

Homework-3 - Stats 519

Aji John

Problem 8

Problem 8 (8 points). Perform an analysis of the atmospheric carbon dioxide (CO₂) time series from Mauna Loa, Hawaii, along the same lines as the analysis of the accidental deaths (AD) series starting with lecture overhead III-82 (the 3rd set of R code on the course Web site has the code used to analyze the AD series). You can read the CO₂ data directly into R using `co2 <- read.table("http://faculty.washington.edu/dbp/s519/Data/co2-1958-2017.txt")`.

Please feel free to alter choices that were made in the analysis of the AD series if you deem them to be inappropriate for your analysis of the CO₂ series. Create and turn in plots that correspond to overheads III-85, III-87, III-89 (but just show the seasonal pattern for a single year, as is done in III-107), III-92, III-94, III-96, III-97 and III-98, along with brief descriptions of the steps you took in your analysis (please also turn in the code you used to do your analysis). Finally, state briefly your conclusions about how well the simple modeling approach worked for the CO₂ series.

```
co2 <- read.table("http://faculty.washington.edu/dbp/s519/Data/co2-1958-2017.txt")
str(co2)

## 'data.frame':    708 obs. of  2 variables:
## $ V1: num  1958 1958 1958 1958 1958 1959 ...
## $ V2: num  316 317 318 317 316 ...
### define function to do filtering ...

filter.with.padding <- function(x,the.filter,iter=1)
{
  q <- (length(the.filter)-1)/2
  n <- length(x)
  w <- stats::filter(c(rep(x[1],q),x,rep(x[n],q)),the.filter)[(q+1):(q+n)]
  if(iter > 1) for(i in 2:iter) w <- filter(c(rep(w[1],q),w,rep(w[n],q)),the.filter)[(q+1):(q+n)]
  return(w)
}

plot.ACFest <- function(ts, main=NULL, n.lags=40)
{
  ts.acf <- acf(ts, lag.max=n.lags, plot=FALSE)
  n.ts <- length(ts)
  xs <- 1:n.lags
  ys <- ts.acf$acf[2:(n.lags+1)]
  plot(xs,ys,typ="h",xlab="h (lag)",ylab="ACF",ylim=c(-1,1),col="blue",main=main)
  points(xs,ys,col="red",cex=0.5)
  xs <- 1:n.lags
  xs[1] <- xs[1] - 0.25
  xs[n.lags] <- xs[n.lags] + 0.25
  lines(xs,1.96*sqrt(n.ts-xs)/n.ts,col="magenta",lty="dashed")
  lines(xs,-1.96*sqrt(n.ts-xs)/n.ts,col="magenta",lty="dashed")
}
```

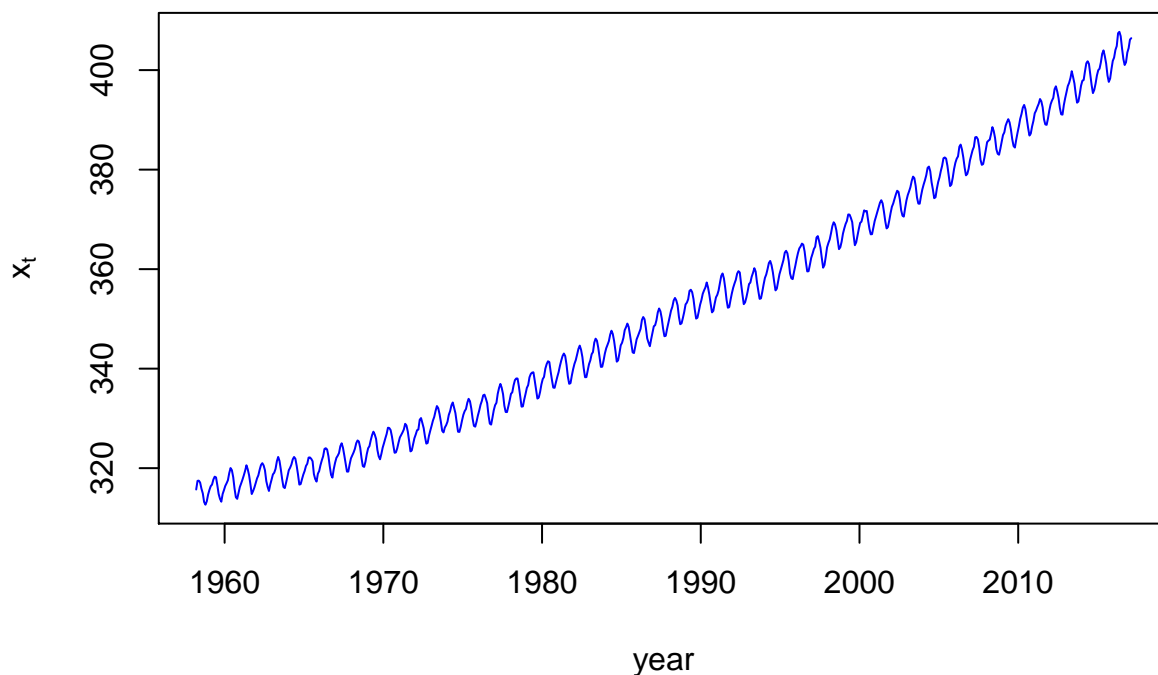
```

abline(h=0,lty="dashed")
CI.hw <- 1.96/sqrt(n.ts)
lines(c(0.75,n.lags+0.25),rep(CI.hw,2),col="blue",lty="dashed")
lines(c(0.75,n.lags+0.25),rep(-CI.hw,2),col="blue",lty="dashed")
return(ts.acf$acf)
}
### overhead III-2

plot(co2$V1,co2$V2,col="blue",xlab="year",typ="l",
      ylab=expression(x[t]),main=expression(paste("2nd Example: ", CO[2],
                                                    " Series from Mauna Loa, Hawaii"))))

```

2nd Example: CO₂ Series from Mauna Loa, Hawaii



III-85 - Seasonal component taken out

Use a smoothing filter to take the seasonal component out

Take 5 points before the center, and 5 after, and penalize the extremes. ### III-85

```

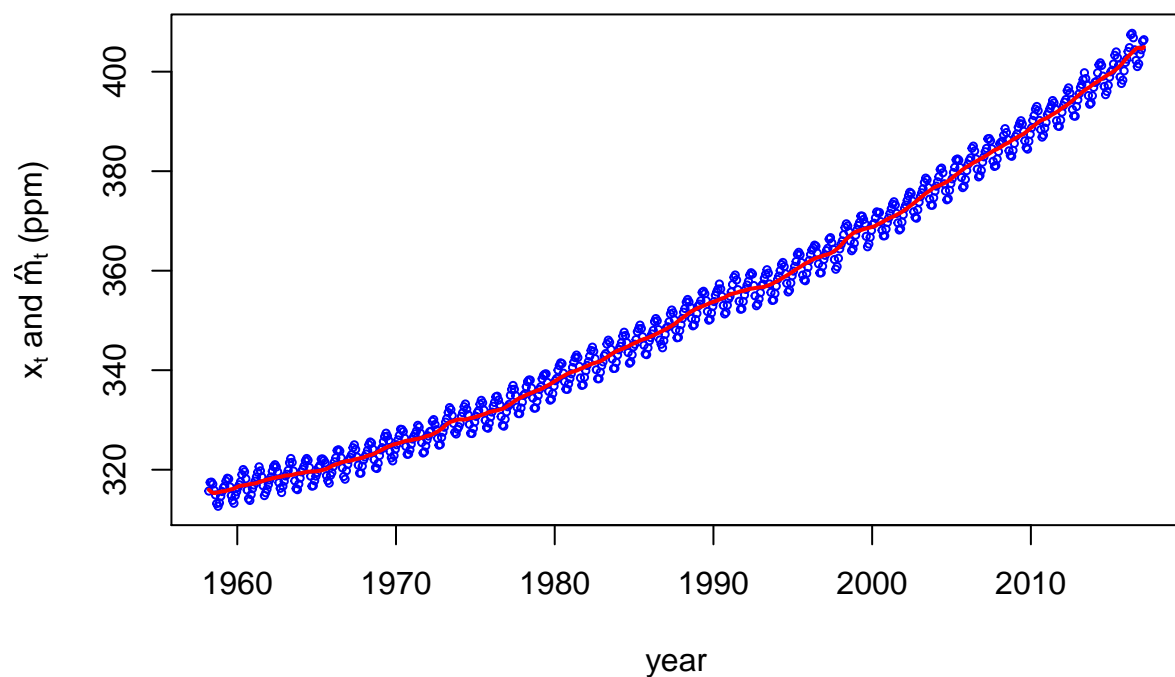
co2 <- read.table("http://faculty.washington.edu/dbp/s519/Data/co2-1958-2017.txt")

m.hat.co2 <- filter.with.padding(co2$V2,c(1/24,rep(1/12,11),1/24))

plot(co2$V1,co2$V2,col="blue",xlab="year",typ="b",
      ylab=expression(paste(x[t], " and ", hat(m)[t], " (ppm)")),main="Monthly CO2 Values",cex=0.5)
lines(co2$V1,m.hat.co2,col="red",lwd=2)

