Time Series Final

Steve Bramhall

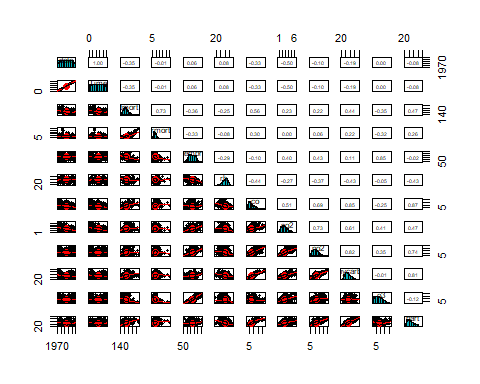
August 13, 2019

data(lap)  
mort=data.frame(date=time(lap),Time=as.factor(seq(1,508,1)),as.matrix(lap))  
  
library(psych)

##   
## Attaching package: 'psych'

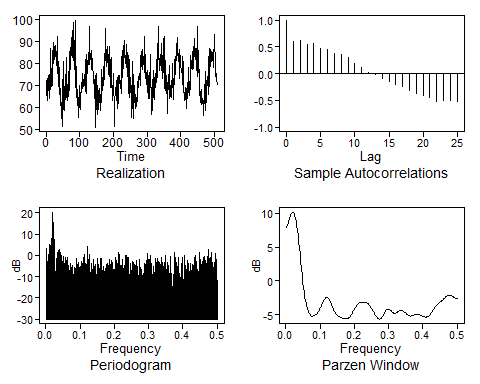
## The following objects are masked from 'package:ggplot2':  
##   
## %+%, alpha

pairs.panels(mort[,-5],   
 method = "pearson", # correlation method  
 hist.col = "#00AFBB",  
 density = TRUE, # show density plots  
 ellipses = TRUE # show correlation ellipses  
 )

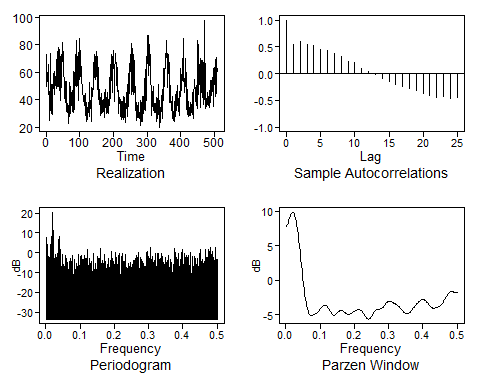
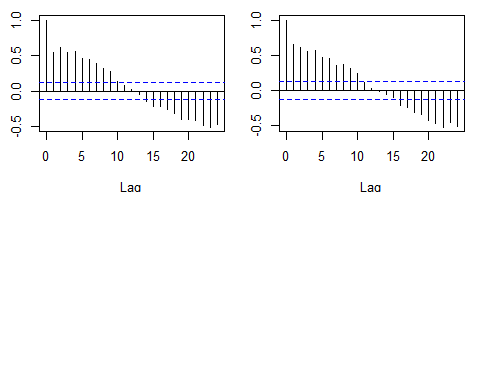


1. Plot the respiratory mortality data.
2. Plot the respiratory mortality data.

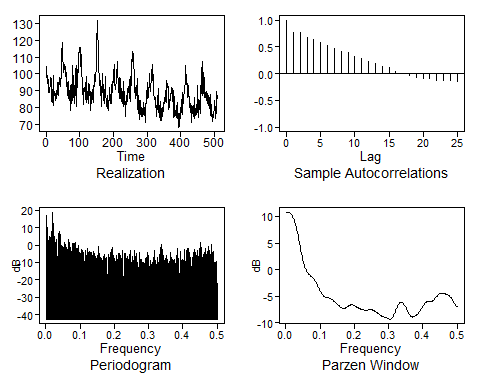
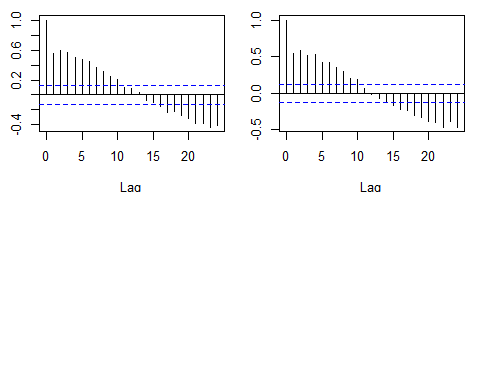
data = read.csv("la\_cmort\_study.csv", header=T)  
  
