Chapter 10: Introduction to Time Series Modeling and Forecasting A

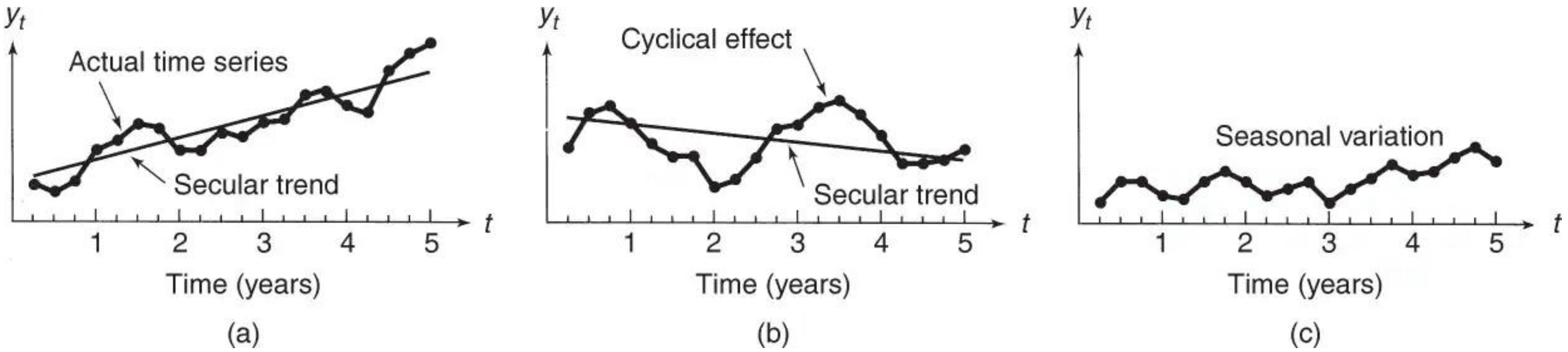
Dr. Daniel B. Rowe
Professor of Computational Statistics
Department of Mathematical and Statistical Sciences
Marquette University

Intro to Time Series Modeling & Forecasting

Time Series Components

If repeated observations on a variable produce a time series, the variable is called a **time series variable**. We use y_t to denote the value of the variable at time t.

Researchers often describe a time series y_t by four components
(1) secular trend, (2) cyclical effect, (3) seasonal variation, and (4) residual effect.



the residual effect, is what remains after other components have been removed.

Intro to Time Series Modeling & Forecasting

Time Series Components

The **secular trend** (T_t) is the tendency of the series to increase or decrease over a long period of time. It is also known as the long-term trend.

The **cyclical fluctuation** (C_t) is the wavelike or oscillating pattern about the secular trend. It is also known as a business cycle.

The **seasonal variation** (S_t) describes the fluctuations that recur during specific portions of the year (e.g., monthly or seasonally).

The **residual effect** (R_t) is what remains after the secular, cyclical, and seasonal components have been removed.

Additive model $y_t = T_t + C_t + S_t + R_t$

Intro to Time Series Modeling & Forecasting

Forecasting: The Regression Approach

SALES35.txt

Example: Data on annual sales y for 35 years.

The model is $E(y) = \beta_0 + \beta_1 t$.

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 = 0.4015 + 4.2956t$$

```
              Estimate Std. Error    t value     Pr(>|t|)
(Intercept) 0.4015126  2.2057083  0.1820334 8.566701e-01
t           4.2956303  0.1068669 40.1960721 1.306377e-29

Analysis of Variance Table

Response: y
          Df Sum Sq Mean Sq F value    Pr(>F)
Model      1  65875   65875  1615.7 < 2.2e-16 ***
Residuals 33   1345      41
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

"s,R-squared,adj R-squared"
 6.3852423 0.9799845 0.9793780
```

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	65875	65875	1615.72	<.0001
Error	33	1345.45355	40.77132		
Corrected Total	34	67221			

Root MSE	6.38524	R-Square	0.9800
Dependent Mean	77.72286	Adj R-Sq	0.9794
Coeff Var	8.21540		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	0.40151	2.20571	0.18	0.8567
T	1	4.29563	0.10687	40.20	<.0001

T	SALES	RESIDUAL
1	4.8	0.102857
2	4	-4.99277
3	5.5	-7.7884
4	15.6	-1.98403
5	23.1	1.220336
6	23.3	-2.87529
7	31.4	0.929076
8	46	11.23345
9	46.1	7.037815
10	41.9	-1.45782
11	45.5	-2.15345
12	53.5	1.550924
13	48.4	-7.84471
14	61.6	1.059664
15	65.6	0.764034
16	71.4	2.268403
17	83.4	9.972773
18	93.6	15.87714
19	94.2	12.18151
20	85.4	-0.91412
21	86.2	-4.40975
22	89.9	-5.00538
23	89.2	-10.001
24	99.1	-4.39664
25	100.3	-7.49227
26	111.7	-0.3879
27	108.2	-8.18353
28	115.5	-5.17916
29	119.2	-5.77479
30	125.2	-4.07042
31	136.3	2.73395
32	146.8	8.938319
33	146.1	3.942689
34	151.4	4.947059
35	150.9	0.151429

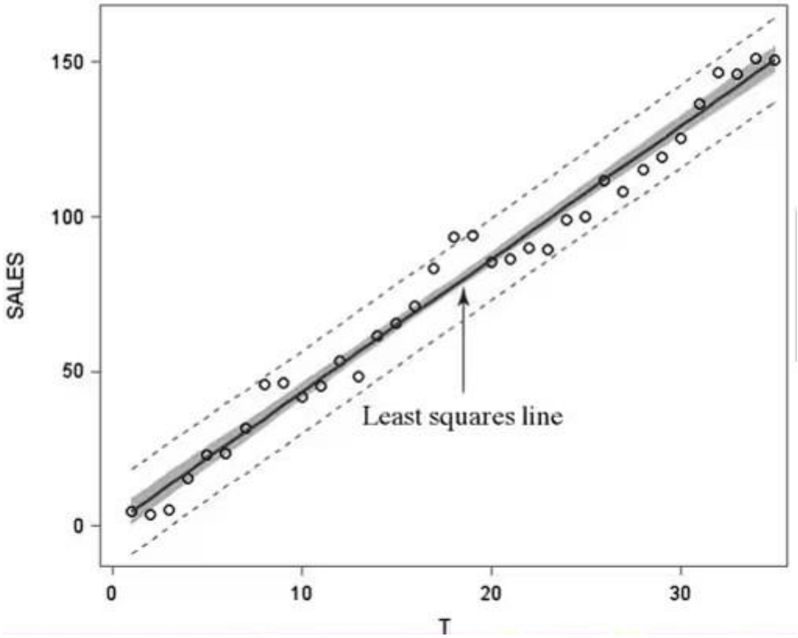
Intro to Time Series Modeling & Forecasting