```

Monthly CO2 Values



###

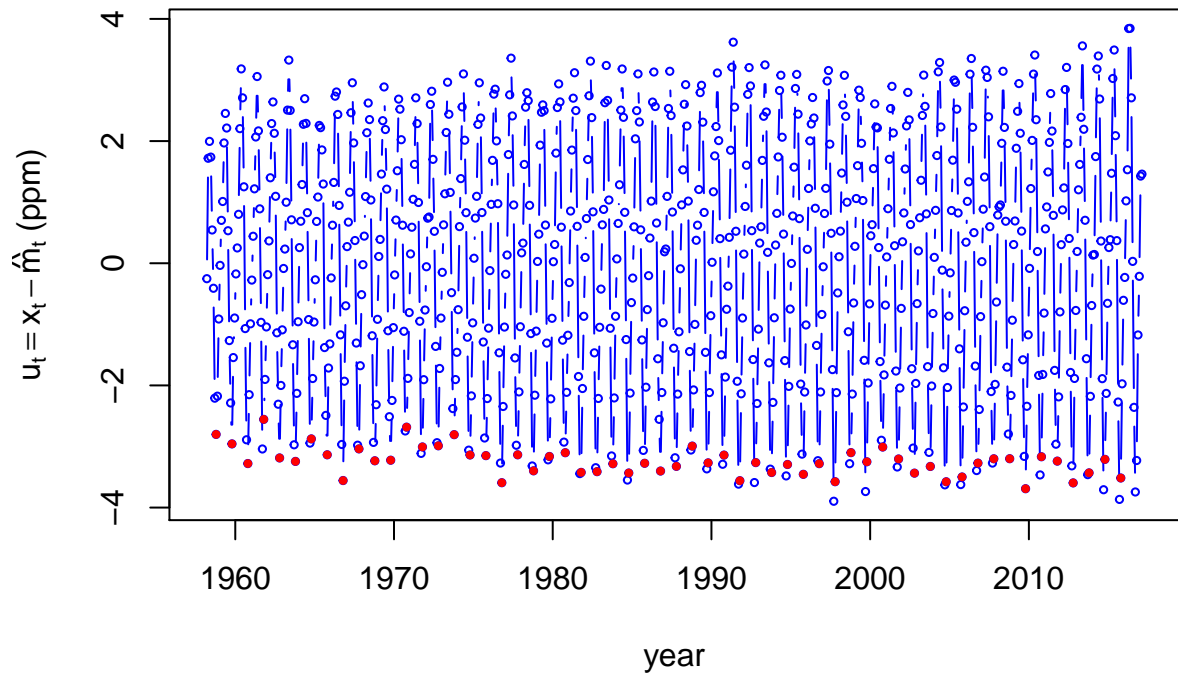
III-87

Removing the trend component and plotting.

```
co2.u <- co2$V2 - m.hat.co2
```

```
plot(co2$V1,co2.u,col="blue",xlab="year",typ="b",ylab=expression(paste(u[t]==x[t]-hat(m)[t]," (ppm)")),
points(co2$V1[seq(8,696,12)],co2.u[seq(8,696,12)],pch=16,col="red",cex=0.6)
```

Preliminary Detrending of CO2 Series



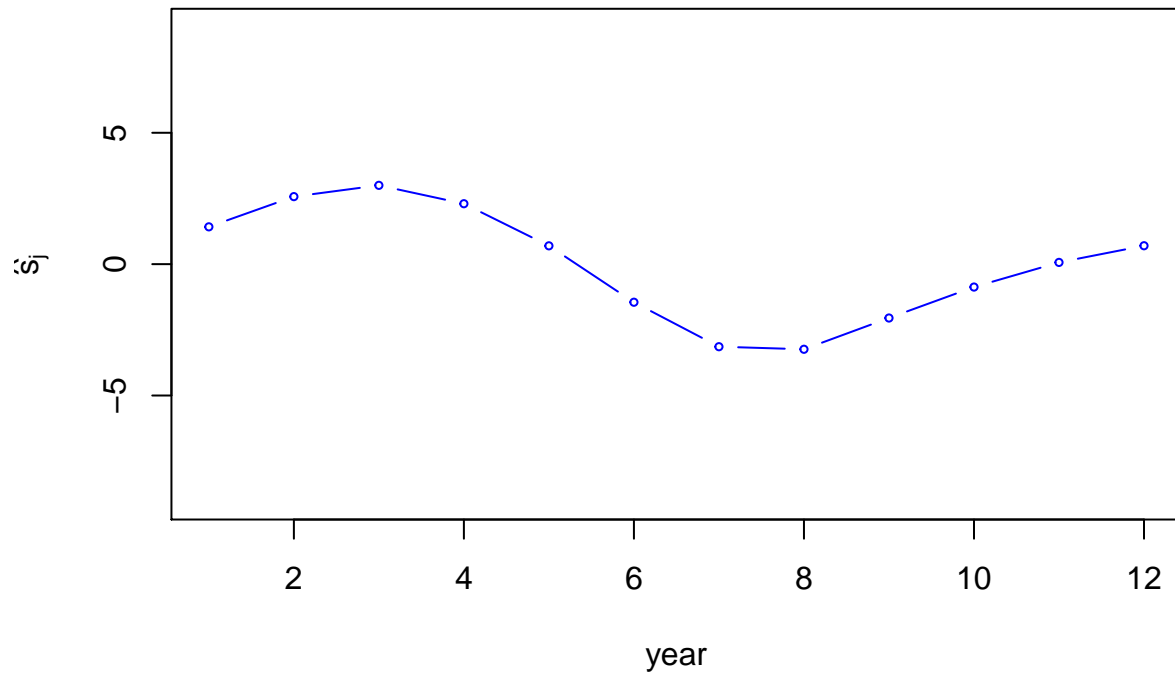
III-89

Extracting for one year to show the seasonal pattern.

```
co2.w.j <- rowMeans(matrix(co2.u,nrow=12))

plot(1:12,co2.w.j- mean(co2.w.j),col="blue",xlab="year",
     typ="b",ylab=expression(hat(s)[j]),
     main=expression(paste("CO2 Step 3: Form Estimate {" ,hat(s)[j] ,"} of Seasonal Pattern")),
     ylim=c(-9,9),cex=0.5)
```