# plot the temp  
plotts.sample.wge(data$temp)



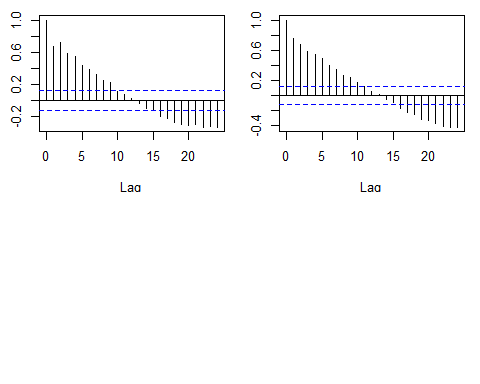
acf(data$temp[1:254]) # acf stationarity check  
acf(data$temp[255:508]) # acf stationarity check  
  
# plot the pollution  
plotts.sample.wge(data$part)



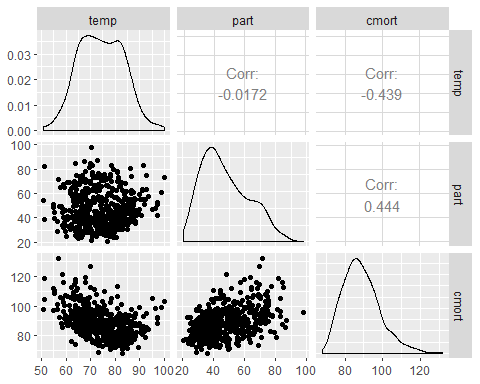
acf(data$part[1:254]) # acf stationarity check  
acf(data$part[255:508]) # acf stationarity check  
  
# plot the cardica mortality  
plotts.sample.wge(data$cmort)



acf(data$cmort[1:254]) # acf stationarity check  
acf(data$cmort[255:508]) # acf stationarity check



ggpairs(data[2:4]) #matrix of scatter plots



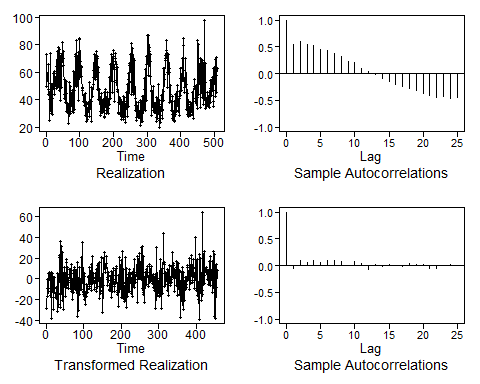
The variables appear independent so we will proceed with a univariate analysis followed by Multiple Linear Regression (MLR) with correlated errors.

### 2. Comment on stationarity or nonstationarity.

The realizations for mortality, particulates, and temperature have some pseudo cyclic behavior with multiple frequencies which is an indicator of correlation. Their means and variances appear constant over time. Their lags do not appear to be dependent on time as indicated by their divided ACFs. There may be some slight wandering with the cardiac mortality realization but this is one realization so the evidence suggests stationarity.

### 3a. Perform a univariate analysis using AR, ARMA, ARMIA or ARUMA. Clearly explain how you arrived at your final model. Build a neural network based model. Build an ensemble model between the two models.

# -- Univariate Analysis of Particulates Using ARUMA --  
  
part52 = artrans.wge(data$part,c(rep(0,51),1)) # since the data is weekly, remove weekly trend



acf(part52) # confirm ACF is ~ white noise, it is!  
  