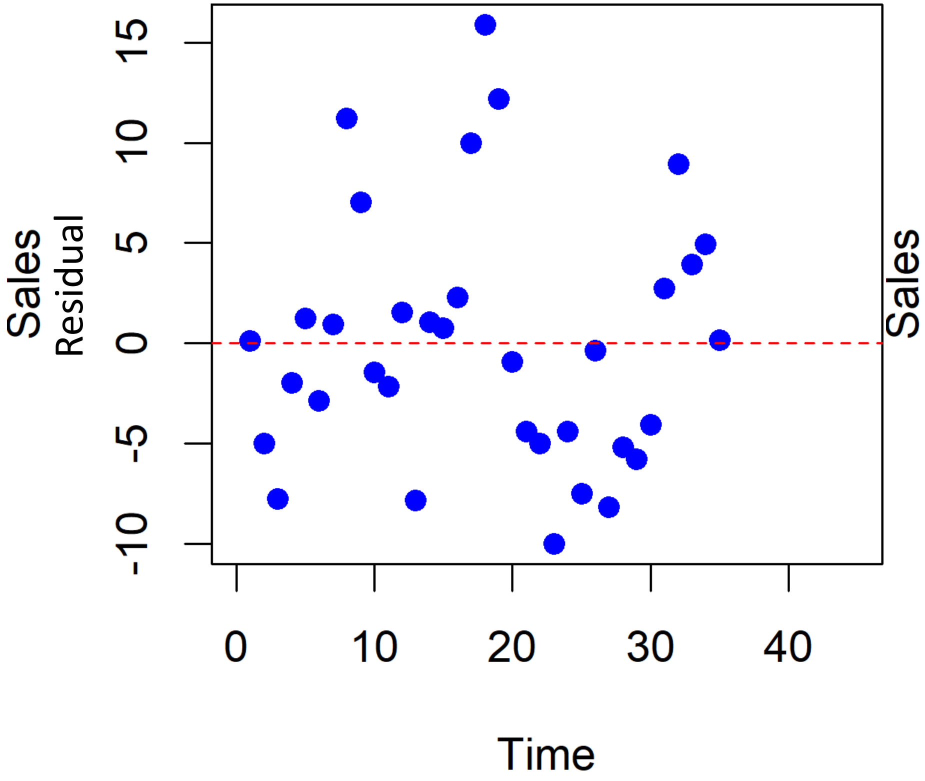
Forecasting: The Regression Approach

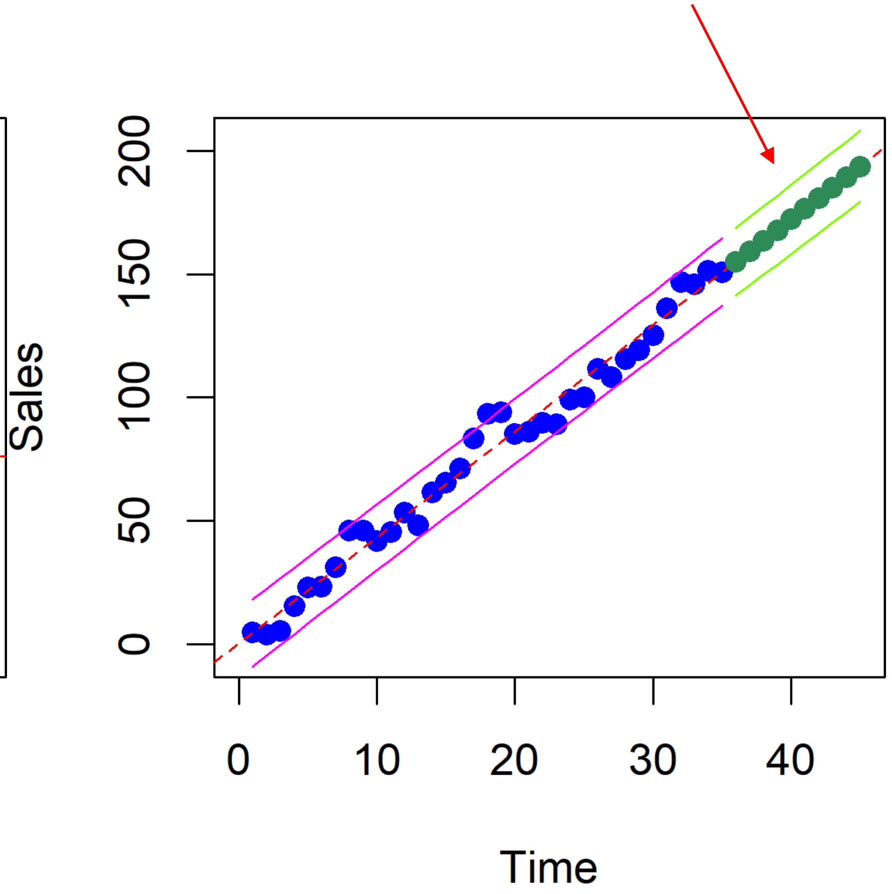
Example: Data on annual sales y for 35 years.

The model is $E(y)=\beta_0+\beta_1 t$.

