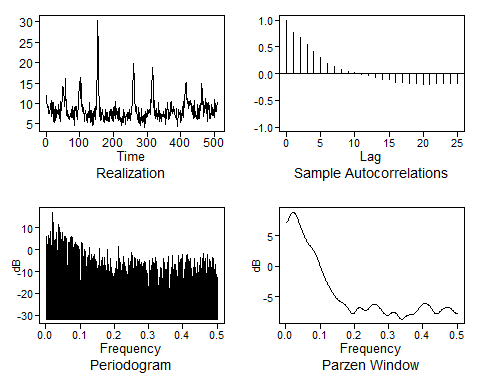
Time Series Final

Steve Bramhall

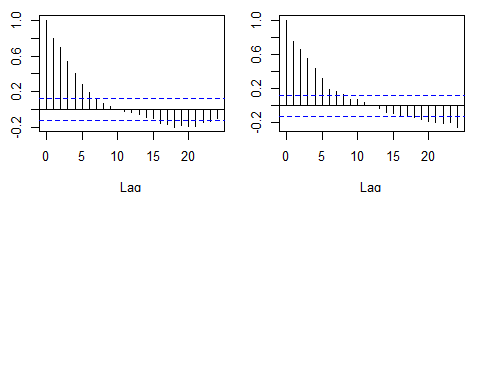
August 13, 2019

1. Plot the respiratory mortality data.

data(lap) # import data  
data=data.frame(date=time(lap),Time=as.factor(seq(1,508,1)),as.matrix(lap)) # create data frame of data  
  
# plot the resp mortality  
plotts.sample.wge(data$rmort)



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

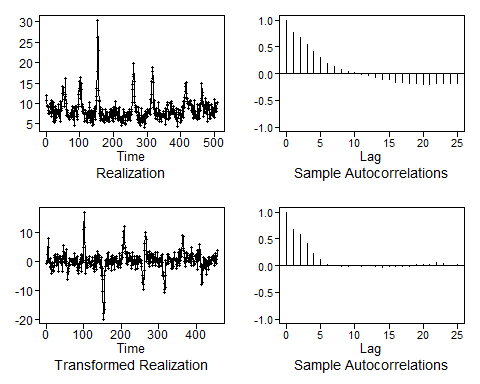


1. Comment on stationarity or nonstationarity.

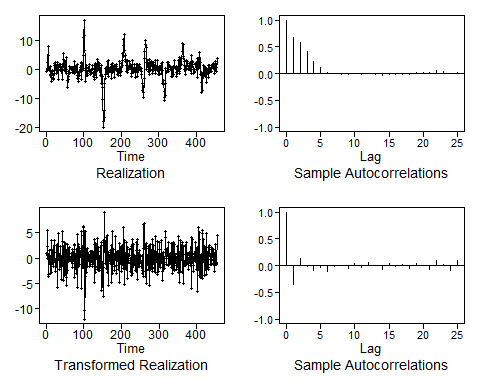
The realization for respiratory mortality shows some pseudo cyclic behavior which makes sense since the data is weekly. The variance does not appear constant over time due to the pseudo cyclic spikes. The lags do not appear to be dependent on time as indicated by their divided ACFs. The evidence suggests nonstationarity. Since the data is weekly this is expected and the seasonality will be removed.

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.

dif1 = artrans.wge(data$rmort,c(rep(0,51),1)) # since the data is weekly, remove weekly trend



dif2 = artrans.wge(dif1,1) # some fairly strong autocorrelation seen so let's remove some trend



acf(dif2,lag.max=50) # now we are closer to being white  
  
# Perform model selection  
aic5.wge(dif2) # AIC picks ARMA(3,2)

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

## p q aic  
## 12 3 2 1.845795  
## 16 5 0 1.846906  
## 4 1 0 1.855495  
## 13 4 0 1.855811  
## 3 0 2 1.858729

aic5.wge(dif2,type = "bic") # BIC picks ARMA(1,0)

