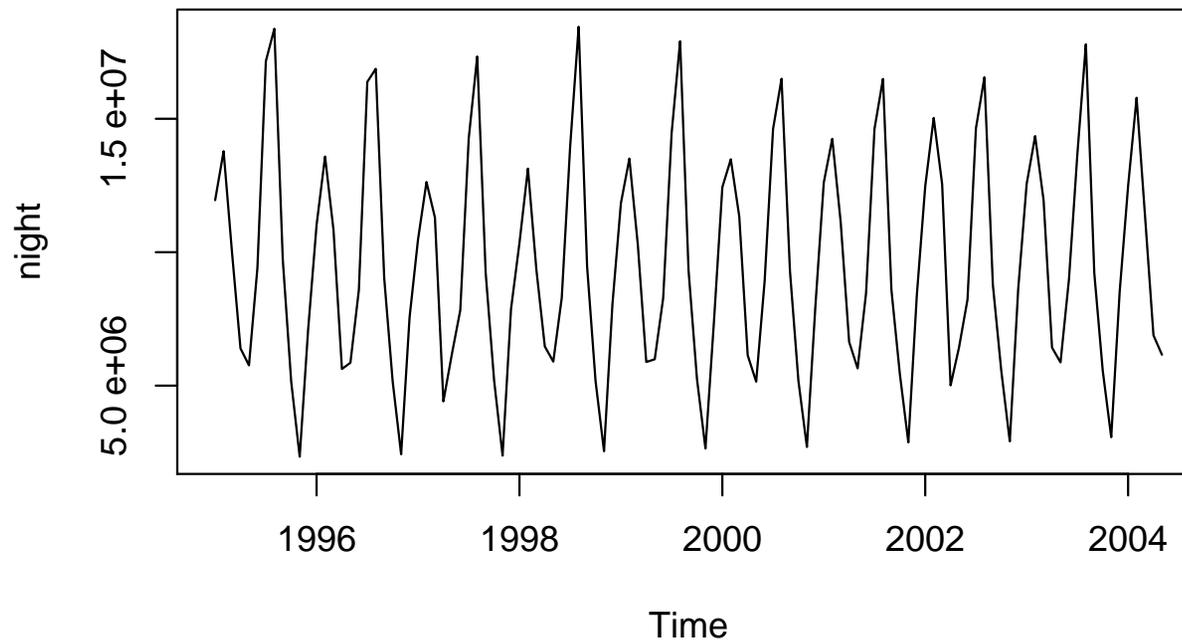


Abbildung 23: Zeitreihenplot der Nächtigungs-Daten.