$$\hat{y}=\hat{\beta}_0+\hat{\beta}_1 x_1 = 0.4015+4.2956t$$





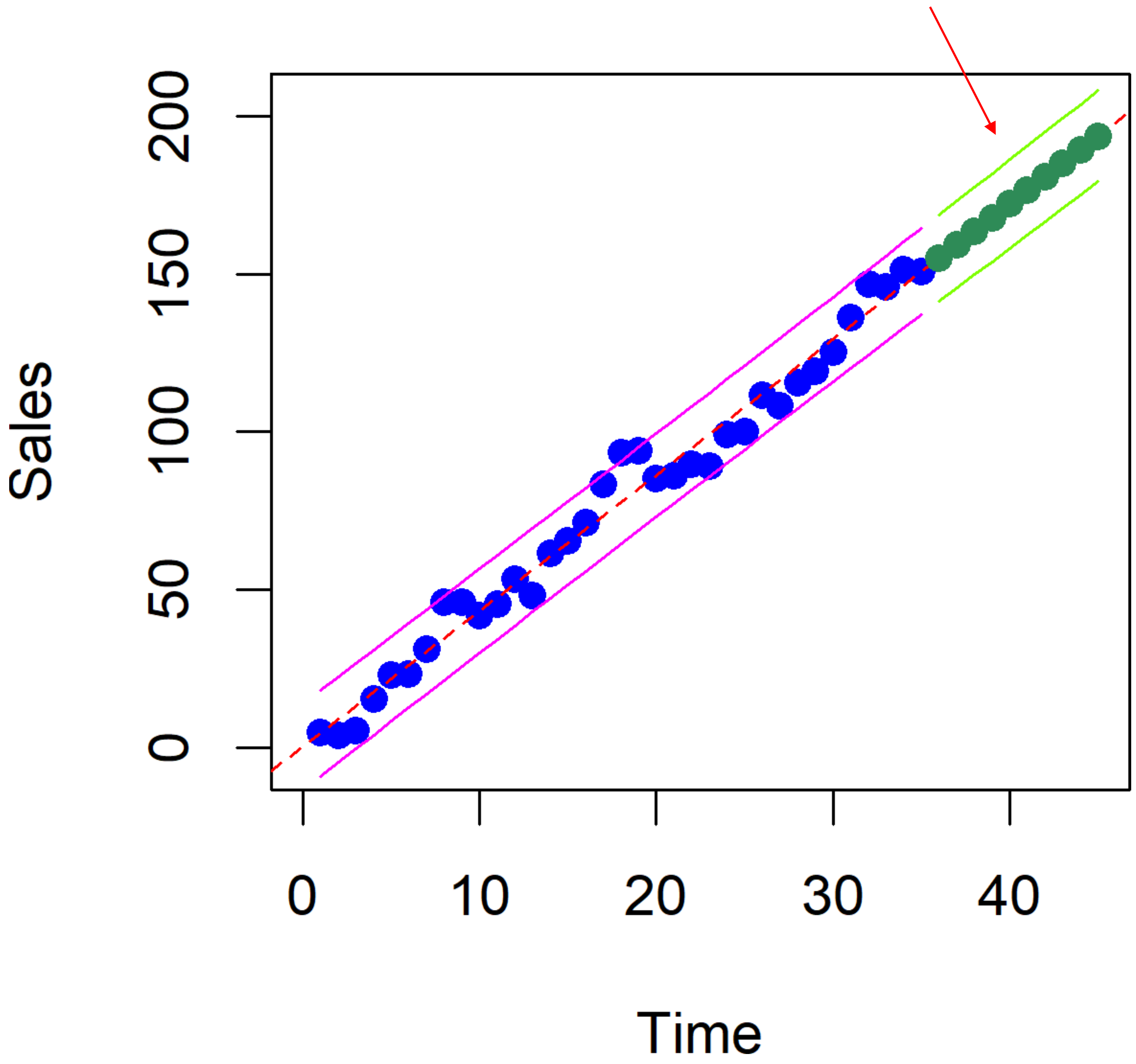


SALES35.txt

T	SALES	RESIDUAL
1	4.8	0.102857
2	4	-4.99277
3	5.5	-7.7884
4	15.6	-1.98403
5	23.1	1.220336
6	23.3	-2.87529
7	31.4	0.929076
8	46	11.23345
9	46.1	7.037815
10	41.9	-1.45782
11	45.5	-2.15345
12	53.5	1.550924
13	48.4	-7.84471
14	61.6	1.059664
15	65.6	0.764034
16	71.4	2.268403
17	83.4	9.972773
18	93.6	15.87714
19	94.2	12.18151
20	85.4	-0.91412
21	86.2	-4.40975
22	89.9	-5.00538
23	89.2	-10.001
24	99.1	-4.39664
25	100.3	-7.49227
26	111.7	-0.3879
27	108.2	-8.18353
28	115.5	-5.17916
29	119.2	-5.77479
30	125.2	-4.07042
31	136.3	2.73395
32	146.8	8.938319
33	146.1	3.942689
34	151.4	4.947059
35	150.9	0.151429

Intro to Time Series Modeling & Forecasting

Forecasting: The Regression Approach



yhat	PI_L	PI_U
4.697	-8.987	18.381
8.993	-4.634	22.619
13.288	-0.284	26.861
17.584	4.062	31.106
21.88	8.405	35.355
26.175	12.744	39.606
30.471	17.08	43.861
34.767	21.413	48.12
39.062	25.743	52.382
43.358	30.068	56.647
47.653	34.391	60.916
51.949	38.709	65.189
56.245	43.025	69.465
60.54	47.337	73.744
64.836	51.645	78.027
69.132	55.949	82.314
73.427	60.25	86.604
77.723	64.548	90.898
82.018	68.842	95.195
86.314	73.132	99.496
90.61	77.418	103.801
94.905	81.702	108.109
99.201	85.981	112.421
103.497	90.257	116.736
107.792	94.53	121.055
112.088	98.798	125.377
116.384	103.064	129.703
120.679	107.326	134.033
124.975	111.584	138.365
129.27	115.839	142.701
133.566	120.091	147.041
137.862	124.339	151.384
142.157	128.585	155.73
146.453	132.826	160.08
150.749	137.065	164.432
155.044	141.3	168.788
159.34	145.532	173.147
163.635	149.761	177.51
167.931	153.987	181.875
172.227	158.21	186.243

T	SALES	PRED_SALES	LOWER95CLI	UPPER95CLI	RESIDUAL
1	4.8	4.697	−8.987	18.381	0.1029
2	4.0	8.993	−4.634	22.619	−4.9928
3	5.5	13.288	−0.284	26.861	−7.7884
4	15.6	17.584	4.062	31.106	−1.9840
5	23.1	21.880	8.405	35.355	1.2203
6	23.3	26.175	12.744	39.606	−2.8753
7	31.4	30.471	17.080	43.861	0.9291
8	46.0	34.767	21.413	48.120	11.2334
9	46.1	39.062	25.743	52.382	7.0378
10	41.9	43.358	30.068	56.647	−1.4578
11	45.5	47.653	34.391	60.916	−2.1534
12	53.5	51.949	38.709	65.189	1.5509
13	48.4	56.245	43.025	69.465	−7.8447
14	61.6	60.540	47.337	73.744	1.0597
15	65.6	64.836	51.645	78.027	0.7640
16	71.4	69.132	55.949	82.314	2.2684
17	83.4	73.427	60.250	86.604	9.9728
18	93.6	77.723	64.548	90.898	15.8771
19	94.2	82.018	68.842	95.195	12.1815
20	85.4	86.314	73.132	99.496	−0.9141
21	86.2	90.610	77.418	103.801	−4.4097
22	89.9	94.905	81.702	108.109	−5.0054
23	89.2	99.201	85.981	112.421	−10.0010
24	99.1	103.497	90.257	116.736	−4.3966
25	100.3	107.792	94.530	121.055	−7.4923
26	111.7	112.088	98.798	125.377	−0.3879
27	108.2	116.384	103.064	129.703	−8.1835
28	115.5	120.679	107.326	134.033	−5.1792
29	119.2	124.975	111.584	138.365	−5.7748
30	125.2	129.270	115.839	142.701	−4.0704
31	136.3	133.566	120.091	147.041	2.7339
32	146.8	137.862	124.339	151.384	8.9383
33	146.1	142.157	128.585	155.730	3.9427
34	151.4	146.453	132.826	160.080	4.9471
35	150.9	150.749	137.065	164.432	0.1514
36	.	155.044	141.300	168.788	.
37	.	159.340	145.532	173.147	.
38	.	163.635	149.761	177.510	.
39	.	167.931	153.987	181.875	.
40	.	172.227	158.210	186.243	.