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

## p q bic  
## 4 1 0 1.873607  
## 3 0 2 1.885896  
## 7 2 0 1.886673  
## 5 1 1 1.886761  
## 2 0 1 1.891391

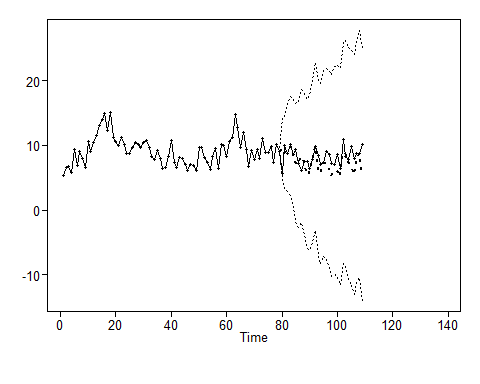
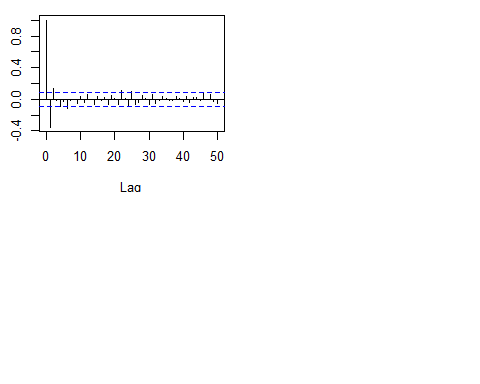
# produce BIC recommended model, AR(1)  
estAR1=est.arma.wge(dif2,p=1,q=0)

##   
## Coefficients of Original polynomial:   
## -0.3535   
##   
## Factor Roots Abs Recip System Freq   
## 1+0.3535B -2.8289 0.3535 0.5000  
##   
##

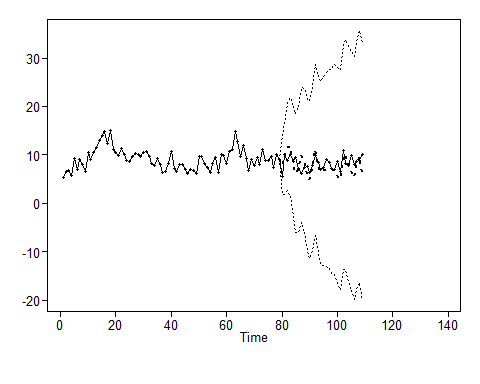
# produce AIC recommended the model, ARMA(3,2)  
estARMA32=est.arma.wge(dif2,p=3,q=2)

##   
## Coefficients of Original polynomial:   
## 0.3322 -0.6921 -0.3198   
##   
## Factor Roots Abs Recip System Freq   
## 1-0.6775B+0.9261B^2 0.3658+-0.9726i 0.9623 0.1927  
## 1+0.3454B -2.8954 0.3454 0.5000  
##   
##

# with the AR(1) parameters, let's create a forecast of the last 30 values  
foreAR1=fore.aruma.wge(data$rmort[400:508],phi=estAR1$phi,theta=estAR1$theta,n.ahead=30,s=52,d=1,lastn=T)



# with the ARMA(3,2) parameters, let's create a forecast of the last 30 values  
foreARMA32=fore.aruma.wge(data$rmort[400:508],phi=estARMA32$phi,theta=estARMA32$theta,n.ahead=30,s=52,d=1,lastn=T)



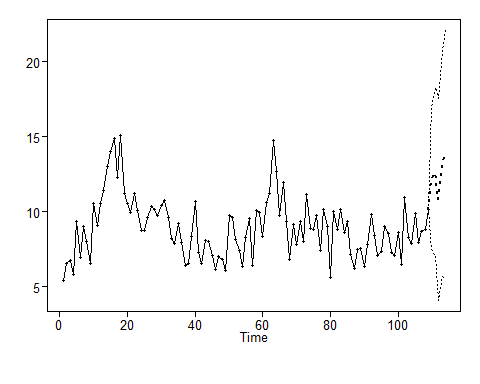
# Get ASE from AR(1)  
ASE1 = mean((data$rmort[(508-30+1):508]-foreAR1$f)^2)  
ASE1

## [1] 3.44117

# Get ASE from ARMA(3,2)  
ASE2 = mean((data$rmort[(508-30+1):508]-foreARMA32$f)^2)  
ASE2

## [1] 3.516085

# The BIC recommended model with an ASE=3.44 is the selected ARUMA model.  
# forecast next 5 points with BIX ARUMA  
foreAR1.new=fore.aruma.wge(data$rmort[400:508],phi=estAR1$phi,theta=estAR1$theta,n.ahead=5,s=52,d=1,lastn=F)



Since the data is weekly, the weekly seasonality was removed. Although trend was not seen, there was still some strong autocorrelations so another difference was performed to obtain data that more closely resembled white noise. Four to five autocorrelations were greater than the limits (with two-three barely out) and out of fifty autocorrelations this is acceptable. The first two autocorrelations and strong which suggests an AR2.

