MATH 2780 Chapter 10A Worksheet

Summary

Repeated observations on a variable y_t at times $t=1,\ldots,n$ produce a **time series**. This time series is decomposed into a long-term secular trend (T_t), a cyclical business cycle wavelike fluctuation (C_t), a seasonal variation (S_t), and a residual effect (R_t), the additive regression model is $y_t = T_t + C_t + S_t + R_t$.
Residuals one time period apart (at times t and t+1) are correlated is called **first-order autocorrelation**.
Model is $y_t=E(y_t)+R_t$, $E(y_t)=\beta_0+\beta_1 x_1+\ldots+\beta_k x_k$, residual R_t has $E(R_t)=0$ and $Var(R_t)=\sigma^2$, but autocorrelated.
Quarterly model $E(y)= \beta_0+\beta_1 t+\beta_1 Q_1+\beta_1 Q_2+\beta_1 Q_3$, $Q_1=1$ if quarter 1, $Q_2=1$ if quarter 2, $Q_3=1$, if quarter 3.
Coefficient and Residual Variance Estimation: The ordinary least squares regression coefficients.

$$Y = X\beta + E$$
$$\hat{\beta} = (X'X)^{-1}X'y$$
$$s^2 = \frac{(y - X\hat{\beta})'(y - X\hat{\beta})}{n-k-1}$$
$$MSE = s^2,\ s = \sqrt{s^2}$$

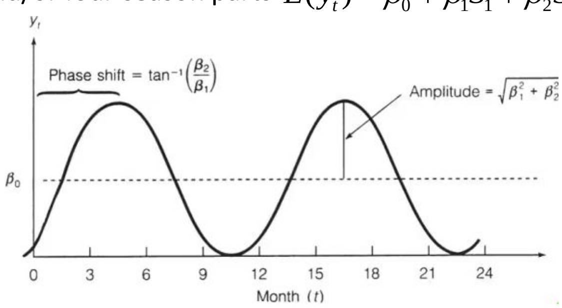
$$Y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{bmatrix}, X = \begin{bmatrix} 1 & x_{11} & x_{21} & \cdots & x_{kn} \\ 1 & x_{12} & x_{22} & \cdots & x_{k2} \\ 1 & x_{13} & x_{23} & \cdots & x_{k3} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{1n} & x_{2n} & \cdots & x_{kn} \end{bmatrix}, \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix}, E = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

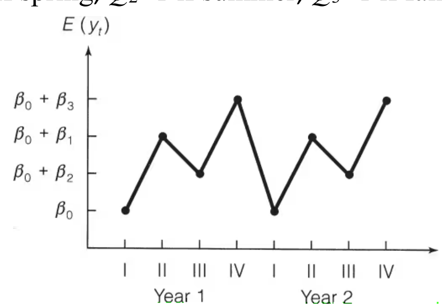
Regression Residuals: Residuals are $\hat{\varepsilon}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_1 - \ldots - \hat{\beta}_k x_k$, $s^2 = \sum (y_i - \hat{y}_i)^2 / (n-k-1)$.
All of the same methods we learned to work with residuals continue to apply.
First-order autoregressive error model: $m=1$, $R_t=\phi R_{t+m}+\varepsilon_t$, $-1<\phi<1$, $AC(R_t,R_{t+m})=\phi^m$.
Forecast limits using AR(1) error model: $\hat{y}_{n+m} \pm 1.96\sqrt{MSE(1+\phi^2+\cdots+\phi^{2(m-1)})}$
Detecting Residual Correlation: The **Durbin-Watson Test**.