Intro to Time Series Modeling & Forecasting

Forecasting: The Regression Approach

```
# SALES35 #
alph <- 0.05;
nf    <- 10 # number to forecast

# read data
mydata <- read.delim("SALES35.txt",header=TRUE,sep="",dec=".")

# parse out variables
n <- nrow(mydata)
k <- ncol(mydata)-2
t <- c(mydata[,1]) #T time
y <- c(mydata[,2]) #y sales
r <- c(mydata[,3]) #R ind model residual
tcrit<-qt(1-alph/2,n-k-1)
df      <- data.frame(cbind(t,y,r))
names(df) <- c("t","y","r")
head(df)
df      <- data.frame(matrix(cbind(y,t,Q1,Q2,Q3),n,5))
names(df) <- c("y","t","Q1","Q2","Q3")

# fit line model and get statistics
mymodel <- lm(y~t)
```

```
# scatter plot with line
plot(t,y,pch=19,col="blue",xlab='Time',ylab='Sales',xlim=c(0,n+nf),ylim=c(-5,205))
abline(mymodel,lty=2,col='red')
# residual plot with 0 line
plot(t,r,pch=19,col="blue",xlab='Time',ylab='Sales',xlim=c(0,n+nf))
abline(lm(r~t),r,lty=2,col='red')
#view model fit regression coefficients
summary(mymodel)$coefficients[,]

#view model fit ANOVA table
temp<-anova(mymodel)
out <- temp
m <- nrow(temp)
out$Df <- with(temp,c(sum(Df[1:(m-1)]),Df[m],rep(NA_real_,m-2)))
out$`Sum Sq` <- with(temp,c(sum(`Sum Sq`[1:(m-1)]),
                     `Sum Sq`[m],rep(NA_real_,m-2)))
out$`Mean Sq` <- with(out,out$`Sum Sq`/out$Df)
out$`F value` <- c(out$`Mean Sq`[1]/out$`Mean Sq`[2],rep(NA_real_,m-1))
out$`Pr(>F)` <- c(pf(out$`F value`[1],out$Df[1],out$Df[2],
               lower.tail = FALSE),rep(NA_real_,m-1))
out <- out[1:2,]
rownames(out) <- c("Model","Residuals")
out
```

Intro to Time Series Modeling & Forecasting

Forecasting: The Regression Approach

```
# view model s, Rsq and adjRsq
print('s,R-squared,adj R-squared')
c(summary(mymodel)$s,summary(mymodel)$r.squared,
  summary(mymodel)$adj.r.squared)

# fit and intervals
c   <-rep(1,n)      #Ones
X   <-matrix(cbind(c,t),nrow=n,ncol=2) #design matrix
W   <-solve(t(X)%*%X)
b   <-W%*%t(X)%*%y
yhat<-X%*%b
# mean function at x0
xnew  <-seq(n+1,n+nf,by=1)
X0    <-cbind(rep(1,nf),xnew)
yhatx0<-X0%*%b
fcast <-data.frame(cbind(xnew,yhatx0))
names(fcast) <- c("xnew","yhat")

# scatter plot with line
plot(t,y,pch=19,col="blue",xlab='Time',ylab='Sales',
     xlim=c(0,n+nf),ylim=c(-5,205))
abline(lm(y~t),lty=2,col='red')
points(fcast$xnew,fcast$yhat,pch=19,col='seagreen')
```

```
# mean function confidence interval at x0
xbar<-mean(t)
#MSE<-anova(mymodel)['Residuals','Mean Sq']
SSxx<-sum(t^2)-(sum(t)^2/n)
s<- summary(mymodel)$sigma
# prediction interval
XCI<-cbind(rep(1,n),t)
tXCI <- matrix(t(XCI),2,n)
yCIx0<-XCI%*%b

yCIL <- matrix(rep(0,n),nrow=n,ncol=1) # preallocate
yCIU <- matrix(rep(0,n),nrow=n,ncol=1) # preallocate
yPIL <- matrix(rep(0,n),nrow=n,ncol=1) # preallocate
yPIU <- matrix(rep(0,n),nrow=n,ncol=1) # preallocate
counter<-0
for (i in (1):(n)){
  counter<-counter+1
  yCIL[counter]<-yCIx0[counter]-tcrit*s*sqrt(  1/n+(t[counter]-xbar)^2/SSxx)
  yCIU[counter]<-yCIx0[counter]+tcrit*s*sqrt(  1/n+(t[counter]-xbar)^2/SSxx)
  yPIL[counter]<-yCIx0[counter]-tcrit*s*sqrt(1+1/n+(t[counter]-xbar)^2/SSxx)
  yPIU[counter]<-yCIx0[counter]+tcrit*s*sqrt(1+1/n+(t[counter]-xbar)^2/SSxx) }
withinPI<-cbind(yhat,yPIL,yPIU)
withinPI
```

Intro to Time Series Modeling & Forecasting

Forecasting: The Regression Approach

```
# scatter plot with line and 95% PI
plot(t,y,pch=19,col="blue",xlab='Time',ylab='Sales',
    xlim=c(0,n+nf),ylim=c(-5,205))
abline(lm(y~t),r,lty=2,col='red')
points(fcast$xnew,fcast$yhat,pch=19,col='seagreen')
#lines(t,yCIL, type = "l", col = "deepskyblue3")
#lines(t,yCIU, type = "l", col = "deepskyblue3")
lines(t,yPIL, type = "l", col = "magenta")
lines(t,yPIU, type = "l", col = "magenta")

# forecasted prediction interval
tF <- seq(n+1,n+nf)
XCIF<-cbind(rep(1,nf),tF)
tXCIF <- matrix(t(XCIF),2,nf)
yCIx0F<-XCIF%*%b
yCILF <- matrix(rep(0,nf),nrow=nf,ncol=1) # preallocate
yCIUF <- matrix(rep(0,nf),nrow=nf,ncol=1) # preallocate
yPILF <- matrix(rep(0,nf),nrow=nf,ncol=1) # preallocate
yPIUF <- matrix(rep(0,nf),nrow=nf,ncol=1) # preallocate
```

```
counter<-0
for (i in (1):(nf)){
  counter<-counter+1
  yCILF[counter]<-yCIx0F[counter]-tcrit*s*sqrt(  1/n+(tF[counter]-
xbar)^2/SSxx)
  yCIUF[counter]<-yCIx0F[counter]+tcrit*s*sqrt(  1/n+(tF[counter]-
xbar)^2/SSxx)
  yPILF[counter]<-yCIx0F[counter]-tcrit*s*sqrt(1+1/n+(tF[counter]-
xbar)^2/SSxx)
  yPIUF[counter]<-yCIx0F[counter]+tcrit*s*sqrt(1+1/n+(tF[counter]-
xbar)^2/SSxx)
}
outsidePI<- cbind(yhatx0,yPILF,yPIUF)
outsidePI
```

Intro to Time Series Modeling & Forecasting

Forecasting: The Regression Approach

```
# scatter plot with line and 95% PI
plot(t,y,pch=19,col="blue",xlab='Time',ylab='Sales',
    xlim=c(0,n+nf),ylim=c(-5,205))
abline(lm(y~t),r,lty=2,col='red')
points(fcast$xnew,fcast$yhat,pch=19,col='seagreen')
#lines(t,yCIL, type = "l", col = "deepskyblue3")
#lines(t,yCIU, type = "l", col = "deepskyblue3")
lines(t,yPIL, type = "l", col = "magenta")
lines(t,yPIU, type = "l", col = "magenta")
lines(tF,yPILF, type = "l", col = "chartreuse1")
lines(tF,yPIUF, type = "l", col = "chartreuse1")

results<-data.frame(rbind(withinPI,outsidePI))
names(results) <- c("yhat","PI_L","PI_U")

write.csv(round(results,digits=3),file="forecasteandpiSales35.csv")
```

Intro to Time Series Modeling & Forecasting

Forecasting: The Regression Approach

QTRPOWER.txt

YEAR	T	POWLOAD
2015	1	103.5
2015	2	94.7
2015	3	118.6
2015	4	109.3
2016	5	126.1
2016	6	116
2016	7	141.2
2016	8	131.6
2017	9	144.5
2017	10	137.1
2017	11	159
2017	12	149.5
2018	13	166.1
2018	14	152.5
2018	15	178.2
2018	16	169
2019	17*	

Example: Quarterly power loads.