# Perform model selection  
aic5.wge(part52) # AIC picks ARMA(2,1)

## ---------WORKING... PLEASE WAIT...   
##   
##   
## Five Smallest Values of aic

## p q aic  
## 8 2 1 5.225047  
## 6 1 2 5.227023  
## 11 3 1 5.228819  
## 14 4 1 5.230602  
## 12 3 2 5.230981

aic5.wge(part52,type = "bic") # BIC picks ARMA(2,1)

## ---------WORKING... PLEASE WAIT...   
##   
##   
## Five Smallest Values of bic

## p q bic  
## 8 2 1 5.261210  
## 1 0 0 5.261226  
## 6 1 2 5.263185  
## 4 1 0 5.271478  
## 2 0 1 5.272038

# Check for white noise  
ljung.wge(part52) # FTR, p-value=.11

## Obs -0.05577662 0.1034012 0.05987676 0.1010256 0.05092448 0.09362802 0.09571357 0.08498643 0.0006229897 0.0780049 0.0384301 -0.0647129 0.02773225 -0.01937532 0.01298579 6.928832e-05 -0.009966296 0.03605867 0.01611233 0.01635121 -0.05353086 -0.04577465 0.01134892 0.0299041

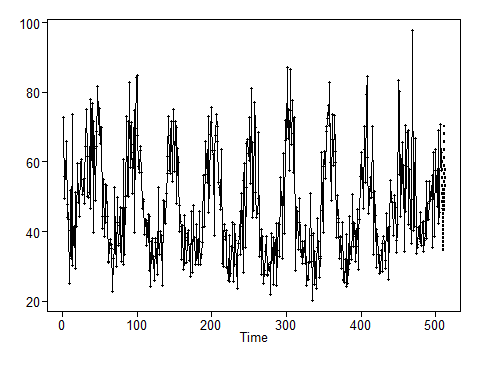
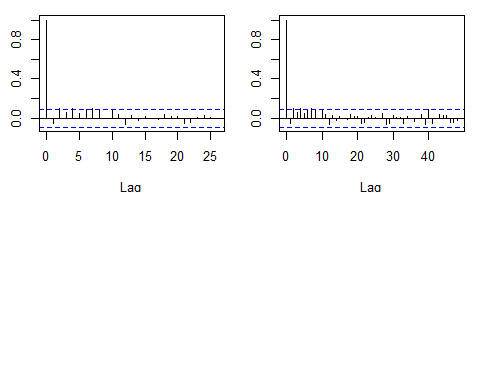
## $test  
## [1] "Ljung-Box test"  
##   
## $K  
## [1] 24  
##   
## $chi.square  
## [1] 35.54288  
##   
## $df  
## [1] 24  
##   
## $pval  
## [1] 0.06074682

ljung.wge(part52, K = 48) # FTR, p-value=.302

## Obs -0.05577662 0.1034012 0.05987676 0.1010256 0.05092448 0.09362802 0.09571357 0.08498643 0.0006229897 0.0780049 0.0384301 -0.0647129 0.02773225 -0.01937532 0.01298579 6.928832e-05 -0.009966296 0.03605867 0.01611233 0.01635121 -0.05353086 -0.04577465 0.01134892 0.0299041 0.004789332 -0.0009208109 0.05183014 -0.0660468 -0.05543925 0.02844053 0.01064092 0.007583799 -0.05369395 0.02009275 -0.003206871 -0.03212606 -0.004742175 0.04009265 -0.06741879 0.0864646 -0.05759432 -0.002036439 0.0327317 0.0237973 0.03137513 -0.04727448 -0.03951558 -0.02110117

## $test  
## [1] "Ljung-Box test"  
##   
## $K  
## [1] 48  
##   
## $chi.square  
## [1] 55.12436  
##   
## $df  
## [1] 48  
##   
## $pval  
## [1] 0.2232435

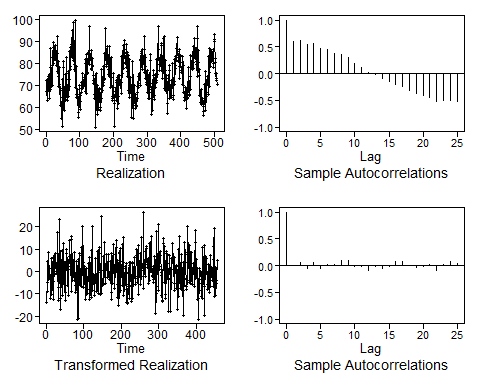
acf(part52,lag.max = 48) # Box-Jenkins, acf looks consistent with white noise as previosly shown  
  
# Plot ARUMA forecast, using ARUMA due to known weekly seasonality of data  
predsPart = fore.aruma.wge(data$part,s = 52, n.ahead=5,limits=F)



Univariate Analysis for Particulates

Since the data is weekly, removing the weekly seasonality occurs first. The differenced data is checked by viewing the resulting ACF with 95% confidence limits and results appear white. Next, the model is selected using the AIC and BIC criteria and both result in a ARMA(2,1) model. The Ljung-Box test with K=24 and K=48 both fail to reject the null and the Box-Jenkins test all support white noise. Since the data is weekly seasonal, an ARUMA forecast is performed and results are saved for later use.

