

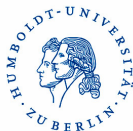
# Stochastik-Praktikum

## Zeitreihenanalyse

Thorsten Dickhaus

Humboldt-Universität zu Berlin

14.10.2010



# Übersicht

- 1 Zeitreihen in R
- 2 Trend und lineare Filter
- 3 Saisonale Analyse mit LOESS
- 4 Spektralanalyse
- 5 ARIMA-Modelle

# Übersicht

- 1 **Zeitreihen in R**
- 2 Trend und lineare Filter
- 3 Saisonale Analyse mit LOESS
- 4 Spektralanalyse
- 5 ARIMA-Modelle

# Einleitung

## Definition Zeitreihe:

Zeitdiskreter stochastischer Prozess  $X_t$ ,  $t \in \{1, \dots, n\}$ .

## Aufgaben der Zeitreihenanalyse:

- Erkennung von zeitlichen Entwicklungen (Trends),
- Erkennung von seriellen oder saisonalen Abhängigkeiten,
- Vorhersage zukünftiger Werte der Zeitreihe (Prognose).

Zeitreihen werden in R als spezielle Objekte gehandhabt.

Zeitabstände zwischen den beobachteten Realisierungen  
äquidistant: Funktion  $t_s$

# R-Code: Zeitreihe Bierproduktion

```
> beer<-read.csv("beer.csv",header=T,dec=".",sep=";")
> beer<-ts(beer[,1],start=1956,freq=12)
> beer
```

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1956	93.2	96.0	95.2	77.1	70.9	64.8	70.1	77.3	79.5	100.6	100.7	107.1
1957	95.9	82.8	83.3	80.0	80.4	67.5	75.7	71.1	89.3	101.1	105.2	114.1
1958	96.3	84.4	91.2	81.9	80.5	70.4	74.8	75.9	86.3	98.7	100.9	113.8
1959	89.8	84.4	87.2	85.6	72.0	69.2	77.5	78.1	94.3	97.7	100.2	116.4
1960	97.1	93.0	96.0	80.5	76.1	69.9	73.6	92.6	94.2	93.5	108.5	109.4
1961	105.1	92.5	97.1	81.4	79.1	72.1	78.7	87.1	91.4	109.9	116.3	113.0
1962	100.0	84.8	94.3	87.1	90.3	72.4	84.9	92.7	92.2	114.9	112.5	118.3
1963	106.0	91.2	96.6	96.3	88.2	70.2	86.5	88.2	102.8	119.1	119.2	125.1
1964	106.1	102.1	105.2	101.0	84.3	87.5	92.7	94.4	113.0	113.9	122.9	132.7
1965	106.9	96.6	127.3	98.2	100.2	89.4	95.3	104.2	106.4	116.2	135.9	134.0
1966	104.6	107.1	123.5	98.8	98.6	90.6	89.1	105.2	114.0	122.1	138.0	142.2
1967	116.4	112.6	123.8	103.6	113.9	98.6	95.0	116.0	113.9	127.5	131.4	145.9
1968	131.5	131.0	130.5	118.9	114.3	85.7	104.6	105.1	117.3	142.5	140.0	159.8
1969	131.2	125.4	126.5	119.4	113.5	98.7	114.5	113.8	133.1	143.4	137.3	165.2
1970	126.9	124.0	135.7	130.0	109.4	117.8	120.3	121.0	132.3	142.9	147.4	175.9
1971	132.6	123.7	153.3	134.0	119.6	116.2	118.6	130.7	129.3	144.4	163.2	179.4
1972	128.1	138.4	152.7	120.0	140.5	116.2	121.4	127.8	143.6	157.6	166.2	182.3
1973	153.1	147.6	157.7	137.2	151.5	98.7	145.8	151.7	129.4	174.1	197.0	193.9
1974	164.1	142.8	157.9	159.2	.....	.....	.....	.....	.....	.....	.....	.....