The model is $E(y)= \underbrace{\beta_0+\beta_1 t}_{\text{Secular Trend}} + \underbrace{\beta_1 Q_1+\beta_1 Q_2+\beta_1 Q_3}_{\text{Seasonal Component}}$.

t=time period from t=1 for quarter I of 2015 to t=16 for quarter IV of 2018.

$$Q_1 = \begin{cases} 1 \; if \; quarter \; I \\ 0 \;\; if \; not \end{cases} \qquad Q_2 = \begin{cases} 1 \; if \; quarter \; II \\ 0 \;\; if \; not \end{cases}$$

$$Q_3 = \begin{cases} 1 \; if \; quarter \; III \\ 0 \;\; if \; not \end{cases} \qquad Base\ Level = quarter\ IV$$

Intro to Time Series Modeling & Forecasting

Forecasting: The Regression Approach

Example: Quarterly power loads.

The model is $E(y)= \beta_0+\beta_1 t + \beta_1 Q_1+\beta_1 Q_2+\beta_1 Q_3$.

$\beta_0+\beta_1 t$: Secular Trend

$\beta_1 Q_1+\beta_1 Q_2+\beta_1 Q_3$: Seasonal Component

```
             Estimate Std. Error   t value     Pr(>|t|)
(Intercept) 90.206250  1.1493140 78.487039 1.787868e-16
t            4.964375  0.0856648 57.951162 4.990600e-15
Q1          10.093125  1.1136425  9.063165 1.957415e-06
Q2          -4.846250  1.0970448 -4.417550 1.032699e-03
Q3          14.364375  1.0869645 13.215128 4.290807e-08
```

```
Response: y
          Df Sum Sq Mean Sq F value    Pr(>F)
Model      4 9101.7 2275.42  968.96 6.262e-14 ***
Residuals 11   25.8    2.35
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
"s,R-squared,adj R-squared"
 1.5324186 0.9971699 0.9961408
```

QTRPOWER.txt

YEAR	T	POWLOAD
2015	1	103.5
2015	2	94.7
2015	3	118.6
2015	4	109.3
2016	5	126.1
2016	6	116
2016	7	141.2
2016	8	131.6
2017	9	144.5
2017	10	137.1
2017	11	159
2017	12	149.5
2018	13	166.1
2018	14	152.5
2018	15	178.2
2018	16	169
2019	17*	

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	9101.67800	2275.41950	968.96	<.0001
Error	11	25.83138	2.34831		
Corrected Total	15	9127.50938			

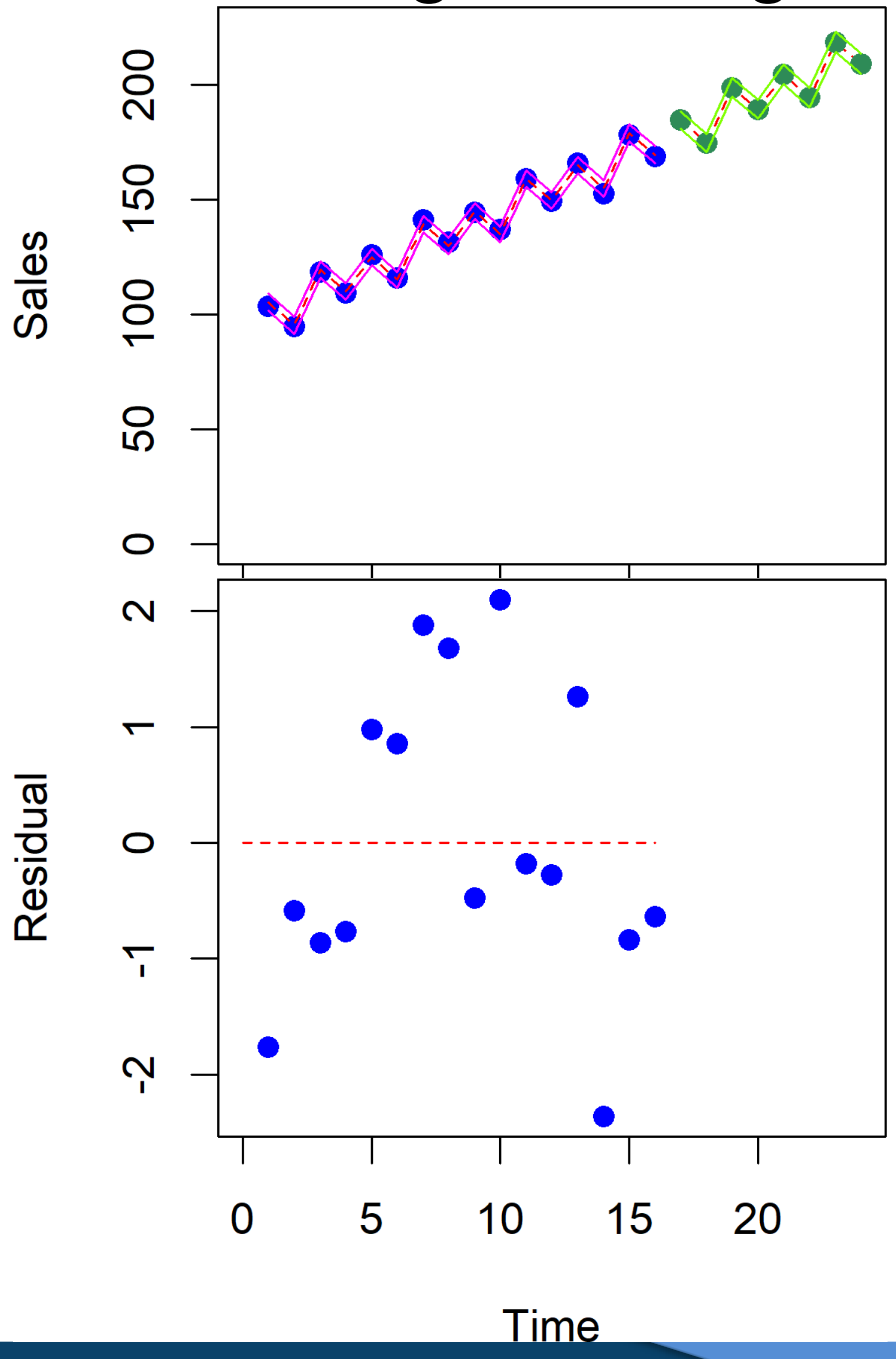
Root MSE	1.53242	R-Square	0.9972
Dependent Mean	137.30625	Adj R-Sq	0.9961
Coeff Var	1.11606		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	90.20625	1.14931	78.49	<.0001
T	1	4.96438	0.08566	57.95	<.0001
Q1	1	10.09313	1.11364	9.06	<.0001
Q2	1	–4.84625	1.09704	–4.42	0.0010
Q3	1	14.36438	1.08696	13.22	<.0001

