Crystal Mosley

Predict 413, Sec 55

Homework #3

All code at the end of file.

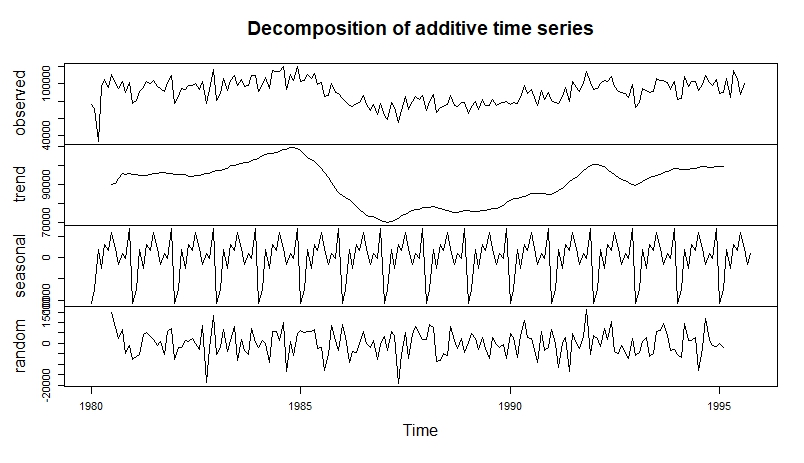
For this assignment, I chose to use *Monthly total number of pigs slaughtered in Victoria. Jan 1980 -August 1995*.

**ETS Model—**

Initial plot of the data:

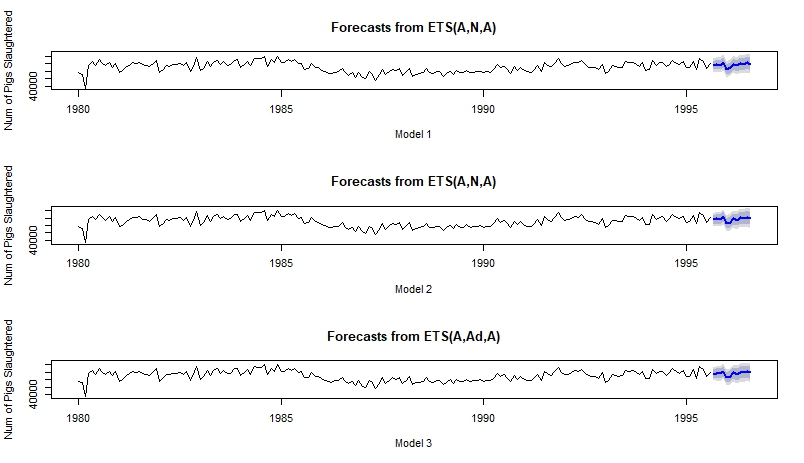


To get a better understanding of this data, I utilized R to determine whether my data was additive or multiplicative in its decomposition. R chose an additive model. Based on this decomposition, there appears to be some seasonality and trending.



In building my ETS models, I let R determine the best version for Model 1, which turned out to be an additive error model, resulting in a forecasted constant of about 15,326 slaughters over the next 12 months. Given my observations in the decomposition, I chose to include an additive seasonality effect in Model 2 and let R determine whether trend should be included. This resulted in a much more flexible forecast that appeared to closely reflect the actual data. Finally, for Model 3, I built an “AAA” model to include trend, the resulting forecast of which appeared to be the same as Model 2.

Forecasts of models:



Then, I compared each model on multiple metrics to determine their accuracy rates and performance status. R’s model (model 1) performed the exact same as model 2. Model 3 out-performed model 1 and 2 on almost every metric except MPE. This suggests that the *Number of Pigs Slaughtered* time series data is seasonal but does not necessarily have a clear trend pattern.

Accuracy Metrics of Models:

accuracy(model1.pred)

ME RMSE MAE MPE MAPE MASE ACF1

Training set 67.41596 8919.631 6406.822 -1.095397 7.74615 0.6410738 0.009262554

accuracy(model2.pred)