# R-Code: Zeitreihe Bierproduktion

```
> tsp(beer)
[1] 1956.000 1995.583 12.000
```

```
> frequency(beer)
[1] 12
```

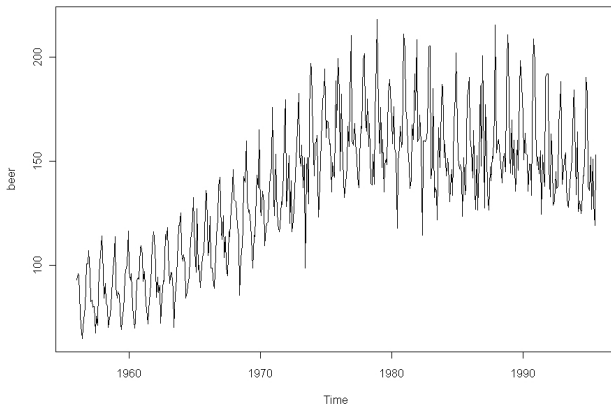
```
> start(beer)
[1] 1956 1
```

```
> end(beer)
[1] 1995 8
```

```
> deltat(beer)
[1] 0.08333333
```

# Grafische Darstellung von Zeitreihen mit R

```
> ts.plot(beer)
```



# R-Code: Bearbeiten von Zeitreihen

```
> z<-aggregate(beer,1,mean)
```

```
> z
```

```
Time Series:
```

```
Start = 1956
```

```
End = 1994
```

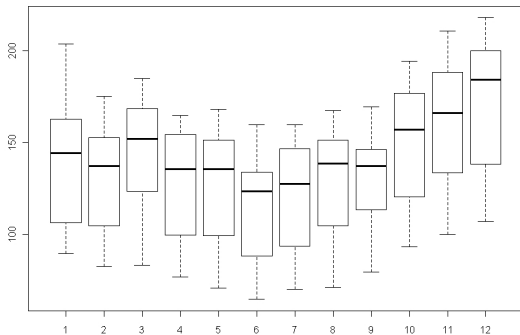
```
Frequency = 1
```

```
[1] 86.04167 87.20000 87.92500 87.70000 90.36667 93.64167 95.36667  
[8] 99.11667 104.65000 109.21667 111.15000 116.55000 123.43333 126.83333  
[15] 131.96667 137.08333 141.23333 153.14167 159.57500 161.68333 161.97500  
[22] 163.39167 166.52500 162.35000 160.92500 169.24167 165.05833 154.13333  
[29] 154.25833 154.75833 154.13333 156.38333 163.16667 160.20833 163.20833  
[36] 158.25833 150.41667 147.91667 148.00000
```



# Monatsweiser Vergleich

```
> boxplot(split(beer, cycle(beer)))
```



# Übersicht

- 1 Zeitreihen in R
- 2 Trend und lineare Filter**
- 3 Saisonale Analyse mit LOESS
- 4 Spektralanalyse
- 5 ARIMA-Modelle

# Trendschätzung (zeitliche Entwicklung des Durchschnittes)

Laufender gewichteter Mittelwert aller Beobachtungen:

$$T_t = \sum_{i=-\infty}^{\infty} \lambda_i X_{t+i} \text{ mit Gewichten } \lambda_i, \text{ wobei } \sum \lambda_i = 1.$$

**Linearer Filter:** Bezieht nur eine endliche Anzahl  $j$  zeitlich folgender und vorangehender Nachbarwerte mit ein:

$$T_t^{L,j} = \frac{1}{2j+1} \sum_{i=-j}^j X_{t+i}$$

Beim linearen Filter wird also das laufende arithmetische Mittel von  $(2j+1)$  Werten gebildet (**Gleitende Durchschnitte**).

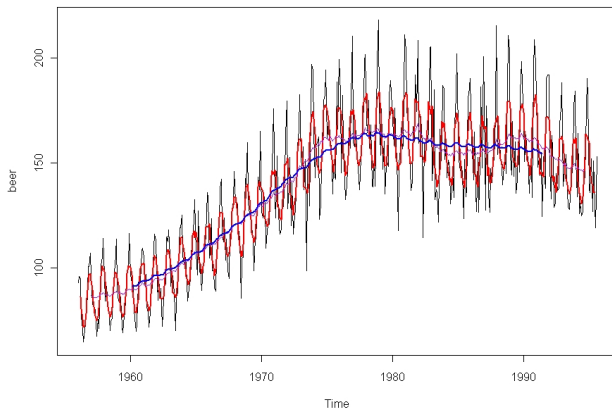
# R-Code: Lineare Filter

```
> ts.plot(beer)

> lf1 <- filter(beer, filter=rep(1/5,5))
> lf2 <- filter(beer, filter=rep(1/25,25))
> lf3 <- filter(beer, filter=rep(1/100,100))

> lines(lf1, col="red", lwd=2)
> lines(lf2, col="purple")
> lines(lf3, col="blue", lwd=2)
```