Intro to Time Series Modeling & Forecasting

Forecasting: The Regression Approach



t	yhat	PI_L	PI_U
1	105.264	101.526	109.001
2	95.289	91.614	98.963
3	119.464	115.844	123.083
4	110.064	106.491	113.636
5	125.121	121.586	128.656
6	115.146	111.64	118.653
7	139.321	135.834	142.809
8	129.921	126.443	133.399
9	144.979	141.501	148.457
10	135.004	131.516	138.491
11	159.179	155.672	162.685
12	149.779	146.244	153.314
13	164.836	161.264	168.409
14	154.861	151.242	158.481
15	179.036	175.362	182.711
16	169.636	165.899	173.374
17	184.694	180.885	188.502
18	174.719	170.832	178.605
19	198.894	194.922	202.866
20	189.494	185.43	193.557

Output Statistics

Obs	YEAR_QTR	Dependent Variable	Predicted Value	Std Error Mean Predict	95% CL Predict		Residual
1	2015_ 1	103.5	105.2638	0.9226	101.3268	109.2007	−1.7637
2	2015_ 2	94.7	95.2887	0.9226	91.3518	99.2257	−0.5887
3	2015_ 3	118.6	119.4638	0.9226	115.5268	123.4007	−0.8637
4	2015_ 4	109.3	110.0637	0.9226	106.1268	114.0007	−0.7637
5	2016_ 5	126.1	125.1213	0.7851	121.3315	128.9110	0.9788
6	2016_ 6	116.0	115.1462	0.7851	111.3565	118.9360	0.8538
7	2016_ 7	141.2	139.3212	0.7851	135.5315	143.1110	1.8788
8	2016_ 8	131.6	129.9212	0.7851	126.1315	133.7110	1.6788
9	2017_ 9	144.5	144.9788	0.7851	141.1890	148.7685	−0.4787
10	2017_10	137.1	135.0038	0.7851	131.2140	138.7935	2.0962
11	2017_11	159.0	159.1787	0.7851	155.3890	162.9685	−0.1787
12	2017_12	149.5	149.7787	0.7851	145.9890	153.5685	−0.2787
13	2018_13	166.1	164.8363	0.9226	160.8993	168.7732	1.2637
14	2018_14	152.5	154.8612	0.9226	150.9243	158.7982	−2.3612
15	2018_15	178.2	179.0362	0.9226	175.0993	182.9732	−0.8362
16	2018_16	169.0	169.6362	0.9226	165.6993	173.5732	−0.6362
17	2019_17	.	184.6938	1.1493	180.4777	188.9098	.
18	2019_18	.	174.7188	1.1493	170.5027	178.9348	.
19	2019_19	.	198.8938	1.1493	194.6777	203.1098	.
20	2019_20	.	189.4937	1.1493	185.2777	193.7098	.

QTRPOWER.txt

YEAR	T	POWLOAD
2015	1	103.5
2015	2	94.7
2015	3	118.6
2015	4	109.3
2016	5	126.1
2016	6	116
2016	7	141.2
2016	8	131.6
2017	9	144.5
2017	10	137.1
2017	11	159
2017	12	149.5
2018	13	166.1
2018	14	152.5
2018	15	178.2
2018	16	169
2019	17*	

$$\hat{y} = 90.206250 + 4.964375t + 10.093125Q_1 - 4.846250Q_2 + 14.364375Q_3$$

Intro to Time Series Modeling & Forecasting

Forecasting: The Regression Approach

```
# QTRPOWER #
alph <- 0.05;
nf   <-8 # number to forecast

# read data
mydata <- read.delim("QTRPOWER.txt",header=TRUE,sep="",dec=".")

# parse out variables
n <- nrow(mydata[1:16,])
k <- 4
y    <- c(mydata[1:16,3]) #y power
year <- c(mydata[1:16,1]) #year
t    <- c(mydata[1:16,2]) #t time
Q1 <- c(1,0,0,0,1,0,0,0,1,0,0,0,1,0,0,0)
Q2 <- c(0,1,0,0,0,1,0,0,0,1,0,0,0,1,0,0)
Q3 <- c(0,0,1,0,0,0,1,0,0,0,1,0,0,0,1,0)
tcrit<-qt(1-alph/2,n-k-1)

df       <- data.frame(matrix(cbind(y,t,Q1,Q2,Q3),n,5))
names(df) <- c("y","t","Q1","Q2","Q3")

# fit line model and get statistics
mymodel <- lm(y~t+Q1+Q2+Q3)
b<-matrix(summary(mymodel)$coefficients[,1],k+1,1)
```

```
# view model fit regression coefficients
summary(mymodel)$coefficients[,]

# scatter plot with line
c   <-rep(1,n)      #Ones
X   <-cbind(c,t,Q1,Q2,Q3) #design matrix
plot(t,y,pch=19,col="blue",xlab='Time',ylab='Power',xlim=c(0,n+nf),ylim=c(0,225))
lines(t,X%*%b,type="l",lty=2,col='red')

# residual plot with 0 line
r<-mymodel$residuals
plot(t,mymodel$residuals,pch=19,col="blue",xlab='Time',ylab='Residual',xlim=c(0,
n+nf))
lines(c(0,n),c(0,0),type="l",lty=2,col='red')

#view model fit ANOVA table
temp<-anova(mymodel)
out <- temp
m <- nrow(temp)
out$Df <- with(temp,c(sum(Df[1:(m-1)]),Df[m],rep(NA_real_,m-2)))
out$`Sum Sq` <- with(temp,c(sum(`Sum Sq`[1:(m-1)]),
                     `Sum Sq`[m],rep(NA_real_,m-2)))
out$`Mean Sq` <- with(out,out$`Sum Sq`/out$Df)
```

Intro to Time Series Modeling & Forecasting

Forecasting: The Regression Approach

```
out$`F value` <- c(out$`Mean Sq`[1]/out$`Mean
Sq`[2],rep(NA_real_,m-1))
out$`Pr(>F)` <- c(pf(out$`F value`[1],out$Df[1],out$Df[2],
          lower.tail = FALSE),rep(NA_real_,m-1))
out <- out[1:2,]
rownames(out) <- c("Model","Residuals")
out

# view model s, Rsq and adjRsq
print('s,R-squared,adj R-squared')
c(summary(mymodel)$s,summary(mymodel)$r.squared,
  summary(mymodel)$adj.r.squared)

# fit and intervals
yhat<-X%*%b

# mean function at x0
Q1 <- c(1,0,0,0)
Q2 <- c(0,1,0,0)
Q3 <- c(0,0,1,0)
xnew  <-cbind(seq(n+1,n+nf,by=1),Q1,Q2,Q3)
X0    <-cbind(rep(1,nf),xnew)
yhatx0<-X0%*%b
```

```
fcast <-data.frame(cbind(xnew,yhatx0))
names(fcast) <- c("xnew","yhat")

# scatter plot with line
plot(t,y,pch=19,col="blue",xlab='Time',ylab='Sales',
   xlim=c(0,n+nf),ylim=c(0,225))
lines(t,X%*%b,type="l",lty=2,col='red')
lines(seq(n+1,n+nf,by=1),yhatx0,type="l",lty=2,col='red')
points(seq(n+1,n+nf,by=1),yhatx0,pch=19,col='seagreen')

# mean function confidence interval at x0
xbar<-mean(t)
#MSE<-anova(mymodel)['Residuals','Mean Sq']
SSxx<-sum(t^2)-(sum(t)^2/n)
s<- summary(mymodel)$sigma

# prediction interval
XCI  <- X
tXCI <- matrix(t(XCI),5,n)
yCIx0<-XCI%*%b
```