$d = \sum_{t=2}^{n}(\hat{\varepsilon}_t - \hat{\varepsilon}_{t-1})^2 / \sum_{t=1}^{n}\hat{\varepsilon}_t^2 \approx \underbrace{2(1-\hat{\phi})}_{\text{Large } n}$	H_0: $\phi\le 0$ vs. H_a: $\phi>0$, Reject $d<d_{L,\alpha}$ H_0: $\phi\ge 0$ vs. H_a: $\phi<0$, Reject $(4-d)<d_{L,\alpha}$ H_0: $\phi=0$ vs. H_a: $\phi\ne 0$, Reject $d<d_{L,\alpha/2}$ or $(4-d)<d_{L,\alpha/2}$

1. Range of d: $0 \le d \le 4$.
2. If residuals are uncorrelated, $d \approx 2$.
3. If residuals are positively correlated, $d < 2$, and if the correlation is very strong, $d\approx 0$.
4. If residuals are negatively correlated, $d > 2$, and if the correlation is very strong, $d\approx 4$.

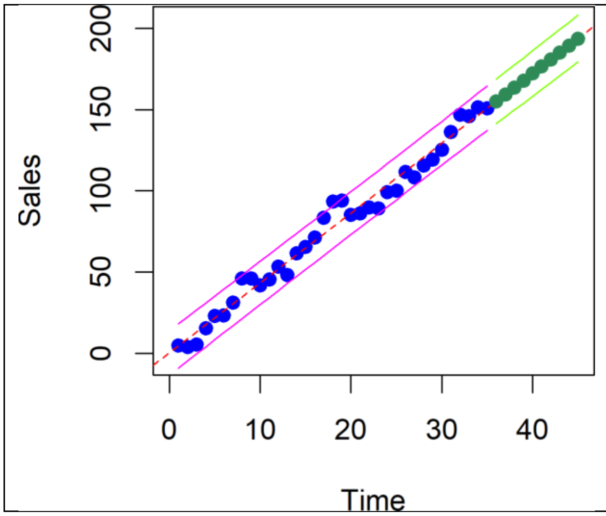
Often the deterministic part has seasonal patterns $E(y_t) = \beta_0 + \beta_1\cos(2\pi t/T) + \beta_2\sin(2\pi t/T)$
and/or four-season parts $E(y_t) = \beta_0 + \beta_1 S_1 + \beta_2 S_2 + \beta_3 S_3$, $Q_1=1$ if spring, $Q_2=1$ if summer, $Q_3=1$ if fall.