CO2 Step 3: Form Estimate $\{\hat{s}_j\}$ of Seasonal Pattern



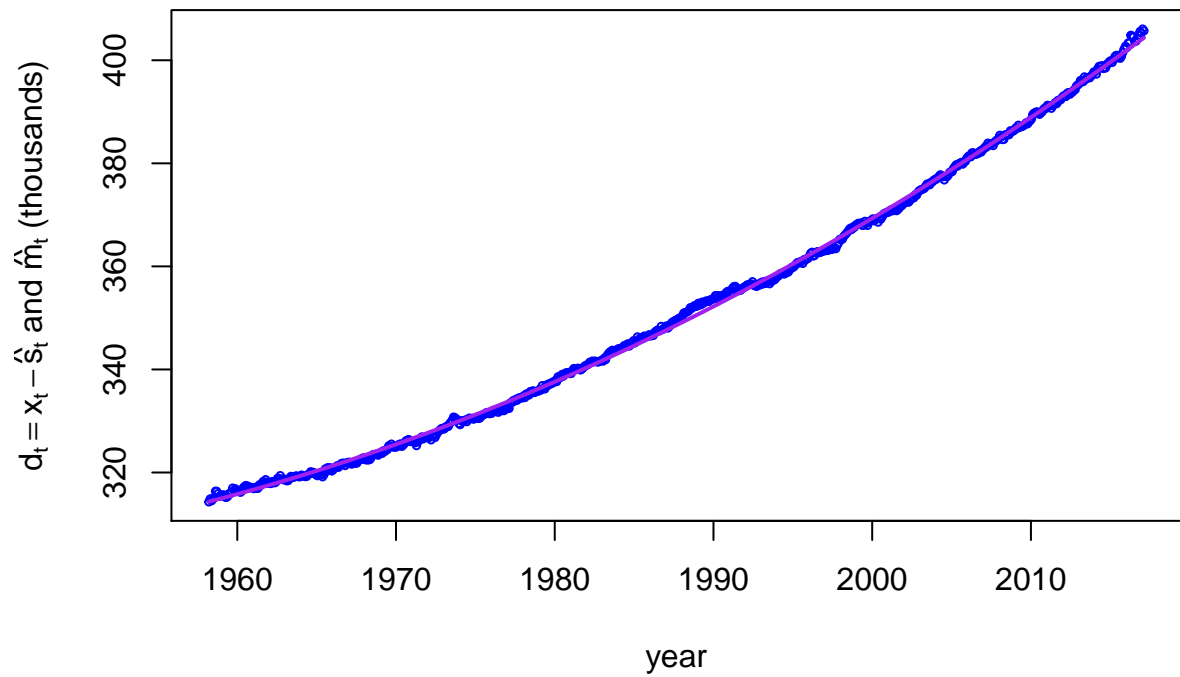
III-92 (Deasonalized data with trend estimate)

```
#get all the years(columns), rows(minths)
#average all the jan, feb... dec
co2.w.j <- rowMeans(matrix(co2.u,nrow=12))
co2.s.j.hat <- rep(co2.w.j - mean(co2.w.j),59)
co2.d <- co2$V2 - co2.s.j.hat

co2.d.reg <- lm(co2.d ~ co2$V1 + I(co2$V1^2))

plot(co2$V1,co2.d,col="blue",xlab="year",typ="b",
      ylab=expression(paste(d[t] == x[t] - hat(s)[t]," and ", hat(m)[t]," (thousands)")),
      main=expression(paste("Deseasonalized Data {" , d[t], "} and Trend Estimate {" ,hat(m)[t],"}")),
      cex=0.5)
lines(co2$V1,fitted(co2.d.reg),col="purple",lwd=2)
```

Deseasonalized Data $\{d_t\}$ and Trend Estimate $\{\hat{m}_t\}$

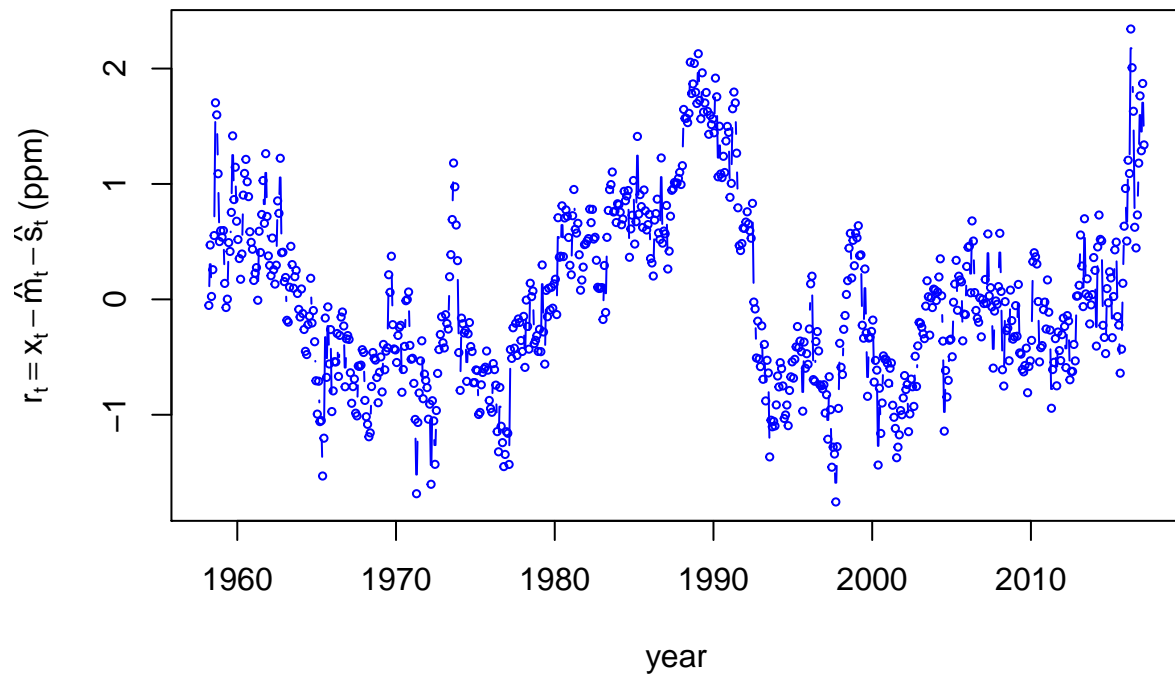


###

III-94 Residuals removed

```
plot(co2$V1, resid(co2.d.reg), col="blue", xlab="year", typ="b",
      ylab=expression(paste(r[t] == x[t] - hat(m)[t] - hat(s)[t], " (ppm)")),
      main=expression(paste("Residuals {" , r[t], "} from Removal of {" , hat(m)[t], "} and {" , hat(s)[t], "}")),
      cex=0.5)
```

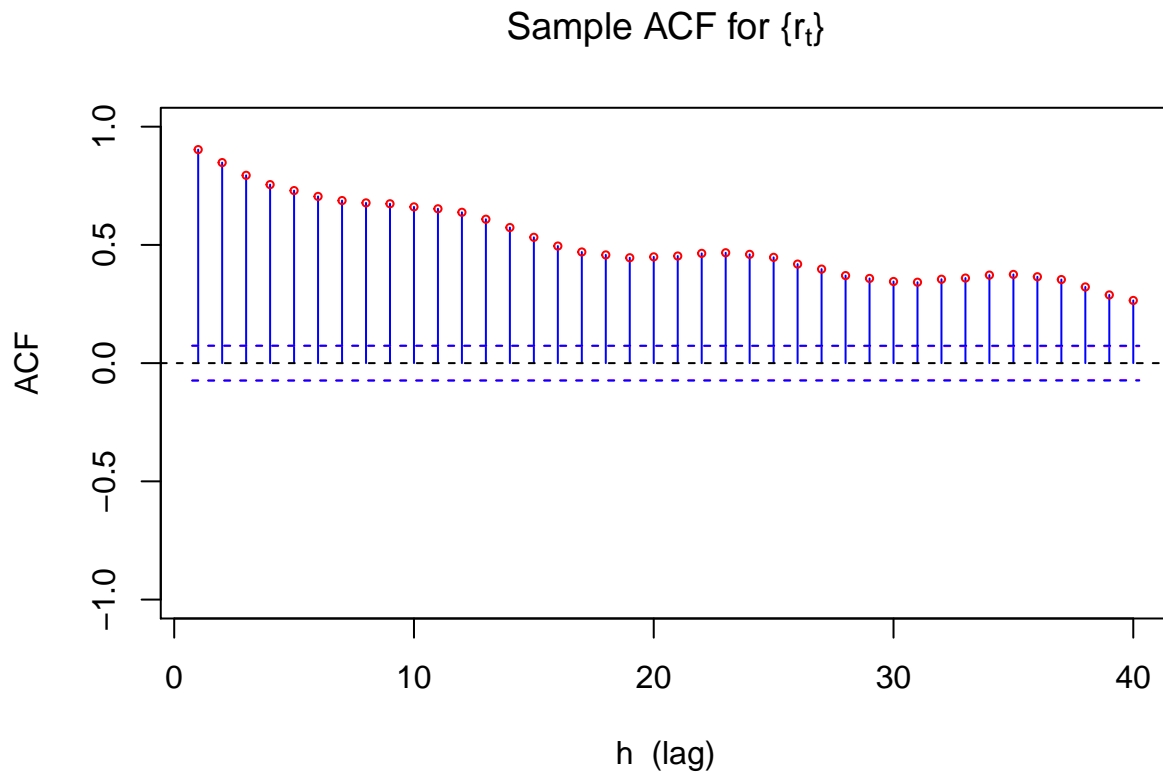
Residuals $\{r_t\}$ from Removal of $\{\hat{m}_t\}$ and $\{\hat{s}_t\}$