Intro to Time Series Modeling & Forecasting

Forecasting: The Regression Approach

```
yCIL <- matrix(rep(0,n),nrow=n,ncol=1) # preallocate
yCIU <- matrix(rep(0,n),nrow=n,ncol=1) # preallocate
yPIL <- matrix(rep(0,n),nrow=n,ncol=1) # preallocate
yPIU <- matrix(rep(0,n),nrow=n,ncol=1) # preallocate
counter<-0
for (i in (1):(n)){
  counter<-counter+1
  yCIL[counter]<-yCIx0[counter]-tcrit*s*sqrt(  1/n+(t[counter]-
xbar)^2/SSxx)
  yCIU[counter]<-yCIx0[counter]+tcrit*s*sqrt(  1/n+(t[counter]-
xbar)^2/SSxx)
  yPIL[counter]<-yCIx0[counter]-tcrit*s*sqrt(1+1/n+(t[counter]-
xbar)^2/SSxx)
  yPIU[counter]<-yCIx0[counter]+tcrit*s*sqrt(1+1/n+(t[counter]-
xbar)^2/SSxx)
}
withinPI<-cbind(yhat,yPIL,yPIU)
withinPI

# scatter plot with line and 95% PI
plot(t,y,pch=19,col="blue",xlab='Time',ylab='Sales',
     xlim=c(0,n+nf),ylim=c(0,225))
lines(t,X%*%b,type="l",lty=2,col='red')
lines(seq(n+1,n+nf,by=1),yhatx0,type="l",lty=2,col='red')
points(seq(n+1,n+nf,by=1),yhatx0,pch=19,col='seagreen')
lines(t,yPIL, type = "l", col = "magenta")
lines(t,yPIU, type = "l", col = "magenta")
# forecasted prediction interval
tF <- seq(n+1,n+nf)
XCIF<-cbind(rep(1,nf),xnew)
tXCIF <- matrix(t(XCIF),2,nf)
yCIx0F<-XCIF%*%b
yCILF <- matrix(rep(0,nf),nrow=nf,ncol=1) # preallocate
yCIUF <- matrix(rep(0,nf),nrow=nf,ncol=1) # preallocate
yPILF <- matrix(rep(0,nf),nrow=nf,ncol=1) # preallocate
yPIUF <- matrix(rep(0,nf),nrow=nf,ncol=1) # preallocate
counter<-0
for (i in (1):(nf)){
  counter<-counter+1
  yCILF[counter]<-yCIx0F[counter]-tcrit*s*sqrt(  1/n+(tF[counter]-xbar)^2/SSxx)
  yCIUF[counter]<-yCIx0F[counter]+tcrit*s*sqrt(  1/n+(tF[counter]-xbar)^2/SSxx)
  yPILF[counter]<-yCIx0F[counter]-tcrit*s*sqrt(1+1/n+(tF[counter]-xbar)^2/SSxx)
  yPIUF[counter]<-yCIx0F[counter]+tcrit*s*sqrt(1+1/n+(tF[counter]-xbar)^2/SSxx)
}
outsidePI<- cbind(yhatx0,yPILF,yPIUF)
outsidePI
```

Intro to Time Series Modeling & Forecasting

Forecasting: The Regression Approach

```
# scatter plot with line and 95% PI
plot(t,y,pch=19,col="blue",xlab='Time',ylab='Sales',
    xlim=c(0,n+nf),ylim=c(0,225))
lines(t,X%*%b,type="l",lty=2,col='red')
lines(seq(n+1,n+nf,by=1),yhatx0,type="l",lty=2,col='red')
points(seq(n+1,n+nf,by=1),yhatx0,pch=19,col='seagreen')
#lines(t,yCIL, type = "l", col = "deepskyblue3")
#lines(t,yCIU, type = "l", col = "deepskyblue3")
lines(t,yPIL, type = "l", col = "magenta")
lines(t,yPIU, type = "l", col = "magenta")
lines(tF,yPILF, type = "l", col = "chartreuse1")
lines(tF,yPIUF, type = "l", col = "chartreuse1")

results<-data.frame(rbind(withinPI,outsidePI))
names(results) <- c("yhat","PI_L","PI_U")

write.csv(round(results,digits=3),file="forecasteandpiPower.csv")
```

Intro to Time Series Modeling & Forecasting

Autocorrelation and Autoregressive Error Models

Autocorrelation is the correlation between time series residuals at different points in time. The special case in which neighboring residuals one time period apart (at times t and $t+1$) are correlated is called **first-order autocorrelation**.

In general, ***m*th-order autocorrelation** occurs when residuals at times t and $(t+m)$ are correlated.

The model is $y_t=E(y_t)+R_t$ where $E(y_t)=\beta_0+\beta_1 x_1+\ldots+\beta_k x_k$ and R_t represents the random residual. We assume that the residual R_t has mean 0 and constant variance σ^2, but that it is autocorrelated.

Intro to Time Series Modeling & Forecasting

Autocorrelation and Autoregressive Error Models

A property commonly observed for autocorrelated residuals is that the size of the autocorrelation between values of the residual R at two different points in time diminishes rapidly as the distance between the time points increases.

Thus, the autocorrelation between R_t and R_{t+1} becomes smaller (i.e., weaker) as the distance m between the time points becomes larger.

First-order autoregressive error model: $R_t = \phi R_{t+m} + \varepsilon_t$, $-1 < \phi < 1$.

A consequence of which is that $\mathrm{AC}(R_t, R_{t+m}) = \phi^m$.