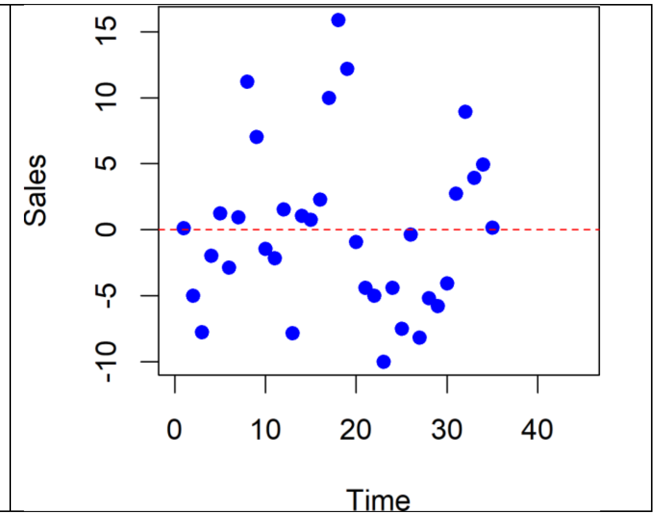




while the residual part R_t depends on the pattern of autocorrelation $R_t = \phi_1 R_{t-1} + \cdots + \phi_1 R_{t-p} + \varepsilon_t$, $E(\varepsilon_t)=0$, $Var(\varepsilon_t)=\sigma^2$, $Cor(\varepsilon_t,\varepsilon_{t-1})=0$. The simplest is AR(1) $R_t = \phi R_{t-1} + \varepsilon_t$. Estimate $\hat{\phi} = \left[\sum \varepsilon_t^2/(n-1)\right]/s_\varepsilon^2$.
We transform to a decorrelated model $y_t^* = \beta_0^* + \beta_1^* t^* + \varepsilon_t$, $\beta_0^* = (1-\hat{\phi})\beta_0$, $\beta_1^* = \beta_1$, $t^* = t - \hat{\phi}(t-1)$.
Estimate using $\tilde{\beta}^* = (X^{*\prime}X^*)^{-1}X^{*\prime}y^*$, then transform back $\tilde{\beta}_0 = \tilde{\beta}_0^*/(1-\tilde{\phi})$, $\tilde{\beta}_1 = \tilde{\beta}_1^*$, $\tilde{\phi} = \hat{\phi}$.
Interpolation: $R_t = y_t - \tilde{\beta}_0 - \tilde{\beta}_1 t$, $\tilde{y}_t = \tilde{\beta}_0 + \tilde{\beta}_1 t + \tilde{\phi}R_{t-1}$, PI: $\tilde{y}_t \pm t_{\alpha,n-k-1}\sqrt{MSE}$,
Extrapolation: $\tilde{R}_n = y_n - \tilde{\beta}_0 - \tilde{\beta}_1 x_n$, $F_{n+1} = \tilde{\beta}_0 + \tilde{\beta}_1 x_{n+1} + \hat{\phi}\tilde{R}_n$, $\tilde{R}_{n+1} = \hat{\phi}\tilde{R}_n$, $F_{n+2} = \tilde{\beta}_0 + \tilde{\beta}_1 x_{n+2} + \hat{\phi}\tilde{R}_{n+1}$,
PI: $\hat{y}_{n+m} \pm t_{\alpha,n-k-1}\sqrt{MSE\sum_{i=1}^{m}\hat{\phi}^{2(i-1)}}$.

Example: Data on annual sales *y* for 35 years. The model is $E(y)=\beta_0+\beta_1 t$. y=annual sales, t=time





SALES35

read data

parse out variables

fit uncorrelated line model and get statistics

get uncorrelated error regression coefficients

scatter plot with uncorrelated error line

uncorrelated error residual plot with 0 line

#AR 1 correlation from uncorrelated error model

fit correlated line model and get statistics

get correlated error regression coefficients

fit and intervals

scatter plot with uncorrelated error line

#view model fit ANOVA table

view model s, Rsq and adjRsq

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

MATH 2780 Chapter 10A Worksheet

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

# parse out variables
n <- nrow(mydata)
t <- c(mydata[,1]) #T time
y <- c(mydata[,2]) #y sales
r <- c(mydata[,3]) #r residual

df      <- data.frame(cbind(t,y,r))
names(df) <- c("t","y","r")

# fit uncorrelated line model and get statistics
mymodel1 <- lm(y~t)

# get uncorrelated error regression coefficients
b0<-mymodel1$coefficients[1]
b1<-mymodel1$coefficients[2]
c(b0,b1)

# scatter plot with uncorrelated error line
plot(t,y,pch=19,col="blue",xlab='Time',ylab='Sales',
xlim=c(0,n))
abline(mymodel1,lty=2,col='red')

# uncorrelated error residual plot with 0 line
plot(t,r,pch=19,col="blue",xlab='Time',ylab='Sales',
xlim=c(0,n))
abline(lm(r~t),r,lty=2,col='red')

#AR 1 correlation from uncorrelated error model
phihat<-cor(r[1:(n-1)],r[(1+1):n])
phihat

# fit correlated line model and get statistics
yast<-y[2:n]-phihat*y[1:n-1]
tast<-t[2:n]-phihat*t[1:n-1]
mymodel2 <- lm(yast~tast)

# get correlated error regression coefficients
b0ast<-mymodel2$coefficients[1]/(1-phihat)
b1ast<-mymodel2$coefficients[2]
bast<-c(b0ast,b1ast)
bast