# Bierproduktion: Lineare Filter



# Übersicht

- 1 Zeitreihen in R
- 2 Trend und lineare Filter
- 3 Saisonale Analyse mit LOESS**
- 4 Spektralanalyse
- 5 ARIMA-Modelle

# Saisonschätzung

## Ziel:

Additive Aufspaltung einer Zeitreihe in Trend, eine saisonale (periodische) Komponente und Residuen (Rauschanteil).

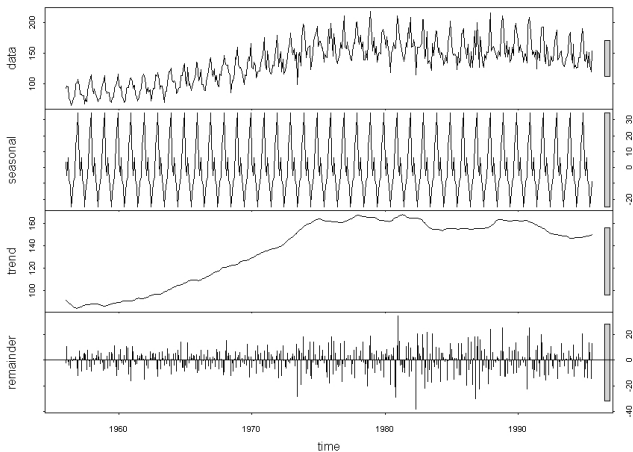
## Methode für die Saisonschätzung:

LOESS / LOWESS (lokale polynomiale Regression)

R-Implementierung der Zerlegung: Funktion `stl`

```
> plot(stl(beer, s.window="periodic"))
```

# Bier-Zeitreihe: STL-Zerlegung





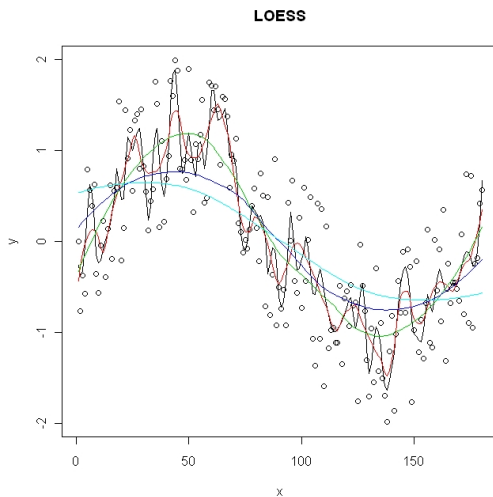
# R-Code: LOESS Illustration

```
> x<- 1:180
> y<- sin(pi*x/90) + runif(length(x),-1,1)
> plot(x,y, main="LOESS")

> yloess1<-loess(y ~ x, span=0.05, data.frame(x=x, y=y))
> yloess2<-loess(y ~ x, span=0.1, data.frame(x=x, y=y))
> yloess3<-loess(y ~ x, span=0.5, data.frame(x=x, y=y))
> yloess4<-loess(y ~ x, span=1, data.frame(x=x, y=y))
> yloess5<-loess(y ~ x, span=1.5, data.frame(x=x, y=y))

> ypredict5<- predict(yloess5, data.frame(x=x))
> lines(x,ypredict5, col=5)
```

# Illustration LOESS



# Übersicht

- 1 Zeitreihen in R
- 2 Trend und lineare Filter
- 3 Saisonale Analyse mit LOESS
- 4 Spektralanalyse**
- 5 ARIMA-Modelle

# Das Autokorrelogramm

Autokorrelogramm und Autokovarianzfunktion:

$$acf_t(k) = \text{Cov}(X_t, X_{t+k})$$

Partielle Autokorrelationen:

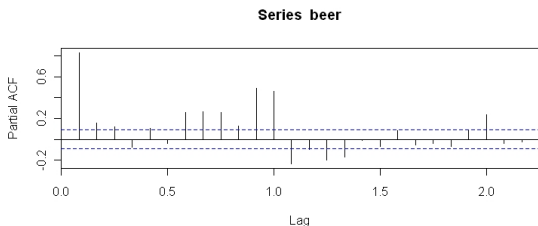
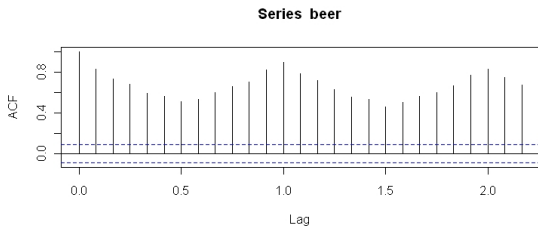
Korrelationen ohne den Einfluss der linearen Abhängigkeiten von  $X_{t+1}, \dots, X_{t+k-1}$ .

Beispiel: Partielle Autokorrelation von  $X_1$  und  $X_3$ :

$$\rho c_{1,3} = \frac{c_{1,3} - c_{1,2}c_{2,3}}{\sqrt{(1 - c_{1,2}^2)(1 - c_{2,3}^2)}},$$

wobei  $c_{i,j}$  die empirische (Auto-)Kovarianz bezeichnet.

# Autokorrelogramm der Bier-Zeitreihe



# Spektralmaß

Die Autokovarianzfunktion einer **stationären** Zeitreihe kann man mit dem **Spektrum**  $F$  darstellen als

$$\text{acf}_t(k) \equiv \gamma_t = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i\omega t} dF(\omega).$$

$F$  ist ein endliches Maß auf  $(-\pi, \pi)$ .

Existiert die **Spektraldichte**  $f$ , so kann man  $\gamma_t$  über die Fouriertransformierte darstellen:

$$\gamma_t = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i\omega t} f(\omega) d\omega = \int_{-1/2}^{1/2} \exp(2\pi i\omega_f t) f(2\pi\omega_f) d\omega_f.$$

$\omega_f$  bezeichnet die Kreisfrequenz / Periodenfrequenz.

# Periodogramm

Entwicklung der Spektraldichte  $f$  in eine Fourierreihe ergibt (Autokovarianzfunktion einer stationären Zeitreihe gerade):

$$f(\omega) = \sum_t \gamma_t e^{-i\omega t} = \gamma_0 \left( 1 + 2 \sum_{t \geq 1} \gamma_t \cos \omega t \right).$$

Periodogramm bei einer Frequenz  $\omega$ :

Quadrierte Korrelation der Zeitreihe und Sinus/Kosinus-Wellen:

$$I(\omega) = \frac{1}{n} \left| \sum_{t=1}^n e^{-i\omega t} X_t \right|^2 = \frac{1}{n} \left( \left( \sum_{t=1}^n X_t \sin \omega t \right)^2 + \left( \sum_{t=1}^n X_t \cos \omega t \right)^2 \right)$$

# Schätzung des Periodogramm

Bezeichnen wir mit  $\hat{\gamma}_t$  die empirische Autokovarianzfunktion:

$$\hat{\gamma}_t = \frac{1}{n} \sum_{j=1}^{n-t} (X_{t+j} - \bar{X}) (X_j - \bar{X}).$$

(Schätzer der Autokovarianzen aus den Beobachtungen)

Geschätztes Periodogramm: Fouriertransformierte

$$\hat{l}(\omega) = \sum_t \exp(-i\omega t) \hat{\gamma}_t.$$



# Schätzung der Spektraldichte

## Schätzer für die Spektraldichte:

Geglättete Version des Periodogramms durch einen  
Kernschätzer mit Bandweite  $h$ :

$$\hat{f}(\omega) = \frac{1}{n} \int \mathcal{K}\left(\frac{\lambda - \omega}{h}\right) \hat{I}(\lambda) d\lambda \approx \frac{2\pi}{nh} \sum_k \mathcal{K}\left(\frac{\omega_k - \omega}{h}\right) \hat{I}(\omega_k).$$