Then AIC was used to select a model and an AR(3,2) model was selected. Using BIC as the selection criteria, an AR(1) model was selected. Since the AIC and BIC did not agree on the same model both were used to create an ARUMA forecast. It should be noted that the estimated ARMA(3,2) had a root with an Absolute Reciprical = 0.9623 at a System Freq = .1927. Both estimated models has system frequencies at 0.5 with Absolute Recipricals at ~0.35. The parameters for the AR(1) and ARMA(3,2) were used to create ARUMA forecasts for the last 30 weeks.

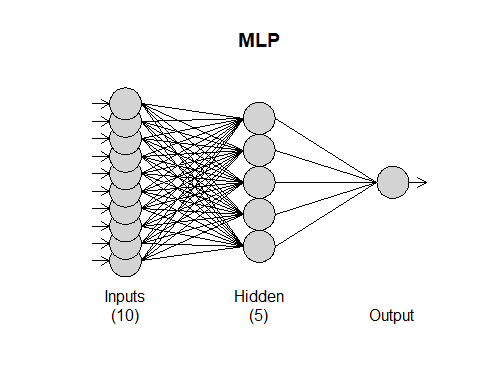
The ASE was calculated for both models using their predicted values and the actual values. The ARUMA model with AR(1) parameters (BIC recommended) produced an ASE = 3.44117 and the ARUMA model with ARMA(3,2) parameters (AIC recommended) produced an ASE = 3.516085. The selected model will be the one with the lower ASE score.

ARUMA with estimated AR(1) paramters -> ASE = 3.44

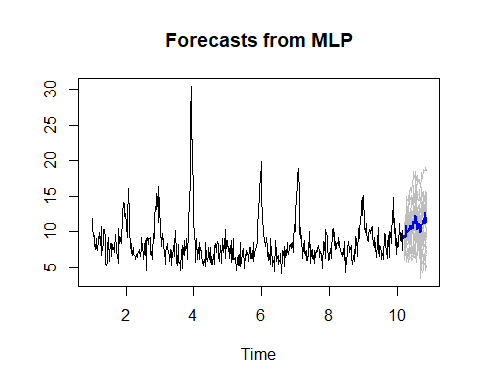
set.seed(2)  
  
# training data  
rmortTrain = ts(data$rmort[1:478],frequency=52)   
  
# test data - last 30 weeks   
rmortTest = ts(data$rmort[479:508],frequency=52)  
  
# fit the NN model, let the function identify the best diff order(s)  
fit.mlp=mlp(rmortTrain,difforder=NULL,allow.det.season = FALSE)  
fit.mlp

## MLP fit with 5 hidden nodes and 20 repetitions.  
## Series modelled in differences: D1.  
## Univariate lags: (1,5,6,7,18,21,30,36,42,47)  
## Forecast combined using the median operator.  
## MSE: 1.6346.

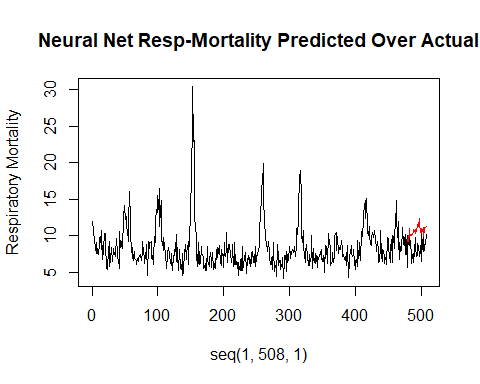
# plot the NN  
plot(fit.mlp)



# forecast the test data + 5 more weeks  
fore.mlp=forecast(fit.mlp,h=35)  
  
# plot the forecast of the test data  
plot(fore.mlp)



# Plot forecast of test data over actual data  
plot(seq(1,508,1), data$rmort, type = "l",xlim = c(0,508), ylab = "Respiratory Mortality", main = "Neural Net Resp-Mortality Predicted Over Actual")  
lines(seq(479,508,1), fore.mlp$mean[1:30], type = "l", col = "red")