# 11. Zeitreihen mit Trend und Saisonalität

---

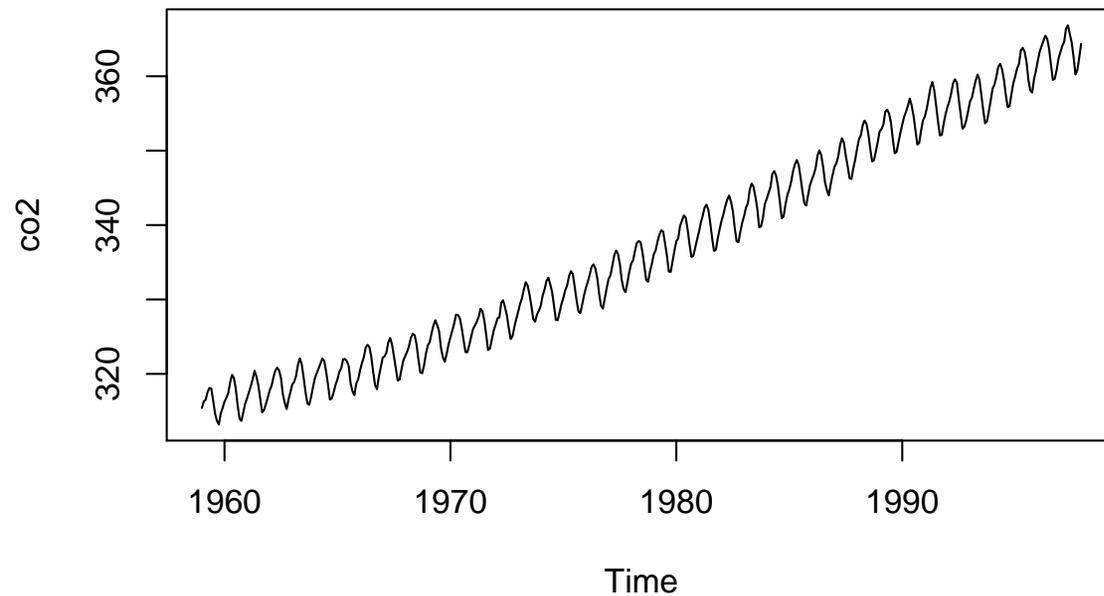


Abbildung 24: Zeitreihenplot der CO<sub>2</sub>-Daten.

# 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

---

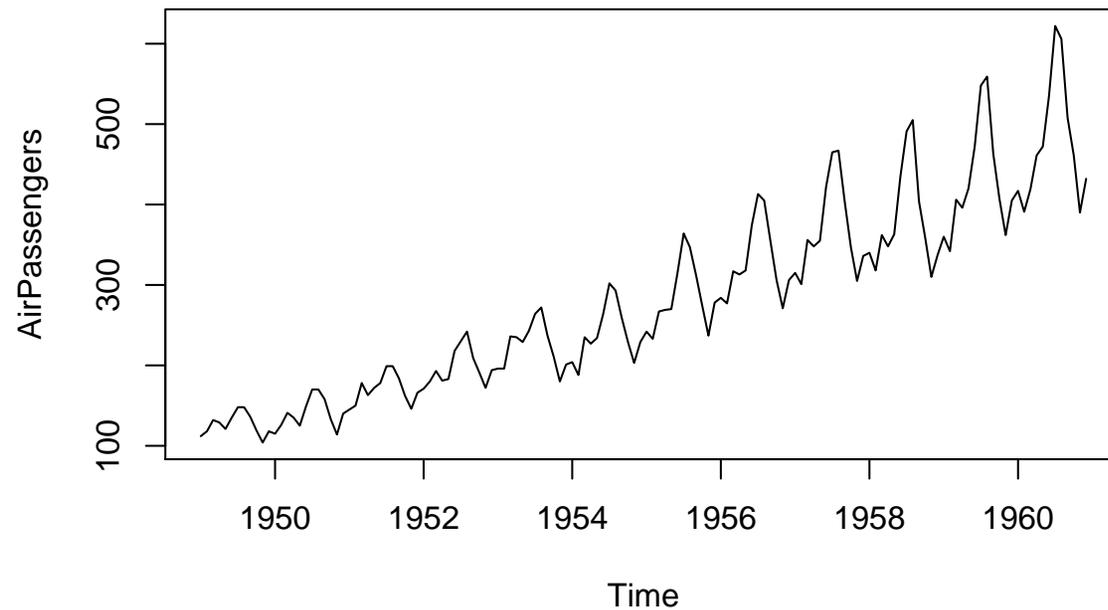


Abbildung 25: Zeitreihenplot der Flugpassagier-Daten.