III-96

** Fails null hypothesis ** Plotting the residuals and showing the 95% CIs

```
phi <- plot.ACFest(resid(co2.d.reg),
  expression(paste("Sample ACF for {", r[t], "}")))[2]
```

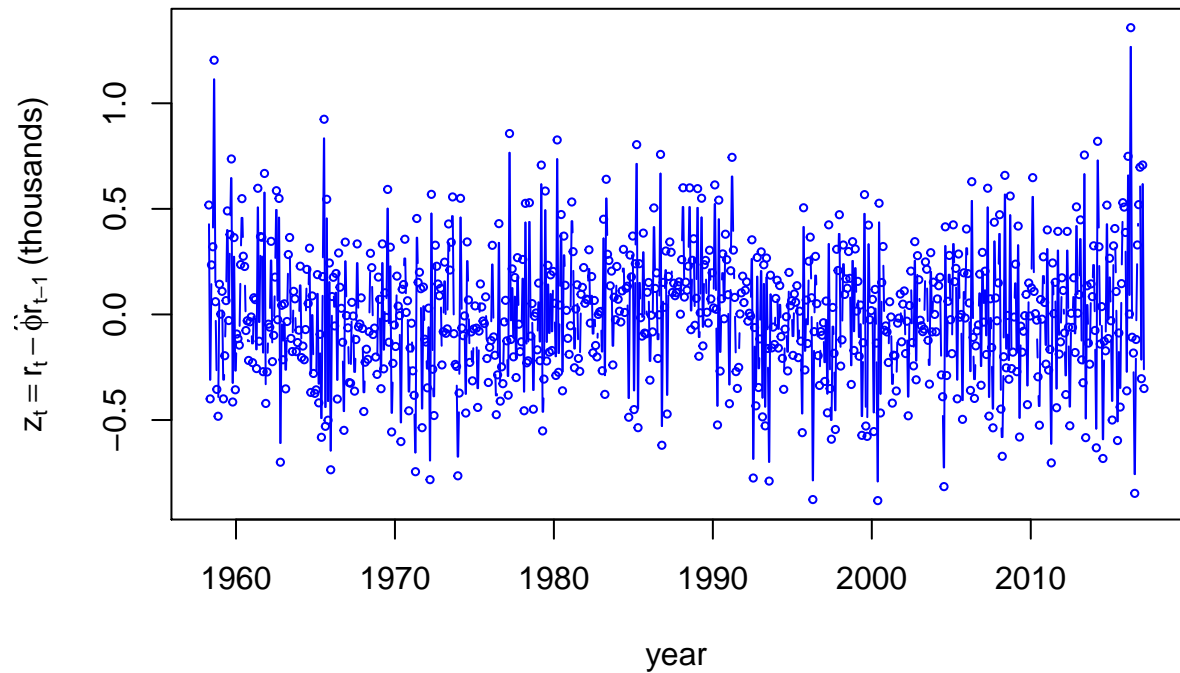


III-97

Testing whether the resultant is a AR(1) model.

```
co2.z <- resid(co2.d.reg)[-1] - phi*resid(co2.d.reg)[-length(resid(co2.d.reg))]  
  
plot(co2$V1[-1],co2.z ,col="blue",xlab="year",  
      typ="b",ylab=expression(paste(z[t]==r[t]-hat(phi)*r[t-1]," (thousands)")),  
      main=expression(paste("Residuals ",  
                             z[t]==r[t]-hat(phi)*r[t-1]," from Fitted AR(1) Model")),  
      cex=0.5)
```


Residuals $z_t = r_t - \hat{\phi}r_{t-1}$ from Fitted AR(1) Model



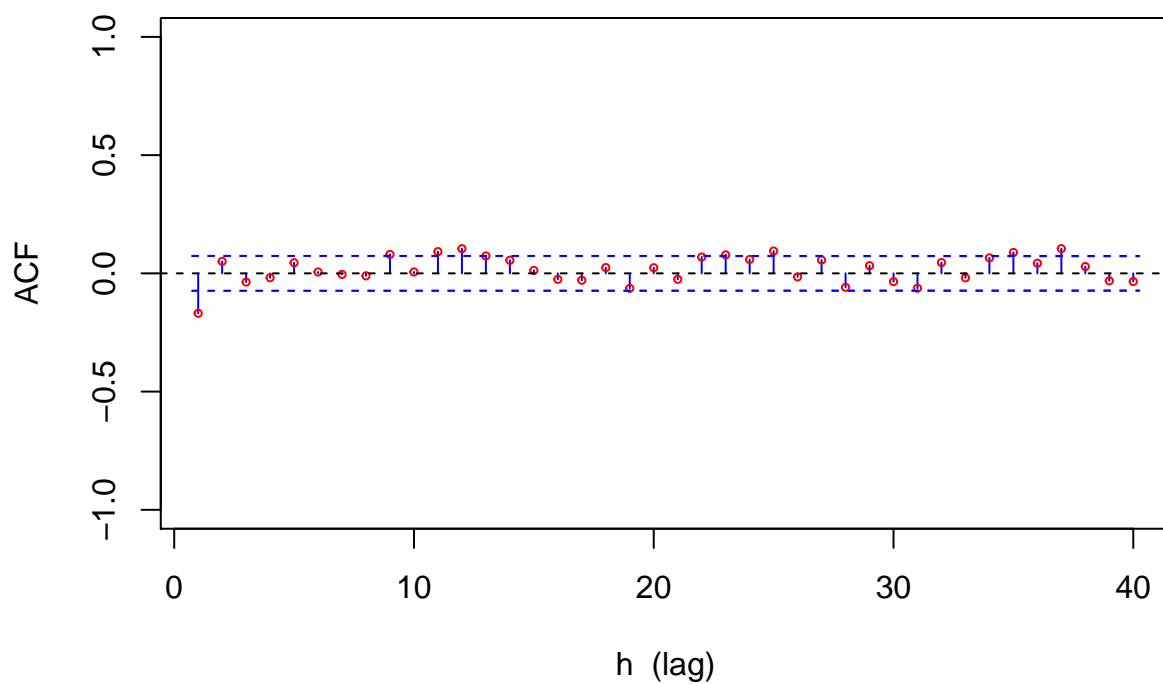
###

III-98

ACF for residuals from the fitted AR(1) model, very good, but two exceptions (lag 1 more pronounced)

```
plot.ACFest(co2.z, expression(paste("Sample ACF for {", z[t], "}")))[2]
```

Sample ACF for $\{z_t\}$



```
## [1] -0.1687572
```

Problem 9

Problem 9 (8 points). Here we consider a time series $\{x_t\}$ measuring ambient noise in the ocean from one second to the next. You can read this series directly into R using `on.ts <- scan("http://faculty.washington.edu/dbp/s519/Data/ocean-noise.txt")`.

Alternatively, you can access the data either via a link on item 15 in the list on the Data page of the course Web site or by going directly to <http://faculty.washington.edu/dbp/s519/Data/ocean-noise.txt>

Create plots of (a) the time series $\{x_t\}$, (b) its unit-lag scatter plot (i.e., x_{t+1} versus x_t) and (c) its sample ACF out to lag 20, along with lines showing 95% confidence intervals for the ACF under the null hypothesis that $\{x_t\}$ is a realization of an IID noise process. Examine the null hypothesis by subjecting $\{x_t\}$ to the portmanteau, turning point, difference-sign, rank and runs tests. State the results of each test and your overall conclusion about the viability of the IID noise hypothesis. How do these formal tests compare with the informal test given in your plot of the sample ACF? (The 4th set of R code on the course Web site has code used to compute the five tests for the examples considered in the course overheads.)