# -- Univariate Analysis of Temperatures Using ARUMA --  
  
temp52 = artrans.wge(data$temp,c(rep(0,51),1)) # since we know the data is weekly, remove weekly trend



acf(temp52) # confirm ACF is ~ white noise, it is!  
  
# model selection  
aic5.wge(temp52) # AIC picks ARMA(0,0)

## ---------WORKING... PLEASE WAIT...   
##   
##   
## Five Smallest Values of aic

## p q aic  
## 1 0 0 4.153692  
## 6 1 2 4.156893  
## 8 2 1 4.157374  
## 4 1 0 4.158061  
## 2 0 1 4.158063

aic5.wge(temp52,type = "bic") # BIC picks ARMA(0,0)

## ---------WORKING... PLEASE WAIT...   
##   
##   
## Five Smallest Values of bic

## p q bic  
## 1 0 0 4.162733  
## 4 1 0 4.176142  
## 2 0 1 4.176144  
## 7 2 0 4.186350  
## 3 0 2 4.186633

# Check for white noise  
ljung.wge(temp52) # FTR, p-value=.129

## Obs 0.004099555 0.05624436 -0.05326951 0.0573932 -0.04861631 0.03098495 0.02718822 0.1004989 0.08851697 -0.01558239 -0.01477092 -0.08380333 -0.01536697 -0.05652122 -0.01491832 0.07303893 0.07379244 0.0110402 -0.03145286 -0.02226394 0.02004529 -0.0838127 0.02392291 0.07002157

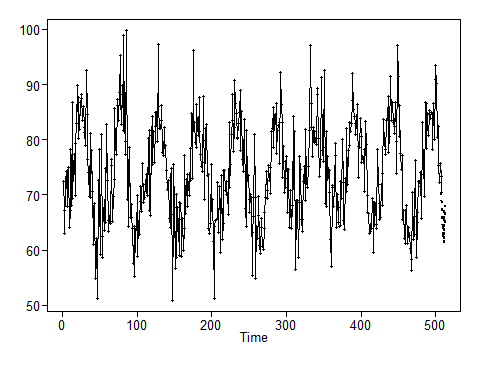
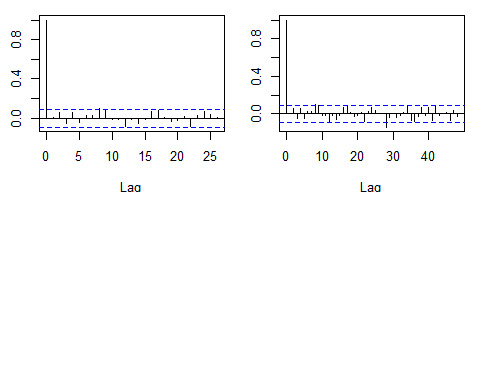
## $test  
## [1] "Ljung-Box test"  
##   
## $K  
## [1] 24  
##   
## $chi.square  
## [1] 31.90274  
##   
## $df  
## [1] 24  
##   
## $pval  
## [1] 0.1294229

ljung.wge(temp52, K = 48) # Barely reject Ho, p-value=.032

## Obs 0.004099555 0.05624436 -0.05326951 0.0573932 -0.04861631 0.03098495 0.02718822 0.1004989 0.08851697 -0.01558239 -0.01477092 -0.08380333 -0.01536697 -0.05652122 -0.01491832 0.07303893 0.07379244 0.0110402 -0.03145286 -0.02226394 0.02004529 -0.0838127 0.02392291 0.07002157 0.03883168 0.004168604 0.005689672 -0.141479 -0.03792635 0.007452253 -0.03892399 -0.01262782 0.01859873 0.09041441 -0.06857444 -0.07895039 -0.02881891 0.06810352 -0.0138232 0.06763404 -0.06651562 0.07729948 -0.01803334 -0.0007800483 0.01314484 -0.06913632 0.03518071 -0.0288905