# calc ASE based on the forecasted and actual test data  
ASE3 = mean((data$rmort[479:508]-fore.mlp$mean[1:30])^2)  
ASE3

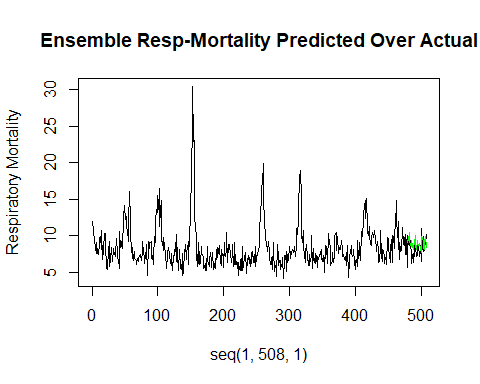
## [1] 7.810705

Next, a multilayer perceptron (MLP) function was used to create a neural network model. Default parameters were used in addition to telling the function to identify the best difference order(s) and restricting modeling seasonality with deterministic dummies.

The neural network model has 5 hidden nodes and 20 repetitions. There were 10 univariate lags which are the inputs. The neural network model produced and ASE = 10.23 which is not as good as the ARUMA model. Predicted values appear to be above the actuals.

Neural Network Model -> ASE = 10.23

# build ensemble using ave of the two forecasts  
ensemble = (foreAR1$f + fore.mlp$mean[1:30])/2  
  
# plot the ensemble results  
plot(seq(1,508,1), data$rmort, type = "l",xlim = c(0,508), ylab = "Respiratory Mortality", main = "Ensemble Resp-Mortality Predicted Over Actual")  
lines(seq(479,508,1), ensemble, type = "l", col = "green")



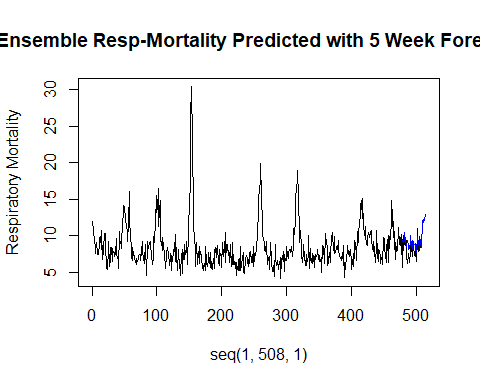
# calc ensemble ASE  
ASE4 = mean((data$rmort[(508-30+1):508] - ensemble)^2)  
ASE4

## [1] 1.985225

# forecast of next 5 weeks  
foreAR1.all = append(foreAR1$f,foreAR1.new$f) # AR1 data with 5 week forecast  
ensemble.new = (foreAR1.all + fore.mlp$mean[1:35])/2  
ensemble.new[31:35] # predicted 5 week values

## t+31 t+32 t+33 t+34 t+35   
## 11.91340 12.23596 11.74706 12.16413 12.91565

plot(seq(1,508,1), data$rmort, type = "l",xlim = c(0,515), ylab = "Respiratory Mortality", main = "Ensemble Resp-Mortality Predicted with 5 Week Forecast")  
lines(seq(479,513,1), ensemble.new, type = "l", col = "blue")



An ensemble model was built using the average of the ARUMA and Neural Network forecasts. The predicted values are a little above the actual values but this model produced the lowest ASE at 2.29. Since this model produced the lowest ASE, it is the recommended model for the univariate analysis or repiratory mortalty. The forecasted next 5 weeks resembles the previous patterns with values of (12.31, 12.65, 12.21, 12.98, 13.39).

Ensemble Model -> ASE = 2.29 (seleced model)

### 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.

# Check for white noise

ljung.wge(dif2) # Reject Ho, p-value=6.7e-11 so Ljung-Box with K=24 indicates not white ljung.wge(dif2, K = 48) # Reject Ho, p-value=4.1e-08 so Ljung-Box with K=48 indicates not white acf(dif2,lag.max = 50) # Box-Jenkins, not quite white but close, lag1 is still a bit strong