# 11. Zeitreihen mit Trend und Saisonalität

---

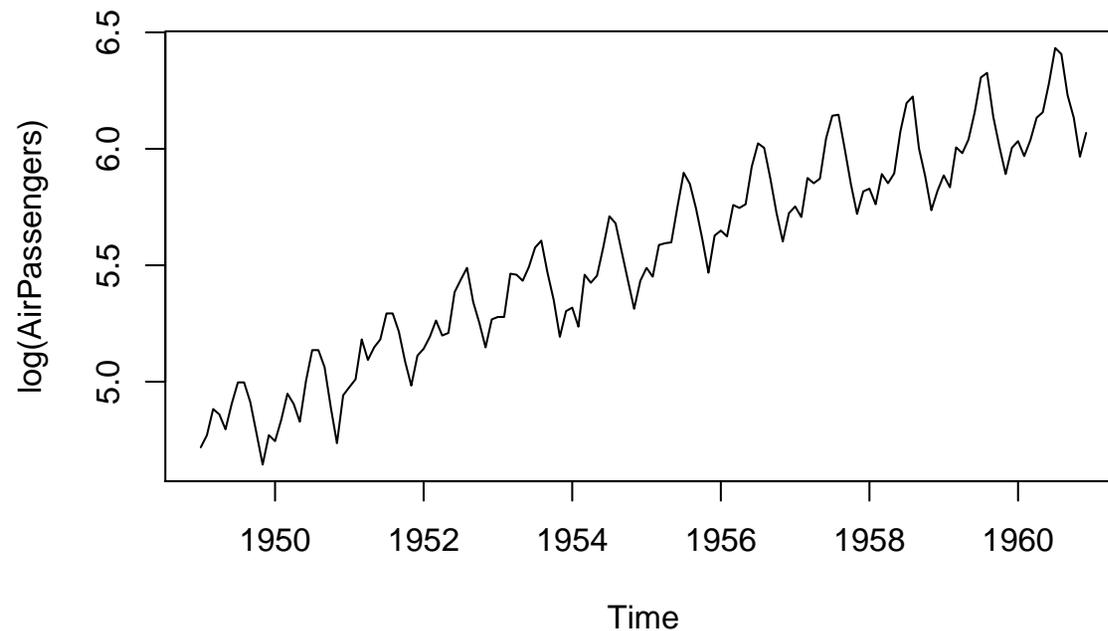


Abbildung 26: Logarithmierten Flugpassagier-Daten.

## 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  $j = 1, \dots, J$  und der zweite Index steht für die Saison  $k = 1, \dots, p$ .  $p$  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  $j = 1, \dots, J$ :

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

Jeweils für ein ganzes Jahr wird ein konstanter Wert angenommen. Schätzung der Saisonkomponente für jede Saison  $k = 1, \dots, p$  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