Kumulatives Periodogramm:  $\hat{f}(\omega_j) / \sum_j^n \hat{f}(\omega_j)$

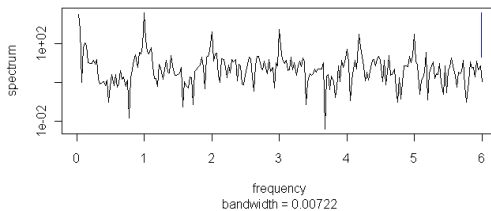
# R-Code: Spektralanalyse

```
> par(mfrow=c(2,1))
> spectrum(beer,main="beer time series")
> x<-ts(rnorm(500))
> spectrum(x,main="gaussian white noise")

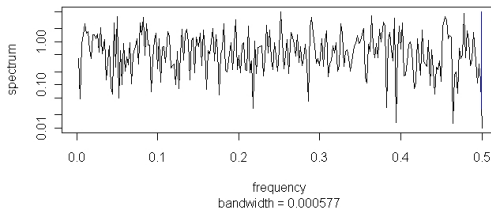
> par(mfrow=c(1,2))
> cpggram(beer,main="beer time series")
> cpggram(x,main="gaussian white noise")
```

# Spektralanalyse

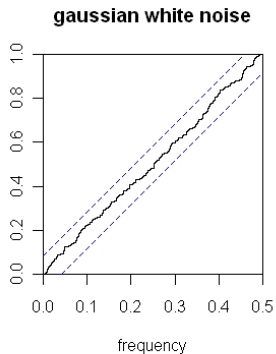
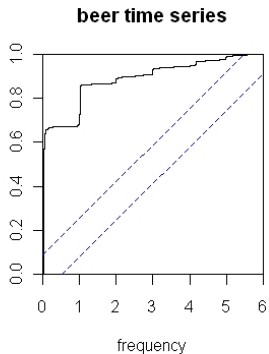
**beer time series**



**gaussian white noise**



# Kumulatives Periodogramm



# Übersicht

- 1 Zeitreihen in R
- 2 Trend und lineare Filter
- 3 Saisonale Analyse mit LOESS
- 4 Spektralanalyse
- 5 ARIMA-Modelle**

# ARIMA Zeitreihen

- **AR( $p$ )-Prozess (autoregressiv):** Zeitreihe  $X_t$ , die von der gewichteten Summe  $p$  vorangehender Werte abhängt, also

$$X_t = \sum_{j=1}^p \alpha_j X_{t-j} + \epsilon_t.$$

- **MA( $q$ )-Prozess (moving average):** Zeitreihe  $X_t$ , die von der gewichteten Summe  $q$  vorangehender Fehler abhängt, also

$$X_t = \sum_{i=0}^q \beta_i \epsilon_{t-i}.$$

Im Allgemeinen setzt man  $\beta_0 = 1$ .

# ARIMA Zeitreihen

- **AR( $p$ )-Prozess (autoregressiv):** Zeitreihe  $X_t$ , die von der gewichteten Summe  $p$  vorangehender Werte abhängt, also

$$X_t = \sum_{j=1}^p \alpha_j X_{t-j} + \epsilon_t.$$

Wenn die  $p$ -ten Einheitswurzeln des AR-Polynoms

$$1 - \sum_{j=1}^p \alpha_j z^j$$

außerhalb des Einheitskreises liegen, sind AR( $p$ ) Prozesse stationär.

# ARIMA Zeitreihen

- **MA( $q$ )-Prozess (moving average):** Zeitreihe  $X_t$ , die von der gewichteten Summe  $q$  vorangehender Fehler abhängt, also

$$X_t = \sum_{i=0}^q \beta_i \epsilon_{t-i}.$$

Im Allgemeinen setzt man  $\beta_0 = 1$ .

MA( $q$ )-Prozesse sind stationär.

$X_t$  hat die Varianz  $\sigma_\epsilon^2 \sum \beta_i^2$ , wobei  $\sigma_\epsilon^2$  die Varianz des weißen Rauschens  $\epsilon_t$  ist.



# ARIMA Zeitreihen

- Unter einem **ARMA( $p,q$ )-Prozess** versteht man eine Zeitreihe  $X_t$ , die von der gewichteten Summe vorangehender Werte und Fehler abhängt:

$$X_t = \sum_{j=1}^p \alpha_j X_{t-j} + \sum_{i=0}^q \beta_i \epsilon_{t-i} .$$

Man fordert, dass das AR-Polynom und das MA-Polynom  $1 + \sum_{i=1}^q \beta_i$  keine gemeinsamen komplexen Nullstellen haben.