(a) Ocean noise time series

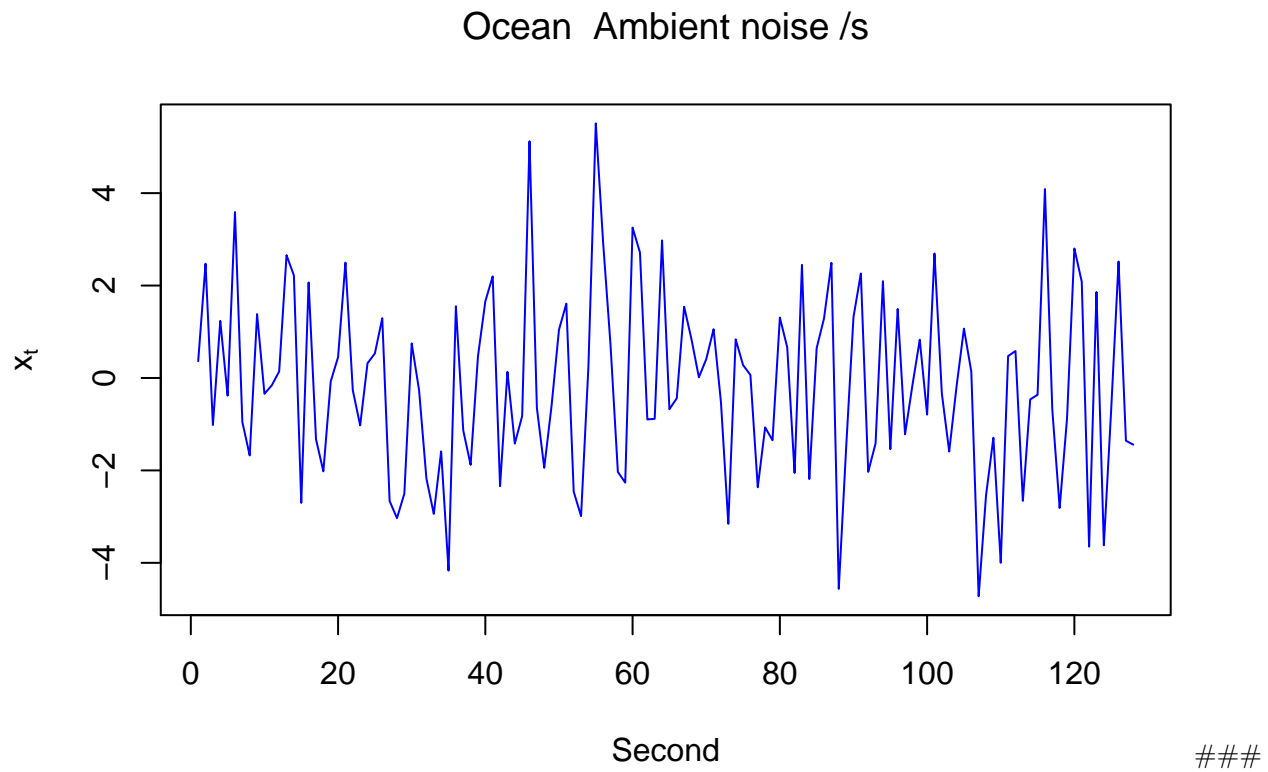
```
onnoise <- read.table("http://faculty.washington.edu/dbp/s519/Data/ocean-noise.txt")

onnoise$t <- 1:128
str(onnoise)

## 'data.frame':   128 obs. of  2 variables:
## $ V1: num  0.362 2.474 -1.019 1.236 -0.383 ...
## $ t : int   1 2 3 4 5 6 7 8 9 10 ...
### define function to do filtering ...

plot(onnoise$t,onnoise$V1,col="blue",xlab="Second",typ="l",
      ylab=expression(x[t]),main=expression(paste("Ocean ",
                                                    " Ambient noise /s")))

```



(b) Ocean noise unit-lag scatter

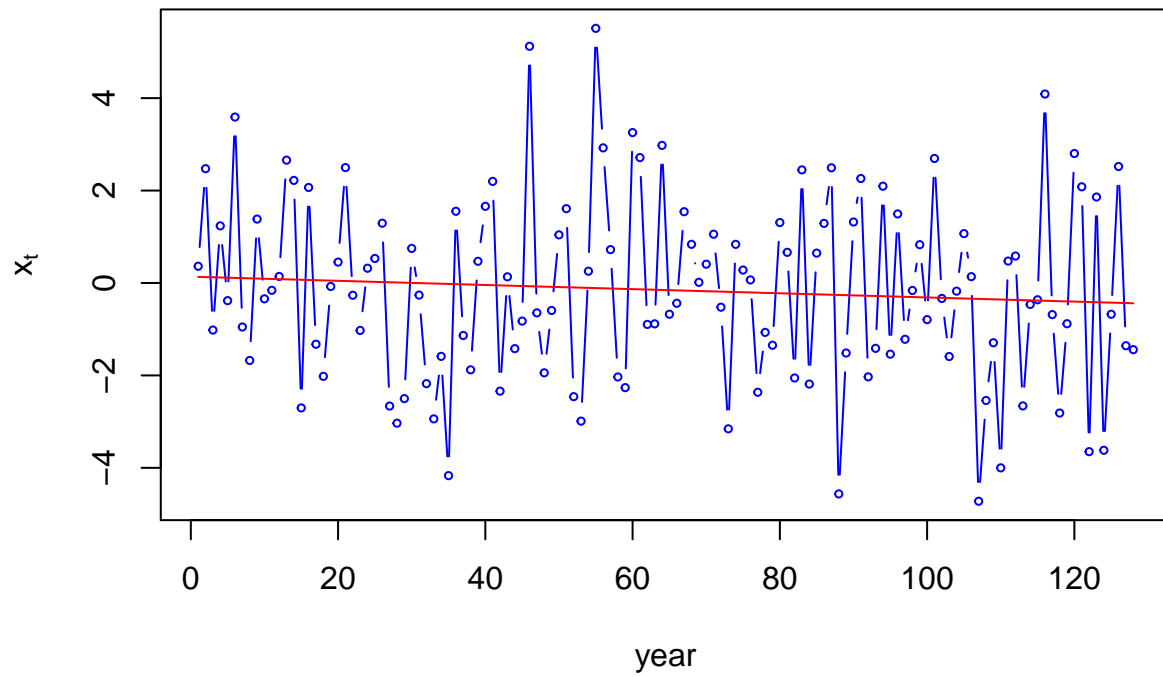
Part i

```
onoise.reg <- lm(onoise$V1 ~ onoise$t)

N.onoise <- length(onoise$t)

plot(onoise$t,onoise$V1 ,col="blue",xlab="year",
      typ="b",ylab=expression(x[t]),main="Ocean floor ambient noise",cex=0.5)
lines(onoise$t,fitted(onoise.reg),col="red")
```

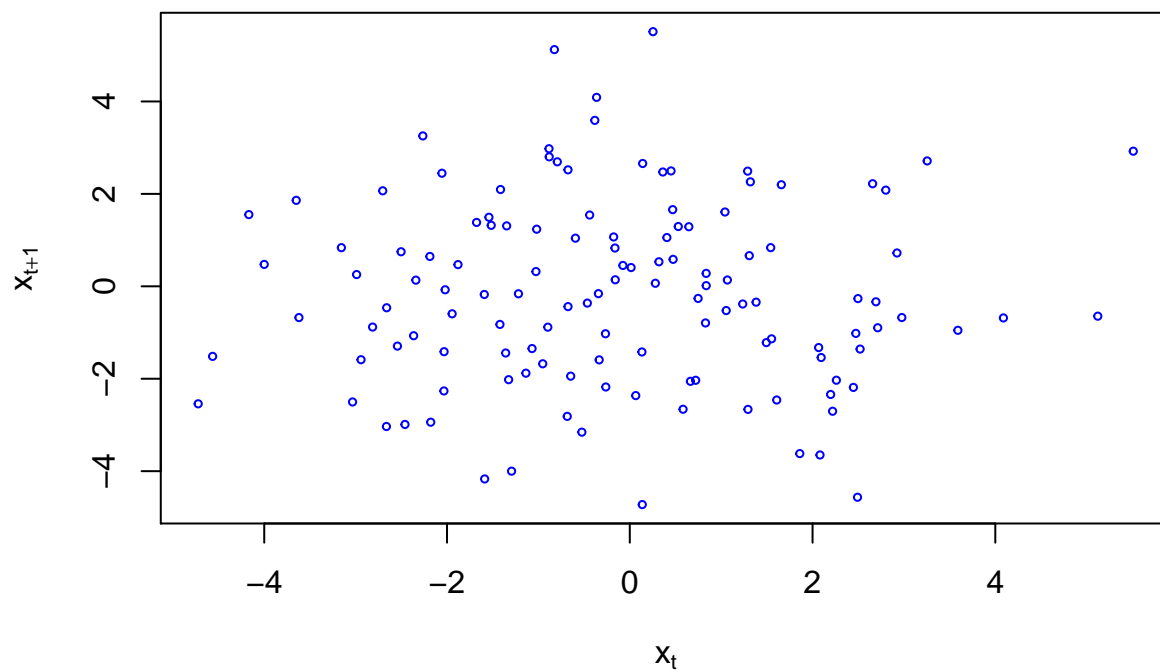
Ocean floor ambient noise



Part ii

```
plot(onoise$V1[-N.onoise],onoise$V1[-1],col="blue",
     xlab=expression(x[t]),typ="p",
     ylab=expression(x[t+1]),
     main=expression(paste("Unit Lag Scatter Plot of Ocean noise {", x[t],"}")),cex=0.5)
```