---

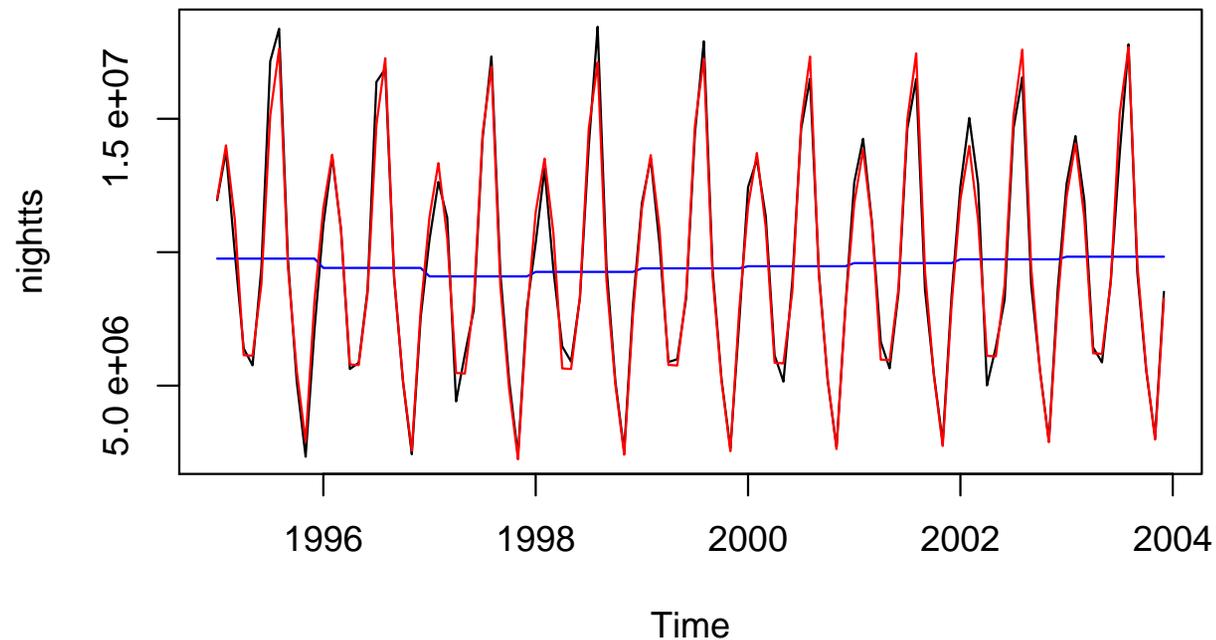


Abbildung 27: Methode d. kl. Trends für Nächtigungs-Daten.

## 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 klassischen 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  $x_t$ ,  $t = 1, \dots, T$  wird gemäß dem klassischen Komponentenmodell in den Trend  $\hat{m}_t$ , die Saison  $\hat{s}_t$  und die Restkomponente  $\varepsilon_t$  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  $t = q + 1, \dots, T - q$  berechnet:  
für ungerade Periode  $p = 2q + 1$ :

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

für gerade Periode  $p = 2q$ :