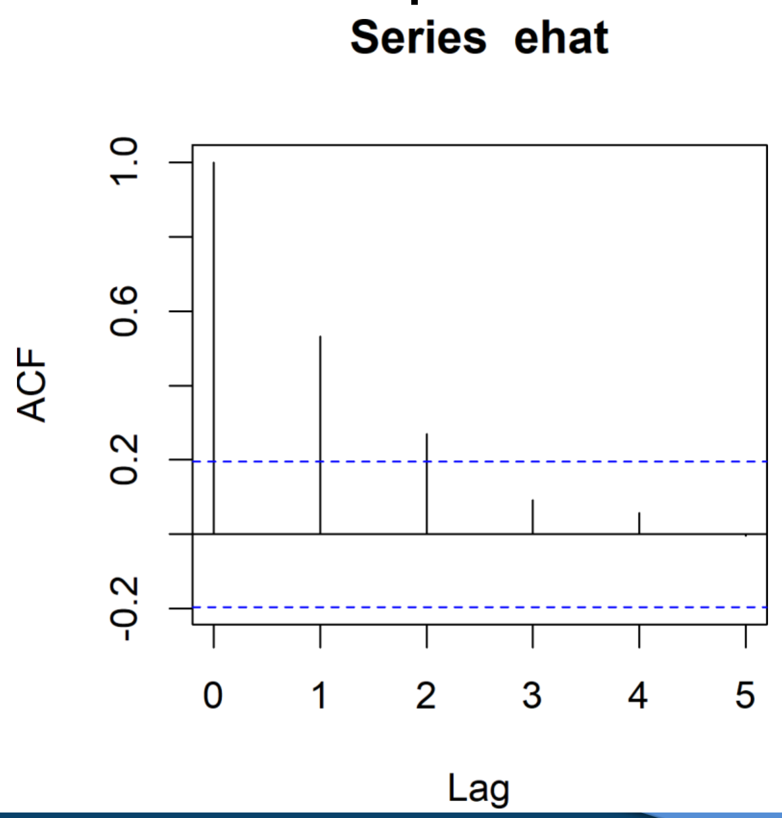
Intro to Time Series Modeling & Forecasting

Autocorrelation and Autoregressive Error Models

First-order autoregressive error model: $R_t = \phi R_{t+1} + \varepsilon_t$, $-1 < \phi < 1$.

A consequence of which is that $AC(R_t, R_{t+m}) = \phi^m$.

Recall: Chapter 8b worksheet with

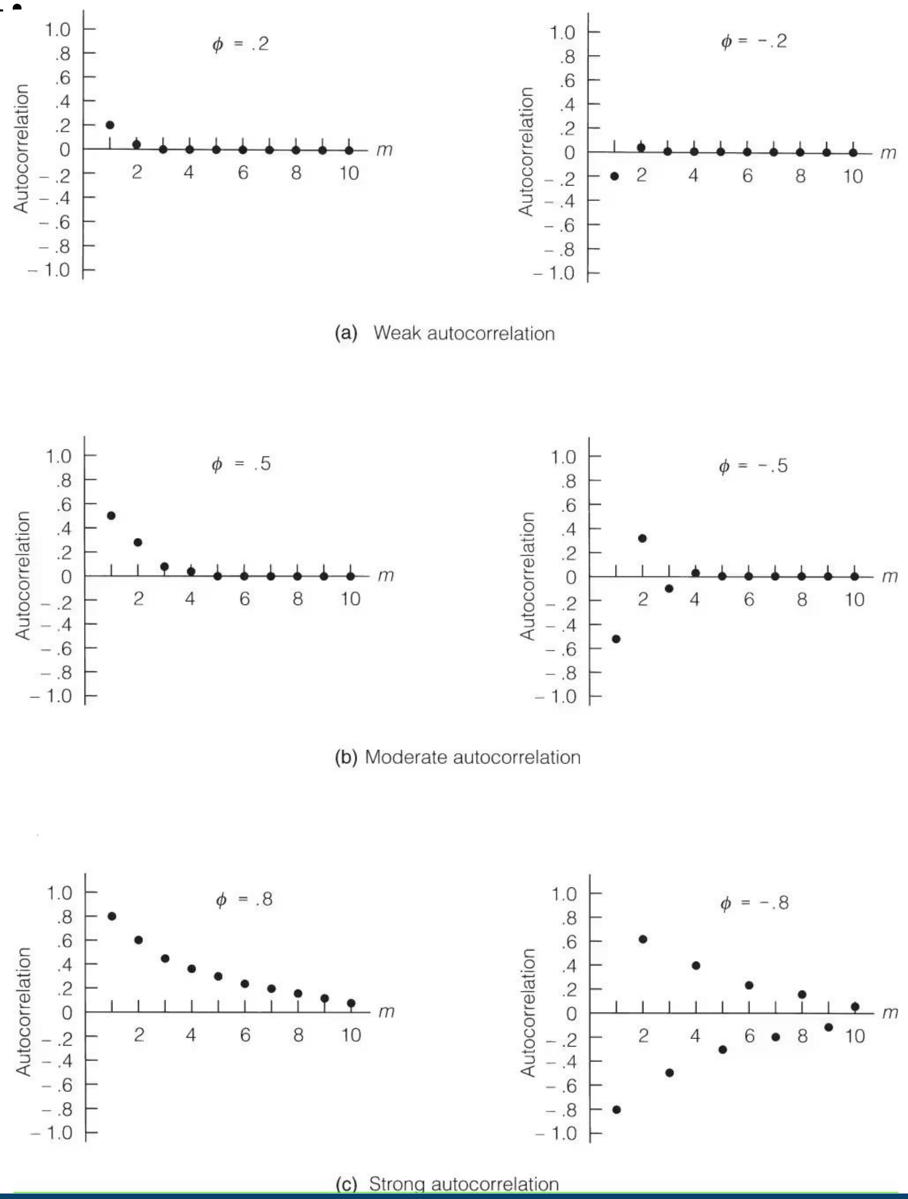


ϕ =0.5. y[i]=b0+b1*t[i]+phi*y[i-1]+ e[i]

$AC(R_t, R_{t+m}) = (0.5)^m$, m=0,1,2,3,4,5.

1.0, 0.5, 0.25, 0.125, 0.0625, 0.0312

Durbin-Watson Test



(a) Weak autocorrelation

(b) Moderate autocorrelation

(c) Strong autocorrelation

Intro to Time Series Modeling & Forecasting

Autocorrelation and Autoregressive Error Models

First-order autoregressive error model: $R_t = \phi R_{t+1} + \varepsilon_t$, $-1 < \phi < 1$.

A consequence of which is that $\text{AC}(R_t, R_{t+m}) = \phi^m$.

Recall: Chapter 8b worksheet with

$$d = \sum_{t=2}^{n} (\hat{\varepsilon}_t - \hat{\varepsilon}_{t-1})^2 \Big/ \sum_{t=1}^{n} \hat{\varepsilon}_t^2 \approx \underbrace{2(1-\hat{\phi})}_{\text{Large } n} \quad , 0 \leq d \leq 4.$$

H_0: $\phi \leq 0$ vs. H_a: $\phi > 0$, Reject $d < d_{L,\alpha}$

H_0: $\phi \geq 0$ vs. H_a: $\phi < 0$, Reject $(4-d) < d_{L,\alpha}$

H_0: $\phi = 0$ vs. H_a: $\phi \neq 0$, Reject $d < d_{L,\alpha/2}$ or $(4-d) < d_{L,\alpha/2}$

Intro to Time Series Modeling & Forecasting

Homework:
Read Chapter 10
Problems #: 8 (INTRATE30), 11 (GRAPHICAL), 16

Submit at minimum one file with all your answers and another with your code.

Intro to Time Series Modeling & Forecasting

Questions?