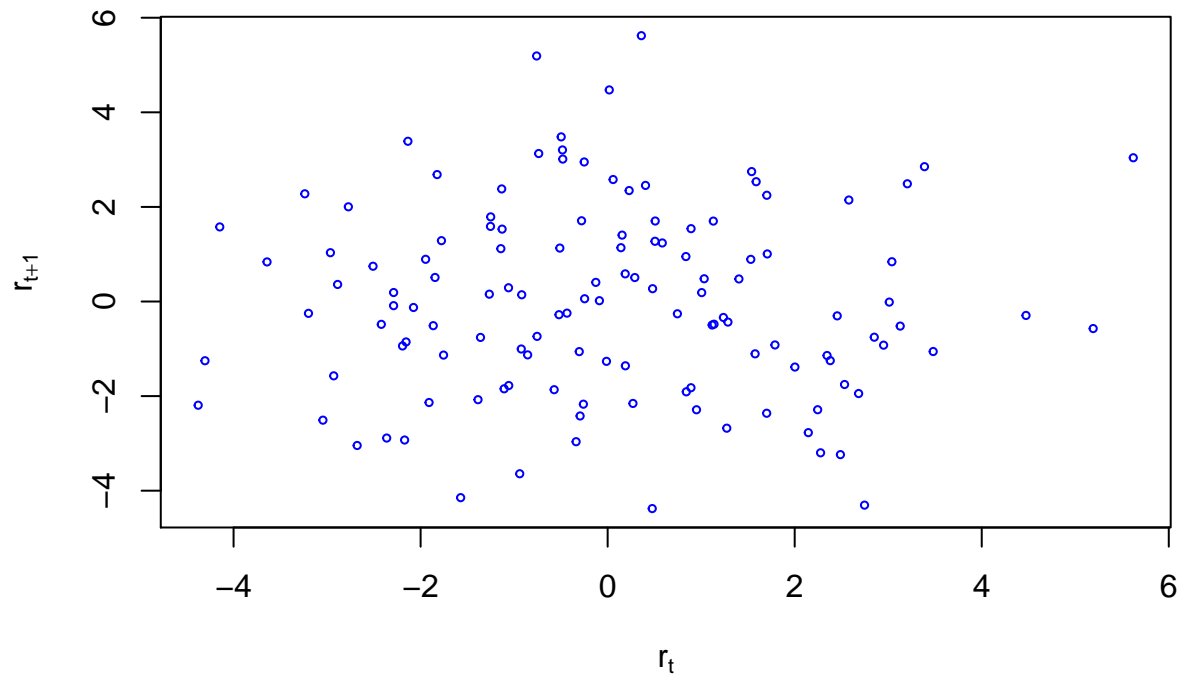
Unit Lag Scatter Plot of Ocean noise $\{x_t\}$



Alternate way to look at it by using residuals

```
plot(resid(onoise.reg)[-N.onoise],resid(onoise.reg)[-1],
     col="blue",xlab=expression(r[t]),typ="p",
     ylab=expression(r[t+1]),
     main=expression(paste("Unit Lag Scatter Plot of Residuals {", r[t],"}")),cex=0.5)
```

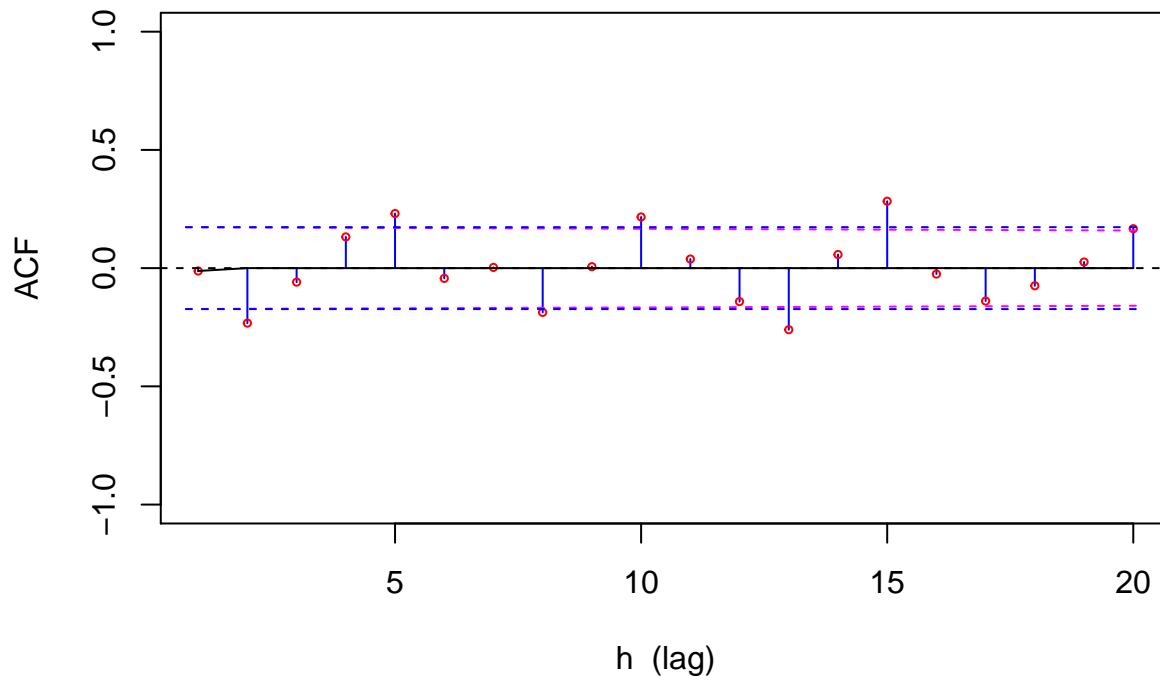
Unit Lag Scatter Plot of Residuals $\{r_t\}$



(c) its sample ACF out to lag 20, along with lines showing 95% confidence intervals for the ACF under the null hypothesis that $\{x_t\}$ is a realization of an IID noise process

```
onoise.reg.acf <- plot.ACFest(resid(onoise.reg),
                             expression(paste("Sample ACF for Residuals {", r[t], "}")), 20)
hs <- 1:20
lines(hs, onoise.reg.acf[2]^hs)
```

Sample ACF for Residuals $\{r_t\}$



Few sticklers ar lag 15, 10 ,5,, so would reject the null hypothesis by looking at the ACF plot.