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

## 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

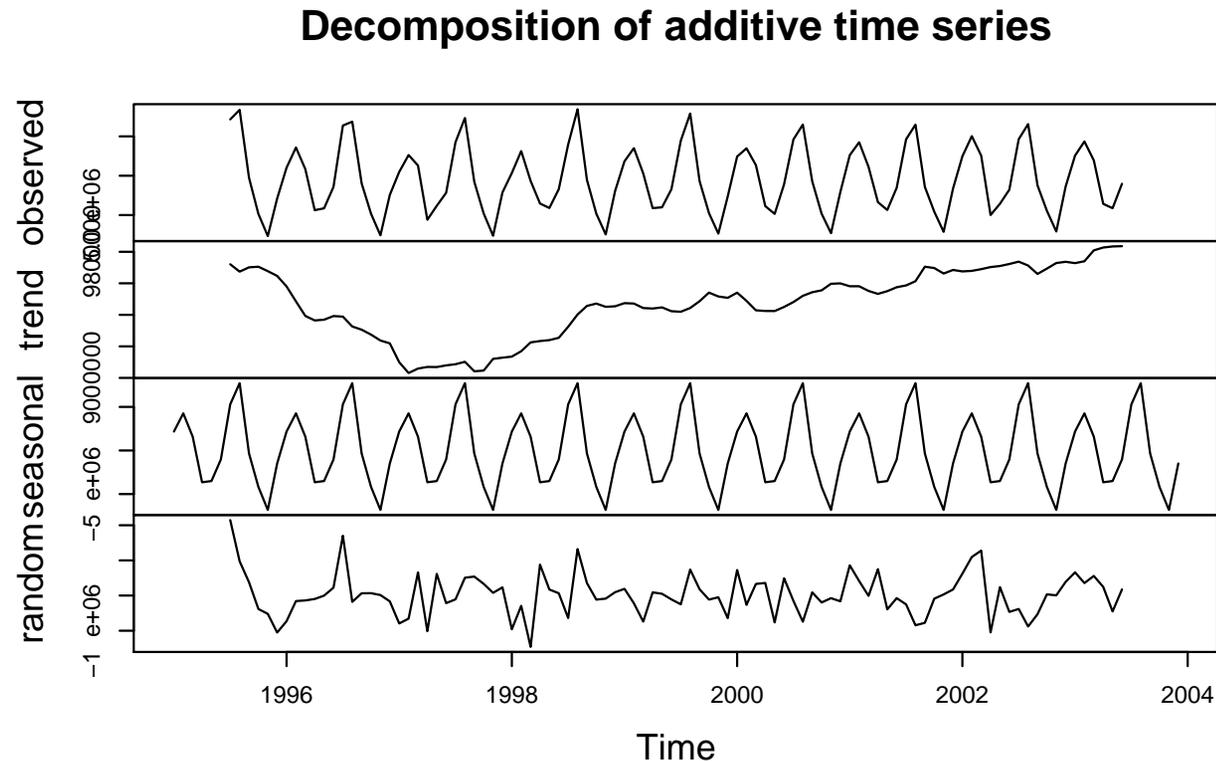


Abbildung 28: Saison-Trend Zerl. Nächtigungs-Daten mit MA.

## 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  $\varepsilon_t$ .

## 11.3. Holt-Winters Verfahren

---

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

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 ( $h = 1, 2, \dots, p$ ):

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

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

## 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

---

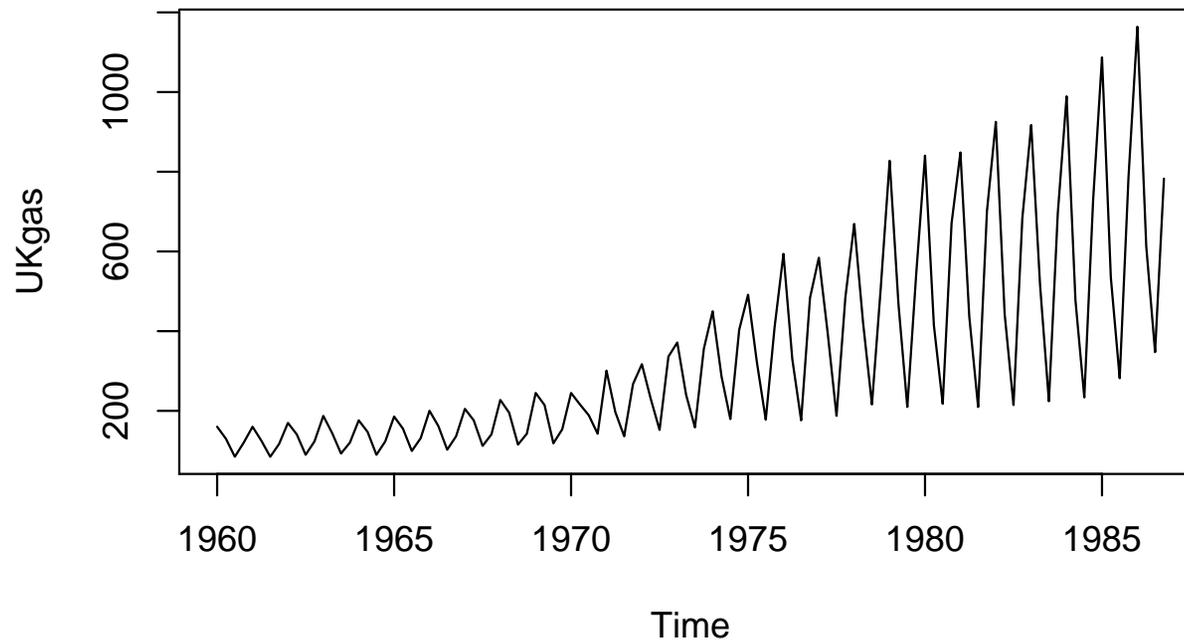


Abbildung 29: Die UKgas-Daten.