- Eine Zeitreihe heißt **ARIMA( $p,d,q$ )-Prozess** (auto regressive integrated moving average), wenn die Zeitreihe, die durch Bildung der  $d$ -ten Differenzen entsteht, ein ARMA( $p,q$ )-Prozess ist.

# Box-Jenkins-Verfahren

- 1 Modellidentifikation (Ordnungen bestimmen),
- 2 Modellanpassung bzw. Parameterschätzung (Feinschätzung des Modells)
- 3 Modellüberprüfung bzw. Diagnose der Modellgüte

Nach Schritt 3 erfolgt ggf. eine Rückkehr zu Schritt 1.

Die Modellidentifikation führt über das Autokorrelogramm und das kumulative Autokorrelogramm.

## Spezialfälle:

- Einfacher Random Walk:  $(0,1,0)$
- Exponential Smoothing:  $(0,1,1)$

# ACF und PACF von ARMA( $p, q$ )-Prozessen

## AR-Prozesse:

- AR(1) hat Autokorrelationen  $\gamma_k = \alpha^{|k|}$
- AR( $p$ ): Alle Autokovarianzen ungleich Null, fallen aber für größere Lags ab.
- Yule-Walker-Gleichung:  $\gamma_k = \sum_1^p \alpha_i \gamma_{k-i}$ ,  $k > 0$ .  
⇒ Alle **partiellen Autokovarianzen** für  $k > p$  **gleich Null**.

## MA-Prozesse:

- **MA(1)**: Autokorrelation  $\gamma_1 = \beta_1 / (1 + \beta_1^2)$  zum Lag 1 und Null für größere Lags.
- MA( $q$ ): Alle **Autokovarianzen** für  $k > q$  **gleich Null**
- Partielle Autokorrelationen abfallend, aber ungleich Null.

# R-Code: Box-Jenkins, 1.Schritt

```
> ar(beer, max.order=100)
```

Call:

```
ar(x = beer, max.order = 100)
```

Coefficients:

1	2	3	4	5	6	7	8
0.2497	0.0797	-0.0658	-0.0349	0.0968	0.0160	-0.0439	0.0178
9	10	11	12	13	14	15	16
0.0107	-0.0973	0.1044	0.4177	0.0245	0.0106	0.0257	-0.0042
17	18	19	20	21	22	23	24
-0.0196	-0.0362	-0.0486	0.0477	-0.0025	-0.0200	0.1808	0.0822
25							
-0.0986							

Order selected 25 sigma^2 estimated as 5573

# R-Code: Box-Jenkins, 2.Schritt

```
> arima(beer, order=c(25,1,0))
```

Call:

```
arima(x = beer, order = c(25, 1, 0))
```

Coefficients:

	ar1	ar2	ar3	ar4	ar5	ar6	ar7	ar8
	-0.8138	-0.7461	-0.8000	-0.8405	-0.7393	-0.7010	-0.7282	-0.6987
s.e.	0.0467	0.0602	0.0698	0.0791	0.0883	0.0947	0.0998	0.1054
	ar9	ar10	ar11	ar12	ar13	ar14	ar15	ar16
	-0.6753	-0.7666	-0.6638	-0.2427	-0.1851	-0.132	-0.0912	-0.0893
s.e.	0.1103	0.1150	0.1206	0.1245	0.1245	0.124	0.1201	0.1147
	ar17	ar18	ar19	ar20	ar21	ar22	ar23	ar24
	-0.1045	-0.1485	-0.2082	-0.1669	-0.1664	-0.1937	-0.0017	0.1181
s.e.	0.1102	0.1052	0.0997	0.0946	0.0883	0.0793	0.0705	0.0607
	ar25							
	0.0583							
s.e.	0.0472							

sigma^2 estimated as 4931: log likelihood = -2703.55, aic = 5459.09

## Modellüberprüfung (3. Schritt)

Korrekt spezifiziertes Modell: **Residuen unkorreliert !**

Der Box–Pierce und Ljung–Box Test (**Portmanteau-Tests**) sind Signifikanztests auf die Hypothese der Unkorreliertheit der Residuen. Die Teststatistik ist  $n \sum_t \hat{\gamma}_t^2$  und unter der Nullhypothese approximativ  $\chi^2$ -verteilt.