Lets continue with portmanteau test.

```
source("http://faculty.washington.edu/dbp/s519/R-code/diagnostic-tests.R")
plot.figPortmanteau <- function(Qs, quantas, p=0, main=NULL)
{
  N.Qs <- length(Qs)
  hs <- (p+1):(N.Qs+p)
  plot(hs,quantas,typ="n",xlab="h (lag)",ylab=expression(paste(Q[LB],"(h) and 95% quantile")),xlim=c(
  points(hs,Qs,pch=16,col="blue")
  segments(seq(p+0.75,by=1,length=N.Qs),quantas,seq(p+1.25,by=1,
    length=N.Qs),quantas,lwd=2,col="red")
}

#Housekeeping
onoise.reg.acf2 <- acf(resid(onoise.reg), lag.max=420, plot=FALSE)$acf
round(phi <- onoise.reg.acf2[2],3) # -0.012

## [1] -0.012
onoise.zs <- resid(onoise.reg)[-1] - phi*resid(onoise.reg)[-N.onoise]

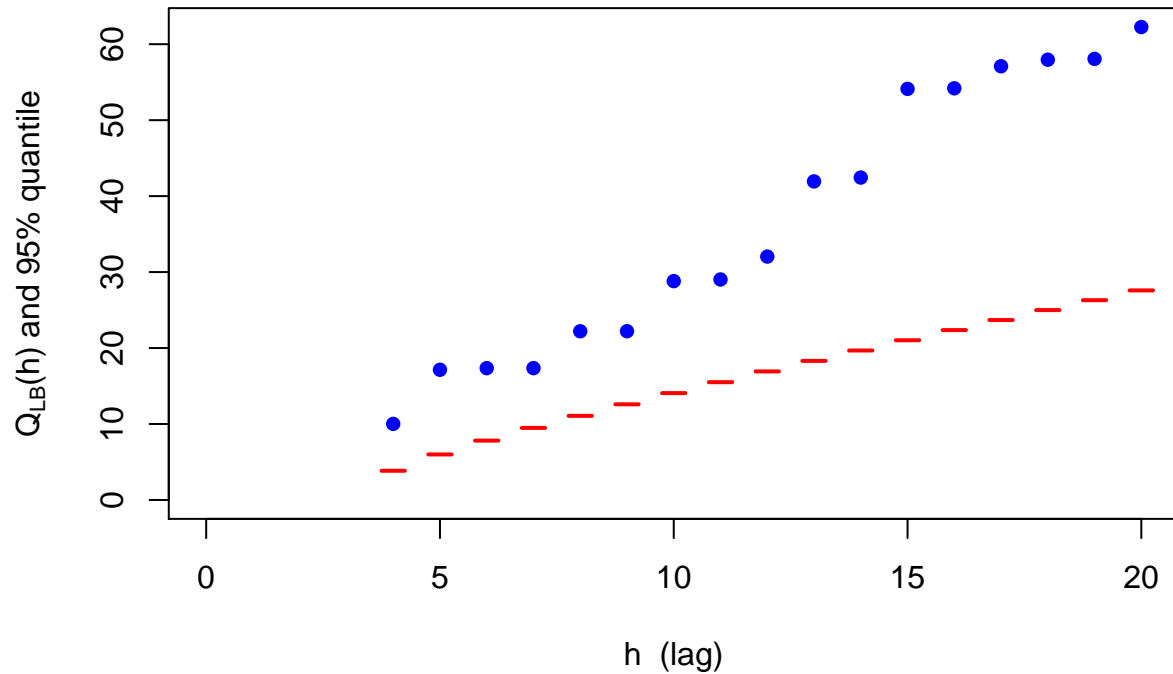
### overhead IV-13

onoise.acf.zs <- acf(onoise.zs, lag.max=20, plot=FALSE)$acf[-1]
Q.zs <- sapply(4:20,function(h) {portmanteau.test.LB(onoise.acf.zs[1:h],length(onoise.zs))})
quantas.4.20 <- sapply(4:20,function(dof) {qchisq(0.05,dof,lower=FALSE)})
quantas.4.20 <- sapply((4:20)-3,function(dof) {qchisq(0.05,dof,lower=FALSE)})

plot.figPortmanteau(Q.zs,quantas.4.20,3,
```

```
main=expression(paste("Portmanteau Tests of Ocean noise {" , z[t],"}"))
```

Portmanteau Tests of Ocean noise $\{z_t\}$



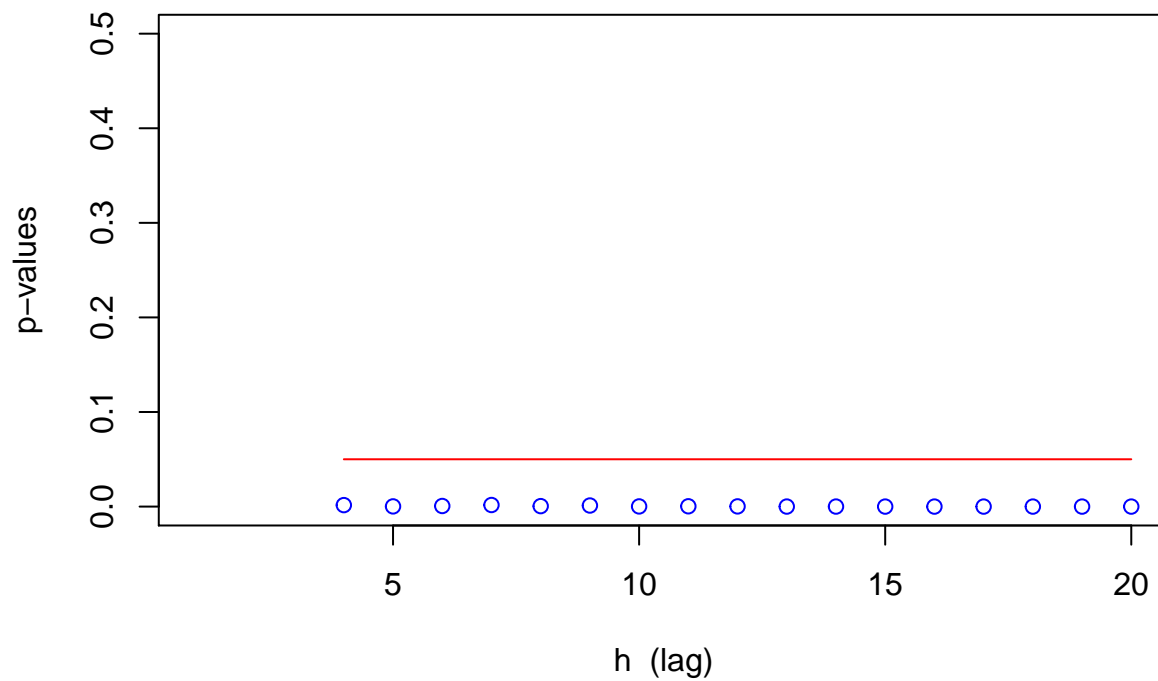
Dots are above the steps, so reject the null hypothesis

Lets look at p-values

```
### overhead IV-14
```

```
plot(4:20,1-apply(pchisq,Q.zs,(4:20)-3),typ="p",xlab="h (lag)",ylab="p-values",xlim=c(1,20),ylim=c(0,1),
      main=expression(paste("p-values for Portmanteau Tests of Ocean Noise {" , z[t],"}")))
lines(c(4,20),rep(0.05,2),col="red")
```


p-values for Portmanteau Tests of Ocean Noise $\{z_t\}$



Diff sign test

```
(results <- difference.sign.test(onoise.zs))
```

```
## $test.sum
## [1] 64
##
## $test
## [1] 0.3061862
##
## $p.value
## [1] 0.7594629
##
## $mu
## [1] 63
##
## $var
## [1] 10.66667
```

```
results$test.sum      # 64
```

```
## [1] 64
```

```
length(onoise.zs)     # 127
```

```
## [1] 127
```

```
results$mu            # 63
```

```
## [1] 63
```

```

round(results$var,1)          # 10.7

## [1] 10.7
round(sqrt(results$var),1)    # 3.3

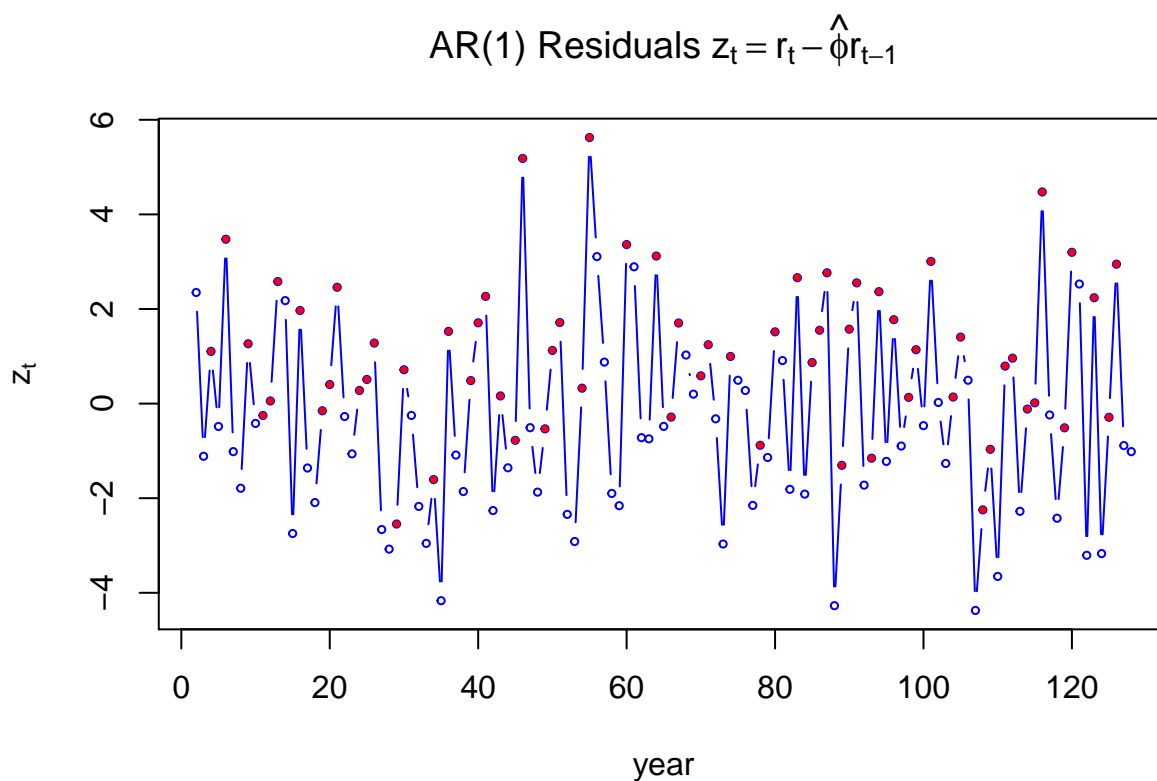
## [1] 3.3
results$test                  # 0.30

## [1] 0.3061862
results$p.value               # .7594629

## [1] 0.7594629
### overhead IV-21

plot(onoise$t[-1],onoise.zs,col="blue",xlab="year",typ="b",
     ylab=expression(z[t]),main=expression(paste("AR(1) Residuals ",
     z[t] == r[t]-hat(phi)*r[t-1])),cex=0.5)
gg <- diff(onoise.zs) > 0
points((onoise$t[c(-1,-2)])[gg],(onoise.zs[-1])[gg],pch=16,col="red",cex=0.5)