## $test  
## [1] "Ljung-Box test"  
##   
## $K  
## [1] 48  
##   
## $chi.square  
## [1] 67.73833  
##   
## $df  
## [1] 48  
##   
## $pval  
## [1] 0.0317127

acf(temp52,lag.max = 48) # Box-Jenkins, acf looks consistent with white noise as previosly shown  
  
# although Ljung0-Box with K=48 barely rejects Ho, K=24 FTR Ho and Box-Jenkins ACF is consisten with white noise, looks good  
  
# Plot ARUMA Forecast  
predsTemp = fore.aruma.wge(data$temp, s=52, n.ahead=5, limits=F)



Univariate Analysis for Temperature

Since the data is weekly, removing the weekly seasonality occurs first. The differenced data was checked by viewing the resulting ACF with 95% confidence limits and results appear white. Next, the model was selected using the AIC and BIC criteria and both resulted in a ARMA(0,0) model. The Ljung-Box test with K=24 failed to reject the null but the Ljung-Box test with K=48 barely rejects Ho with a p-value=.032. However, the Box-Jenkins test support white noise so no concerns. Since the data is weekly seasonal, an ARUMA forecast is performed and results are saved for later use.

# -- Model cmort based on predicted part, predicted temp, and Week using MLR with Correlated Errors --  
  
ksfit = lm(cmort~temp+part+Week, data = data) # get linear fit to access residuals  
  
# get model for residuals  
phi = aic.wge(ksfit$residuals) # aic selects AR(2)  
phi

## $type  
## [1] "aic"  
##   
## $value  
## [1] 3.528269  
##   
## $p  
## [1] 2  
##   
## $q  
## [1] 0  
##   
## $phi  
## [1] 0.2477388 0.3581387  
##   
## $theta  
## [1] 0  
##   
## $vara  
## [1] 33.66498

# fit arima with residual phis, remove weekly seasonality and incl ext vars temp, part, week  
attach(data)

## The following objects are masked from package:astsa:  
##   
## cmort, part

fit = arima(cmort,order=c(phi$p,0,0), seasonal=list(order=c(1,0,0),period=52), xreg = cbind(temp, part, Week))  
fit

##   
## Call:  
## arima(x = cmort, order = c(phi$p, 0, 0), seasonal = list(order = c(1, 0, 0),   
## period = 52), xreg = cbind(temp, part, Week))  
##   
## Coefficients:  
## ar1 ar2 sar1 intercept temp part Week  
## 0.3909 0.4149 0.082 87.5975 0.0258 0.1407 -0.0292  
## s.e. 0.0441 0.0404 0.049 4.0150 0.0530 0.0290 0.0086  
##   
## sigma^2 estimated as 28.81: log likelihood = -1575.14, aic = 3166.29

# The intercept is significantly different from 0, the particulates and Time are significant, but the temperature is not.  
  
# Check residuals for white noise  
ljung.wge(fit$residuals) # Barely reject, p-value=.048

## Obs -0.008773997 -0.006778 0.02847037 0.001592821 0.04173621 0.03014614 -0.02519791 -0.03614844 0.08896799 -0.04873154 0.003349583 0.04272511 -0.05378803 -0.03121467 0.1104465 -0.0511791 -0.05781976 -0.05203572 -0.1226383 -0.01661192 0.03274971 -0.06125045 -0.08319911 0.009083812

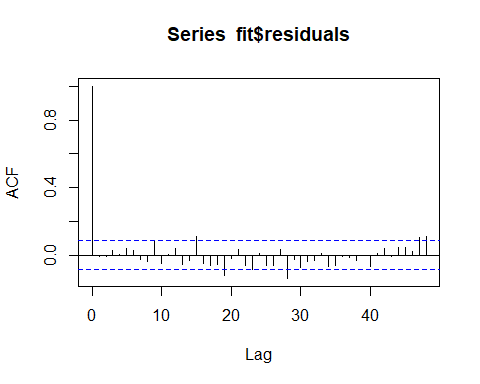
## $test  
## [1] "Ljung-Box test"  
##   
## $K  
## [1] 24  
##   
## $chi.square  
## [1] 36.59418  
##   
## $df  
## [1] 24  
##   
## $pval  
## [1] 0.04801085