ME RMSE MAE MPE MAPE MASE ACF1

Training set 67.41596 8919.631 6406.822 -1.095397 7.74615 0.6410738 0.009262554

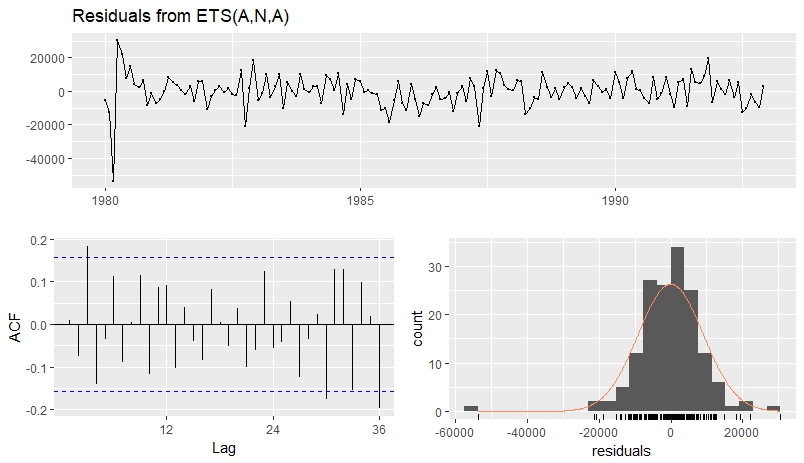
accuracy(model3.pred)

ME RMSE MAE MPE MAPE MASE ACF1

Training set -107.1603 8893.625 6378.186 -1.294863 7.725895 0.6382084 0.009024906

**Auto.Arima Model—**

Next, I decided to build an auto arima model. The residuals reveal that there is a presence of autocorrelation (3 lags), as is indicated by the ACF plot below, as well as the significant p-value resulting from the Ljung-Box test (p-value = 0.0004242).



Next, an auto.arima( ) model was fit using the full dataset and setting stepwise=FALSE and approximation = FALSE to get a more accurate fit. R chose to fit an ARIMA(2,1,0)(2,0,0)[12] with Drift model. R built a seasonal ARIMA model, as to be expected since we noted the strong presence of seasonality in the data. This model assumes 2 AR order, 1 differencing, and 0 Moving Average Order (ARIMA(2,1,0)). The second part (2,0,0), includes additional seasonal terms to account for the seasonality. The [12] means that this is monthly data. The residuals were checked, and although it appears that not as many “spikes” fall outside of the ACF significance bounds, there is still a significant presence of autocorrelation as indicated by the p-value from the Ljung-Box test (p-value = 0.7623).

Next, use the last 6 months of the test dataset. Based on the below accuracy metrics, the ETS model performed slightly better on the test data set with the exceptions of MAE/MASE/Theil’s U. When the two forecasts are graphed and compared to the actuals, it is apparent that both the ARIMA and ETS models are kind of neck and neck in performance.

accuracy(ets\_model\_fc, pigTest)

ME RMSE MAE MPE MAPE MASE ACF1 Theil's U

Training set -47.09261 9036.815 6462.945 -1.302944 8.011256 0.6298813 0.0080831 NA

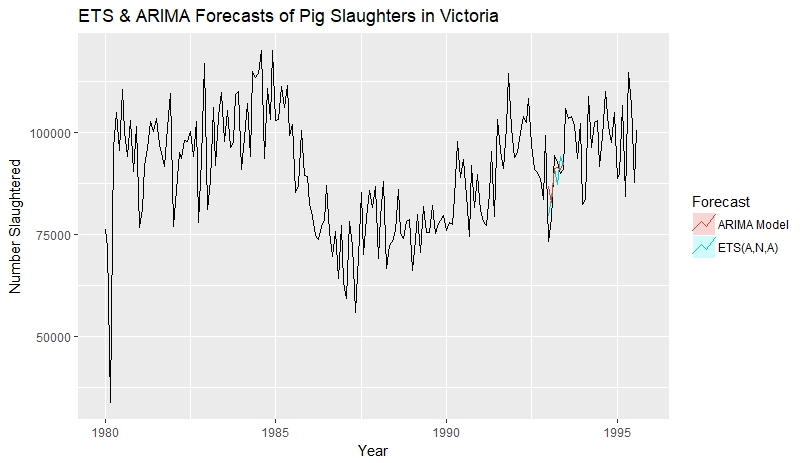
Test set -1237.68725 4512.650 4164.305 -1.824097 4.955387 0.4058549 0.1093265 0.5093421

accuracy(auto\_arima\_model\_fc, pigTest)

ME RMSE MAE MPE MAPE MASE ACF1 Theil's U

Training set -4.872528 9232.632 6984.994 -0.9080213 8.210954 0.6807604 0.03320563 NA

Test set -2581.748379 5987.607 4140.161 -3.5757721 5.234712 0.4035018 0.22164360 0.3428946



**Neural Network Model—**

Lambda is set to “auto” to R can pick the best route. After reviewing the accuracy and performance metrics, the NN outperformed the AA and ETS models. The graph shows that the ARIMA model did slightly better than the ETS model, but not by much.

accuracy(pig\_fit\_nn\_fc, pigTest)

ME RMSE MAE MPE MAPE MASE ACF1 Theil's U

Training set 373.0517 7844.270 6382.803 -0.4001988 7.230203 0.6220706 0.068101679 NA