# fit and intervals
c  <-rep(1,n)     #Ones
X  <-matrix(cbind(c,t),nrow=n,ncol=2) #design ma-
trix
yasthat<-X%*%bast
```

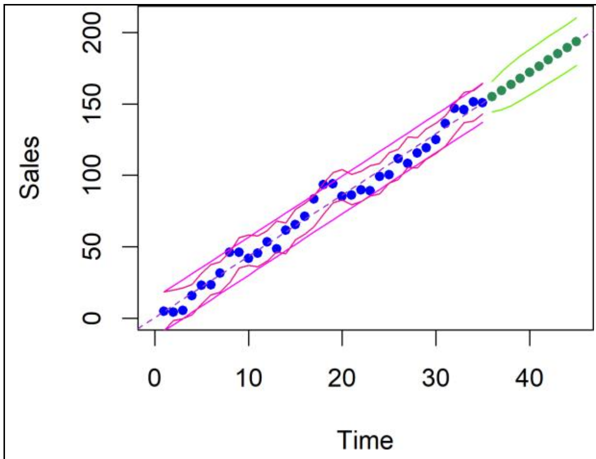
```
# scatter plot with uncorrelated error line
plot(t,y,pch=19,col="blue",xlab='Time',ylab='Sales',
xlim=c(0,n))
abline(lm(y~t),r,lty=2,col='red')
lines(t,yasthat, type = "l", col = "green")

#view model fit ANOVA table
temp<-anova(mymodel2)
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

# view model s, Rsq and adjRsq
print('s,R-squared,adj R-squared')
c(summary(mymodel2)$s,summary(mymodel2)$r.squared,
 summary(mymodel2)$adj.r.squared)
```

MATH 2780 Chapter 10A Worksheet

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



t	y	yhatUcor	FL	FU
1	4.8	4.697	-8.987	18.381
2	4	8.993	-4.634	22.619
3	5.5	13.288	-0.284	26.861
4	15.6	17.584	4.062	31.106
5	23.1	21.880	8.405	35.355
6	23.3	26.175	12.744	39.606
7	31.4	30.471	17.080	43.861
8	46	34.767	21.413	48.120
9	46.1	39.062	25.743	52.382
10	41.9	43.358	30.068	56.647
11	45.5	47.653	34.391	60.916
12	53.5	51.949	38.709	65.189
13	48.4	56.245	43.025	69.465
14	61.6	60.540	47.337	73.744
15	65.6	64.836	51.645	78.027
16	71.4	69.132	55.949	82.314
17	83.4	73.427	60.250	86.604
18	93.6	77.723	64.548	90.898
19	94.2	82.018	68.842	95.195
20	85.4	86.314	73.132	99.496
21	86.2	90.610	77.418	103.801
22	89.9	94.905	81.702	108.109
23	89.2	99.201	85.981	112.421
24	99.1	103.497	90.257	116.736
25	100	107.792	94.530	121.055
26	112	112.088	98.798	125.377
27	108	116.384	103.064	129.703
28	116	120.679	107.326	134.033
29	119	124.975	111.584	138.365
30	125	129.270	115.839	142.701
31	136	133.566	120.091	147.041
32	147	137.862	124.339	151.384
33	146	142.157	128.585	155.730
34	151	146.453	132.826	160.080
35	151	150.749	137.065	164.432
36	*	155.139	144.484	165.794
37	*	159.402	145.968	172.836
38	*	163.678	148.848	178.508
39	*	167.963	152.368	183.558
40	*	172.252	156.223	188.280
41	*	176.543	160.264	192.823
42	*	180.837	164.412	197.262
43	*	185.132	168.621	201.642
44	*	189.427	172.866	205.987
45	*	193.722	177.132	210.312

```
# SALES35 #
nf   <-10
alph <-0.05
k    <-1
tcrit<-qt(1-alph/2,n-k-1)

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

# parse out variables
n <- nrow(mydata)
t <- c(mydata[,1]) #T time
y <- c(mydata[,2]) #y sales
e <- c(mydata[,3]) #e ind model residual

## Uncorrelated model
# fit linear independent model
mymodel <- lm(y~t)
mymodel$coefficients
ee<-mymodel$residuals
cbind(e,ee) # e=ee
yhatUcor<-mymodel$fitted.values
yhatUcor

