Studying the nature of coarse wool prices from 1990-2017

Pranav Gopalkrishna, 1940223

2022-01-19

# INTRODUCTION & AIM

We will be studying coarse wool prices from 1990 to 2017 (recorded monthly). This data is from a dataset containing information about the prices of various agricultural raw materials.

Using time plots and autocorrelation function plots to understand the nature of coarse wool prices from 1990 to 2017 (recorded monthly). We will aim to identify the components of this time series, comment on the level of autocorrelation in this time series, and analyze its stationarity.

# DATA

This dataset records data about the prices of various agricultural raw materials, including coarse wool, cotton, fine wool, sawn wood, etc., from the year 1990 up to 2017. The data points are separated by a monthly time interval i.e. data was observed for every month. In this study, we will only look at coarse wool prices.

setwd("~/Documents/Study/computerScience/programming/r/data/")  
data = read.csv("agriculturalRawMaterial.csv")[c(1, 2)]  
head(data)

**Month | Coarse.wool.Price**  
Apr-90 | 482.34  
May-90 | 447.26  
Jun-90 | 440.99  
Jul-90 | 418.44  
Aug-90 | 418.44  
Sep-90 | 412.18

## Reformatting the columns for obtaining better data summary

### Reformatting ‘Month’ column as date objects

dates = c()  
for(d in data$Month)  
{  
 d = paste("01", d, sep = "-")  
 dates = c(dates, d)  
}  
dates = as.Date(dates, format = "%d-%b-%y")  
# %b => abbreviated month  
# %y => 2 digit year  
# It can recognize century on its own.  
# So for example, '93' will be interpreted as '1993'.  
data$Month = dates

### Reformatting ‘Coarse.wool.Price’ column as numeric values

# Possible commas must be removed as well.  
prices = c()  
for(p in data$Coarse.wool.Price)  
{  
 p = gsub(',', '', p)  
 # Removing commas in the numbers  
 # 1st argument => what to replace  
 # 2nd argument => what to put instead  
 # 3rd argument => full string  
 prices = c(prices, p)  
}  
prices = as.numeric(prices)  
data$Coarse.wool.Price = prices

## Summarizing the data

summary(data)

**Month | Coarse.wool.Price**  
Min. :1990-04-01 | Min. : 247.1   
1st Qu.:1997-10-01 | 1st Qu.: 369.6   
Median :2005-04-01 | Median : 525.1   
Mean :2005-04-01 | Mean : 626.3   
3rd Qu.:2012-10-01 | 3rd Qu.: 847.1   
Max. :2020-04-01 | Max. :1391.5   
 | NA's :34

We can see there are 34 missing values in the price column. Checking in the dataset itself, values for coarse wool prices are available until a certain point, after which we have these missing values. Hence, we can simply remove the tail end of the dataset where these missing values are concentrated.

### Removing empty rows

Month = data$Month[c(1:(length(data$Month) - 34))]  
CoarseWoolPrice = data$Coarse.wool.Price[c(1:(length(data$Coarse.wool.Price) - 34))]  
data = data.frame(Month, CoarseWoolPrice)  
summary(data)

**Month | CoarseWoolPrice**   
Min. :1990-04-01 | Min. : 247.1   
1st Qu.:1997-01-16 | 1st Qu.: 369.6   
Median :2003-11-01 | Median : 525.1   
Mean :2003-10-31 | Mean : 626.3   
3rd Qu.:2010-08-16 | 3rd Qu.: 847.1   
Max. :2017-06-01 | Max. :1391.5

## Creating a time series for coarse wool prices

CoarseWoolPrice = ts(CoarseWoolPrice, start = c(1990, 4, 1), end = c(2017, 6, 1), frequency = 12)

**NOTE**: frequency = 12 -> monthly frequency

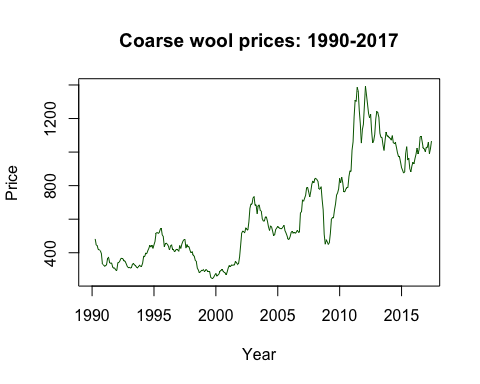
# TIME PLOT AND TIME SERIES COMPONENTS

## Function to create time plots for different year ranges

timeplot = function(min, max)  
{  
 ts.plot(CoarseWoolPrice,  
 main = paste("Coarse wool prices: ", min, "-", max, sep = ''),  
 ylab = "Price",  
 xlab = "Year",  
 xlim = c(min, max),  
 col = "darkgreen")  
}