Test set -4294.4160 9755.969 7805.283 -5.4264552 8.677513 0.7607061 0.004394598 0.723564

accuracy(pig\_fit\_aa\_fc, pigTest)

ME RMSE MAE MPE MAPE MASE ACF1 Theil's U

Training set 122.9786 9871.615 7269.800 -0.6377762 8.362061 0.7085177 0.02832763 NA

Test set 5257.6456 9978.272 8297.453 4.6072646 8.334531 0.8086732 0.07066870 0.7669512

accuracy(pig\_fit\_ets\_fc, pigTest)

ME RMSE MAE MPE MAPE MASE ACF1 Theil's U

Training set 216.713 10221.45 7836.424 -1.143699 9.366839 0.7637411 0.08800373 NA

Test set 3945.807 10582.68 8897.283 3.033598 9.159136 0.8671329 0.10791936 0.809694



**R Code—**

library(fpp2)

library(forecast)

library(ggplot2)

pig\_data <- read.csv("monthly\_pigs\_slaughtered.csv")

pigTS <- ts(pig\_data[,2],start = c(1980,1),frequency = 12)

pigTrain <- window(pigTS,end=c(1992,12))

pigTest <- window(pigTS, start = c(1993,1))

# Initial plot of data

autoplot(pigTS) +

ggtitle('Monthly total number of pigs slaughtered in Victoria. Jan 1980 - August 1995') +

ylab('Units (thousounds)')

#-------------------------------------------------------------------------------------------------

# Build ETS Models

dec <-decompose(pigTS)

plot(dec)

model1 <- ets(pigTS)

model1.pred <- forecast(model1, h=12)

model2 <- ets(pigTS, model = "AZA")

model2.pred <- forecast(model2, h=12)

model3 <- ets(pigTS, model = "AAA")

model3.pred <- forecast(model3, h=12)

par(mfrow=c(3,1))

plot(model1.pred,xlab = "Model 1",ylab = "Num of Pigs Slaughtered")

plot(model2.pred,xlab = "Model 2",ylab = "Num of Pigs Slaughtered")

plot(model3.pred,xlab = "Model 3",ylab = "Num of Pigs Slaughtered")

par(mfrow=c(1,1))

accuracy(model1.pred)

accuracy(model2.pred)

accuracy(model3.pred)

#---------------------------------------------------------------------------------------

# Build auto.arima model

ets\_trainmodel <- ets(pigTrain)

checkresiduals(ets\_trainmodel) #lag at 3

auto\_arima\_trainmodel <- auto.arima(pigTrain, stepwise=FALSE, approximation = FALSE)

checkresiduals(auto\_arima\_trainmodel)

#forecast

ets\_model\_fc <- forecast(ets\_trainmodel, 6)

auto\_arima\_model\_fc <- forecast(auto\_arima\_trainmodel, 6)

accuracy(ets\_model\_fc, pigTest)

accuracy(auto\_arima\_model\_fc, pigTest)

autoplot(pigTS) +

autolayer(ets\_model\_fc, PI=FALSE, series = "ETS(A,N,A)") +

autolayer(auto\_arima\_model\_fc, PI=FALSE, series = "ARIMA Model") +

xlab("Year") + ylab("Number Slaughtered") +

ggtitle("ETS & ARIMA Forecasts of Pig Slaughters in Victoria") +

guides(color=guide\_legend(title="Forecast"))

#--------------------------------------------------------------------------------------

# Build NN Model

pig\_fit\_nn <- nnetar(pigTrain, lambda="auto")

pig\_fit\_aa <- auto.arima(pigTrain, lambda="auto")

pig\_fit\_ets <- ets(pigTrain, lambda="auto")

pig\_fit\_nn\_fc <- forecast(pig\_fit\_nn, h=48)

pig\_fit\_aa\_fc <- forecast(pig\_fit\_aa, h=48)

pig\_fit\_ets\_fc <- forecast(pig\_fit\_ets, h=48)

accuracy(pig\_fit\_nn\_fc, pigTest)

accuracy(pig\_fit\_aa\_fc, pigTest)

accuracy(pig\_fit\_ets\_fc, pigTest)

autoplot(pigTest)+

autolayer(forecast(pig\_fit\_nn, h=48), series = "NN(2,1,2)[12]", PI = F)+

autolayer(forecast(pig\_fit\_aa, h=48), series = "ARIMA(0,1,1)(0,1,1)[12]", PI = F)+

autolayer(forecast(pig\_fit\_ets, h=48), series = "ETS(A,A,A)", PI = F)+

ylab("pigTest") + xlab("Year") + ggtitle("Monthly total number of pigs slaughtered in Victoria. Jan 1980 - August 1995")