## 11.3. Holt-Winters Verfahren

---

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

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

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

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

Die Einschnitt-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  $t = p + 1, \dots, T$ :

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

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

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

Die Einschnitt-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  $t = 1, \dots, p$ . 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:

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

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

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

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

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

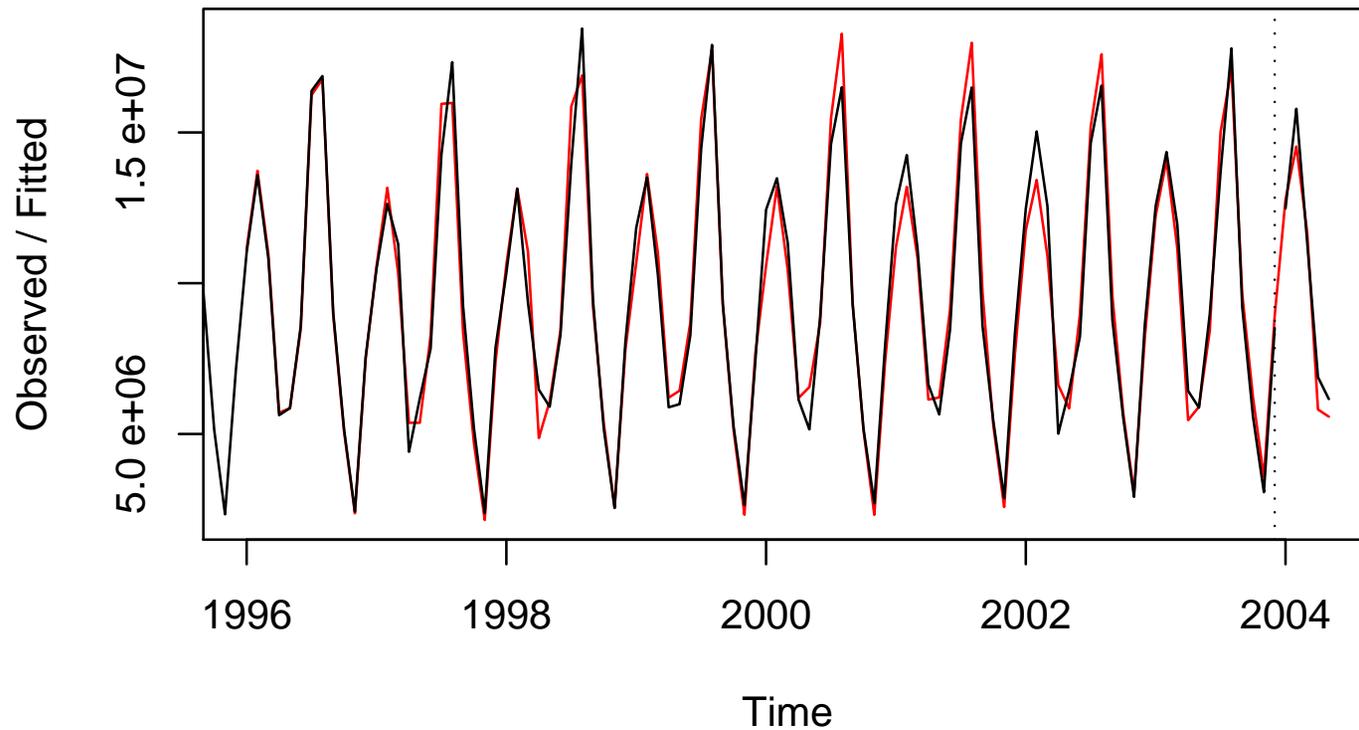
$$\hat{x}_T(h) = \hat{a}_{T+1} + \hat{b}_{T+1} \cdot h + \hat{s}_{T+h-2p}, \quad h = p + 1, \dots, 2p$$

*usw.*

# 11.3. Holt-Winters Verfahren

---

Holt-Winters filtering



## 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  $\alpha$ ,  $\beta$  und  $\gamma$  liegen zwischen 0 und 1.  $\alpha$  dient zur Glättung des Niveaus,  $\beta$  zur Glättung des linearen Trends und  $\gamma$  zur Glättung der Saisonalität.
- Die Glättungsparameter können auch wieder durch Minimierung der Fehlerquadratsumme zwischen Einschnitt-Vorhersage und wahren 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 Funktionen `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 `st1` 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?