## Plotting from 1990-2017

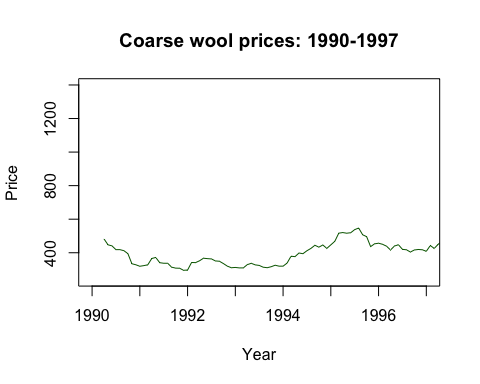
timeplot(1990, 2017)

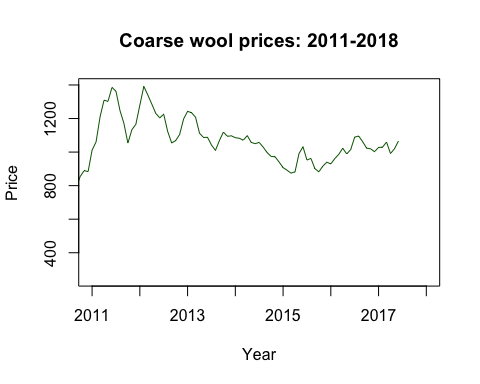


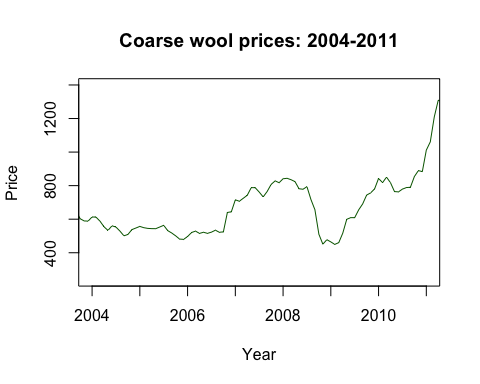
From the above time plot, we observe noticeable upward trend over time, especially from 2000-2012. However, this trend is not constant, and tends to stagnate in certain periods. We also observe multiple irregular fluctuations that vary significantly in size. To observe seasonal fluctuations, we will fo the following plots…

## **Time plots for smaller time periods**

timeplot(1990, 1997)  
timeplot(1997, 2004)  
timeplot(2004, 2011)  
timeplot(2011, 2018)





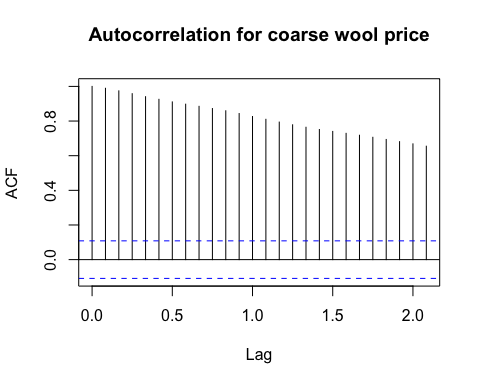


From the above plots, we may observe some seasonal fluctuation, wherein prices are higher at the start of the year and dip towards the end (often slightly), before rising again just before the next year. This pattern is not very consistent or noticeable, but it may be observed for many of the years, upon closer inspection. For the given data, we cannot clearly discern any cyclical fluctuations.

# AUTOCORRELATION FUNCTION PLOT

Autocorrelation is the presence of correlation between observations of the same variable. This tends to occur when previous values of a variable can subsequent future values of the variable. Autocorrelation is often found in time series data, since past values may have an effect on future values. Here, we aim to see if this seems to be the case for coarse wool prices over time.

acf(CoarseWoolPrice, main = "Autocorrelation for coarse wool price")



Here we observe strong correlation between observations that are separated by lags from between 1 to 2. Here, the lags are in monthly intervals. Hence, we can say that coarse wool prices are strongly correlated to the coarse wool prices of the previous two months, i.e. there is high autocorrelation in coarse wool prices at lags 1 and 2.

# CONCLUSIONS

The above interpretations indicate that the time series is non-stationary, since we have

* No constant mean
* No constant variance

However, the autocorrelation in the time series is dependent on lag, since we see a steady decrease in the autocorrelation coefficient with increase in lag. It is possible that for a larger sample, we may observe stationary data for coarse wool prices, since the irregular variations are many and varied, and are seemingly more influential than seasonal or cyclical variations. Hence, over time, we may observe more random data around a constant mean.

Alternatively, we may observe cyclical variations due to yet undiscovered factors. Finally, we must note that while not very significant, there is likely a seasonal component in the data, which makes sense since demand for wool clothing may be expected to go up in colder seasons, and go down in warmer seasons.