# Uncorrelated PI
yPIL <- matrix(rep(0,n),nrow=n,ncol=1)
yPIU <- matrix(rep(0,n),nrow=n,ncol=1)
sUcor<- summary(mymodel)$sigma

xbar<-mean(t)
SSxx<-sum(t^2)-(sum(t)^2/n)
for (i in 1:n){
  yPIL[i]<-yhatUcor[i]-tcrit*sUcor*
sqrt(1+1/n+(t[i]-xbar)^2/SSxx)
  yPIU[i]<-yhatUor[i]
+tcrit*sUcor*sqrt(1+1/n+(t[i]-xbar)^2/SSxx)
  }
withinPI<-cbind(y,yhatUcor,yPIL,yPIU)
withinPI

## Correlated model
# fit correlated line model and get statistics
phihat<-cor(ee[1:(n-1)],ee[(1+1):n])
phihat
yast<-y[2:n]-phihat*y[1:n-1]
tast<-t[2:n]-phihat*t[1:n-1]
mymodel2 <- lm(yast~tast)

# get correlated error regression coefficients
b0ast<-mymodel2$coefficients[1]/(1-phihat)
b1ast<-mymodel2$coefficients[2]
bast<-c(b0ast,b1ast)
bast
```

```
# values entered from book, w/ slight round off
b0    <-0.4058
b1    <-4.2959
phihat<-0.5896
MSE   <-27.42767

Rt     <- matrix(rep(0,n),nrow=n,ncol=1)
for (i in 1:n){
  Rt[i]<-y[i]-b0-b1*i
}
yhatcor<- matrix(rep(0,n),nrow=n,ncol=1)
yhatcor[1]<-yhatUcor[1]# set first value to ucor
yhatcorL<- matrix(rep(0,n),nrow=n,ncol=1)
yhatcorU<- matrix(rep(0,n),nrow=n,ncol=1)
yhatcorL[1]<-yPIL[1] # set first value to ucor
yhatcorU[1]<-yPIU[1] # set first value to ucor
for (i in 2:n){
  yhatcor[i]<-b0+b1*t[i]+phihat*Rt[i-1]
  yhatcorL[i]<-yhatcor[i]-tcrit*sqrt(MSE)
  yhatcorU[i]<-yhatcor[i]+tcrit*sqrt(MSE)
}
cbind(y,Rt,yhatcor,yhatcorL,yhatcorU)

# forecast mean and interval
F  <- matrix(rep(0,nf),nrow=nf,ncol=1)
FL <- matrix(rep(0,nf),nrow=nf,ncol=1)
FU <- matrix(rep(0,nf),nrow=nf,ncol=1)
Rn <-y[n]-b0-b1*n
TMP    <-0
for (i in 1:nf){
  F[i] <-b0+b1*(n+i)+phihat^i*Rn
  TMP  <-TMP+phihat^(i-1)
  FL[i]<-F[i]-tcrit*sqrt(MSE*TMP)
  FU[i]<-F[i]+tcrit*sqrt(MSE*TMP)
}
Fall<-cbind(rep(0,nf),F,FL,FU)
Fall

# scatter plot with uncorr and corr error lines
plot(t,y,pch=19,col="blue",xlab='Time',ylab='Sales',
    cex=0.75,xlim=c(0,n+nf),ylim=c(0,max(FU)))
abline(mymodel,lty=2,col='red')      # ind fit
lines(t,yPIL, type = "l", col = "magenta")
lines(t,yPIU, type = "l", col = "magenta")
abline(b0,b1,lty=2,col='darkorchid') # cor fit
points(seq(1,n,by=1),yhatcorL,col='deeppink2',type="l")
points(seq(1,n,by=1),yhatcorU,col='deeppink2',type="l")
points(seq(n+1,n+nf,by=1),F,pch=19,cex=0.75,col='seagreen')
points(seq(n+1,n+nf,by=1),FL,col='chartreuse2',type="l")
points(seq(n+1,n+nf,by=1),FU,col='chartreuse2',type="l")

results<-rbind(withinPI,Fall)
```