ljung.wge(fit$residuals, K = 48) # Reject Ho, p-value=.003

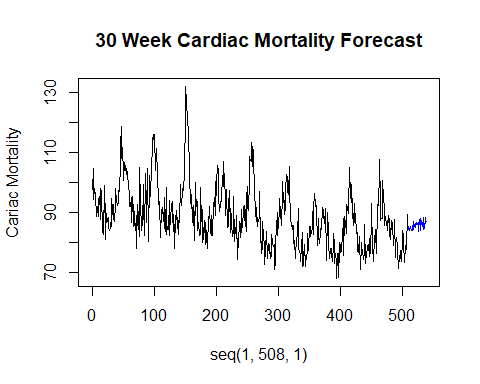
## Obs -0.008773997 -0.006778 0.02847037 0.001592821 0.04173621 0.03014614 -0.02519791 -0.03614844 0.08896799 -0.04873154 0.003349583 0.04272511 -0.05378803 -0.03121467 0.1104465 -0.0511791 -0.05781976 -0.05203572 -0.1226383 -0.01661192 0.03274971 -0.06125045 -0.08319911 0.009083812 -0.06018696 -0.05789479 0.03458024 -0.1371038 -0.02600092 -0.07277285 -0.03981533 -0.03157352 0.009442916 -0.06678764 -0.05777429 -0.004599118 -0.01492029 -0.03032822 -0.004028157 -0.06378089 0.01029915 0.03886128 -0.006943152 0.04753463 0.04475963 0.0230044 0.1033364 0.110844

## $test  
## [1] "Ljung-Box test"  
##   
## $K  
## [1] 48  
##   
## $chi.square  
## [1] 79.51156  
##   
## $df  
## [1] 48  
##   
## $pval  
## [1] 0.002852882

acf(fit$residuals,lag.max = 48) # Box-Jenkins, acf looks



# acf show maybe 3 points out of 40 outside the limits but this really isn't bad  
  
# build DF of predicted variables  
last30 = data.frame(temp = data$temp[479:508], part = data$part[479:508], Week = seq(479,508,1))  
  
# get predictions  
predsCMort = predict(fit,newxreg=last30)  
  
# plot the next 30 cmorts  
plot(seq(1,508,1),cmort,type="l",xlim=c(0,538),ylab="Cariac Mortality",main="30 Week Cardiac Mortality Forecast")  
lines(seq(509,538,1),predsCMort$pred,type="l",col="blue")



ASE = mean((data$cmort[479:508] - predsCMort$pred)^2)  
ASE

## [1] 56.59301

Cmort Multiple Linear Regression with Correlated Errors

The goal is to develop a model to predict cardiac mortality. Since the variables were seen earlier to be independent a MLR model using correlated errors was chosen. A linear fit was performed using particulates, temperature, and time. AIC selecetd an AR(2) model for the residuals. The ARIMA function was used to fit the residuals, applying weekly seasonality and including external varibles temperature, particulate, and time (week #). The fitted results showed the intercept is significantly different from 0, the particulates and time are significant, but the temperature is not.

Next, the residuals were checked for white noise. The Ljung-Box test with K=24 barely rejects with a p-value=.048 and with K=48 it rejects Ho. When reviewing the ACFs, maybe 3 points our of 40 are outside the limits but this is not bad. Judgement is to proceed.

A dataframe of the predicted temperature and particulates from the univariate analysis along with the week number was created. The plot above shows the realization with the predicted values in blue. The ASE for this model is **56.59**.

### 3b. Compare the models and describe which univariate model you feel is the best and why.

### 4a. Perform a multivariate analysis using at least a VAR or MLR with correlated errors and a MLP model. Clearly explain how you arrived at the final model. Use forecasted values of the predictors where appropriate.

### 4b. Fit and evaluate an ensemble model from the models you fit in 4a.

### 4c. Compare these models and describe which multivariate model you feel is the best and why.

### 5. Use the model you feel is most useful to forecat the next 5 weeks of respiratory mortality.