```



Diff test, p-value is .7594629 (at $\alpha = 0.05$), so we fail to reject the null hypothesis

Turning point test

```

(results <- turning.point.test(onoise.zs))

## $test.sum

```

```

## [1] 72
##
## $test
## [1] 2.402362
##
## $p.value
## [1] 0.01628958
##
## $mu
## [1] 83.33333
##
## $var
## [1] 22.25556
results$test.sum          # 72

## [1] 72
length(onoise.zs)         # 127

## [1] 127
round(results$mu,1)        # 83.3

## [1] 83.3
round(results$var,1)       # 22.3

## [1] 22.3
round(sqrt(results$var),1) # 4.7

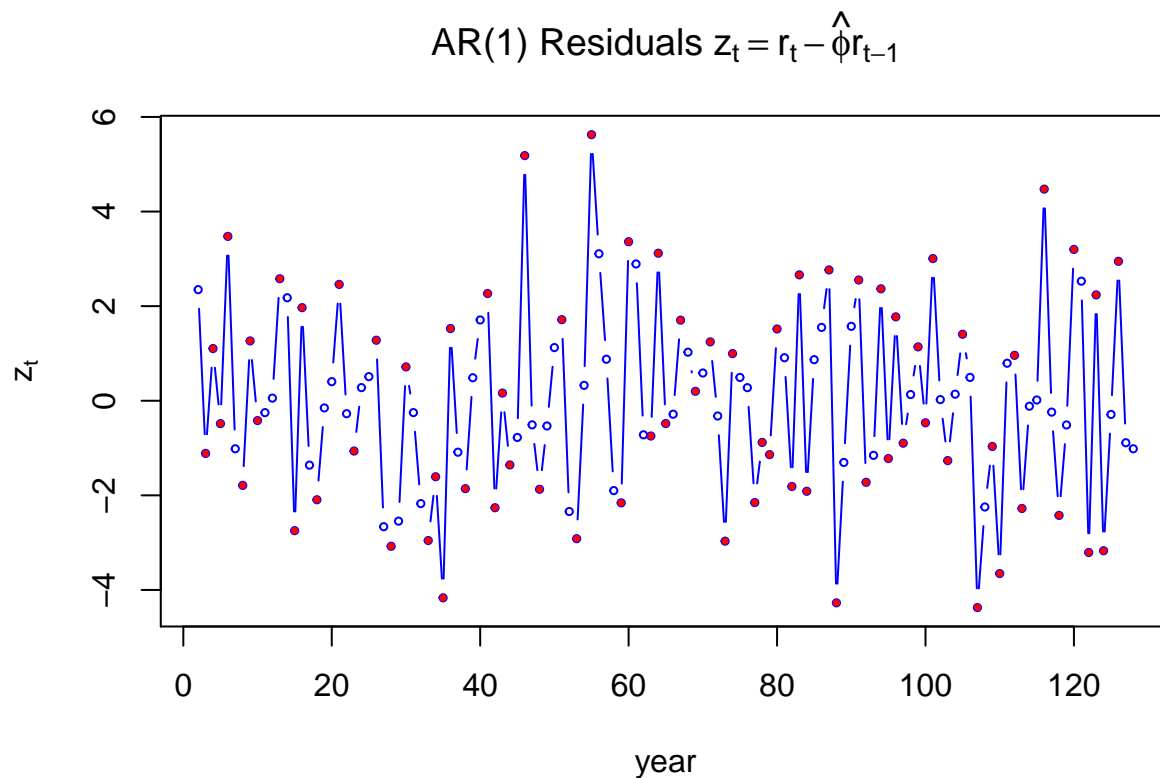
## [1] 4.7
round(results$test,2)      # 2.40

## [1] 2.4
round(results$p.value,2)   # 0.016

## [1] 0.02
### overhead IV-18

plot(onoise$t[-1],onoise.zs,col="blue",xlab="year",typ="b",
     ylab=expression(z[t]),main=expression(paste("AR(1) Residuals ",
     z[t] == r[t]-hat(phi)*r[t-1])),cex=0.5)
onoise.t.shortened <- onoise$t[-c(1,2,length(onoise.zs))]
y <- embed(onoise.zs,3)
gg <- (y[,2] > y[,1] & y[,2] > y[,3]) | (y[,2] < y[,1] & y[,2] < y[,3])
points(onoise.t.shortened[gg],(y[,2])[gg],pch=16,col="red",cex=0.5)

```



Number of turning points - 72. p-value = 0.016, so we reject the null hypothesis($\alpha = 0.05$).

Rank Test

```
(results <- rank.test(onoise.zs))
```

```
## $P
## [1] 4017
##
## $mu
## [1] 4000.5
##
## $var
## [1] 57562.75
##
## $sd
## [1] 239.9224
##
## $test.stat
## [1] 0.06877224
##
## $p.value
## [1] 0.9451709
##
## $reject.null
## [1] FALSE
```

```
results$P # 4017
```

```
## [1] 4017
length(onoise.zs)          # 127

## [1] 127
results$mu                 # 4000.5

## [1] 4000.5
round(results$var,1)       # 57562.75

## [1] 57562.8
round(results$sd,1)        # 239.9224

## [1] 239.9
round(results$test.stat,2) # 0.06877224

## [1] 0.07
round(results$p.value,2)   # 0.9451709
```

[1] 0.95
P-value - 0.94, so we fail to reject our null hypothesis.

Runs test

```
(results <- runs.test(onoise.zs))

## $n.runs
## [1] 66
##
## $n.pos
## [1] 62
##
## $n.neg
## [1] 65
##
## $mu
## [1] 64.46457
##
## $var
## [1] 31.46259
##
## $sd
## [1] 5.609153
##
## $test.stat
## [1] 0.2737371
##
## $p.value
## [1] 0.7842867
##
## $reject.null
## [1] FALSE
```

```

results$n.runs          # 66
## [1] 66
results$n.pos           # 62
## [1] 62
results$n.neg           # 65
## [1] 65
round(results$mu,1)      # 64.46457
## [1] 64.5
round(results$var,1)     # 31.46259
## [1] 31.5
round(results$sd,1)      # 5.609153
## [1] 5.6
round(results$test.stat,1) # 0.2737371
## [1] 0.3
round(results$p.value,3) # 0.7842867
## [1] 0.784

```

P-value - 0.78, so we fail to reject our null hypothesis.

Summary of IID Tests ($\{x_t\}$ is a realization of an IID noise process)

- Informal test based on sample ACF : IID hypothesis unlikely(reject null hypothesis), as handful are outside 95%CI
- portmanteau test: reject null (with $\alpha = 0.05$) for $h = 4$ thru 20
- turning point test: reject null (p-value is 0.016)
- difference-sign test: fail to reject (p-value is .75)
- rank test: fail to reject (p-value is 0.94)
- runs test: fail to reject (p-value is 0.78)

Conclusion

- There is a tie, if informal ACF test is included, but I feel the sample size is constrained to make a more informed decision. I would lean towards rejecting the hypothesis that it is IID noise based on Portmanteau & Turning point test .