Es werden drei Grafiken erzeugt:  
Standardisierte Residuen, deren Autokorrelationen und Ljung–Box  $p$ -Werte für die niedrigsten Lags.

# R-Code: Box-Jenkins 3.Schritt

```
> tsdiag(arima(beer, order=c(25,1,0)))
> Box.test(arima(beer, order=c(25,1,0))$residuals)

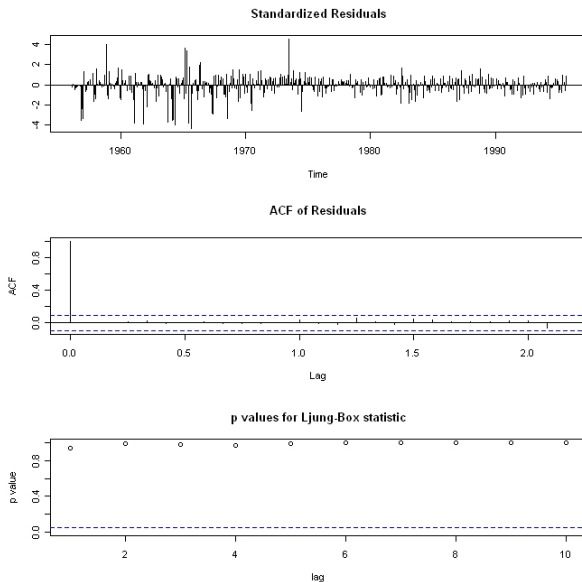
      Box-Pierce test

data:  arima(beer, order = c(25, 1, 0))$residuals
X-squared = 0.0061, df = 1, p-value = 0.938

> Box.test(arima(beer, order=c(25,1,0))$residuals, type="Ljung")

      Box-Ljung test

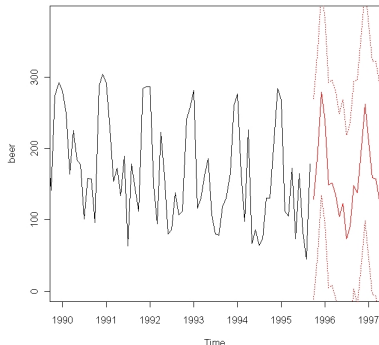
data:  arima(beer, order = c(25, 1, 0))$residuals
X-squared = 0.0061, df = 1, p-value = 0.9378
```








# R-Code: Vorhersage im ARIMA-Modell




```
> fore<-predict(arima(beer, order=c(25,1,0)), n.ahead=24)
> ts.plot(beer, xlim=c(1990,1997))
> lines(fore$pred, col="red")
> lines(fore$pred+2*fore$se, col="red", lty=3)
> lines(fore$pred-2*fore$se, col="red", lty=3)
```



# Literatur

-  Rainer Schlittgen and Bernd H. J. Streitberg  
Zeitreihenanalyse.  
*Oldenbourg Verlag*, 9, 2001.
-  Paul S.P. Cowpertwait and Andrew Metcalfe  
Introductory Time Series with R.  
*Berlin: Springer*, 1, 2009.
-  Robert H. Shumway and David S. Stoffer  
Time Series Analysis and Its Applications: With R  
Examples.  
*Berlin: Springer*, 2, 2006.

# Literatur

-  William S. Cleveland and Clive Loader  
Smoothing by local regression: principles and methods.  
*Statistical Theory and Computational Aspects of Smoothing*, 10-49, 1996.
-  Prabhakar S. Naidu  
Modern Spectrum Analysis of Time Series.  
*CRC-Press*, 1, 1995.
-  Spyros Makridakis and Steve Wheelwright and Rob J. Hyndman  
Forecasting: Methods and Applications.  
*Wiley*, 3, 1998.

# Literatur



Maurice B. Priestley  
Spectral Analysis and Time Series.  
*London: Academic Press, 1981.*



Peter J. Brockwell and Richard A. Davis  
Introduction to Time Series and Forecasting.  
*New York: Springer, 2, 2002.*



Peter J. Diggle  
Time series: A biostatistical introduction.  
*Oxford University Press, 1994.*