6/7/16

Time Series Analysis: Home Sales

Allan Swartz

Introduction. The study was conducted on time series data of the monthly total of new one-family home sales in the US. The span of the series begins in 1973 extending until November of 1995. For this duration, at least 24 houses were sold from month to month reaching a maximum of 89 homes sold. On average, approximately 52 were sold, and the median was 53 homes sold. *Figure 1* shows a visual representation of the data, and it is immediately apparent that any rises in sales is immediately followed by a drop in sales, or that strong positive correlation is followed by negative correlation. It appears to be an ongoing pattern. The two main goals are to find any significant seasonal effects on sales and to determine a reasonable one-year forecast.

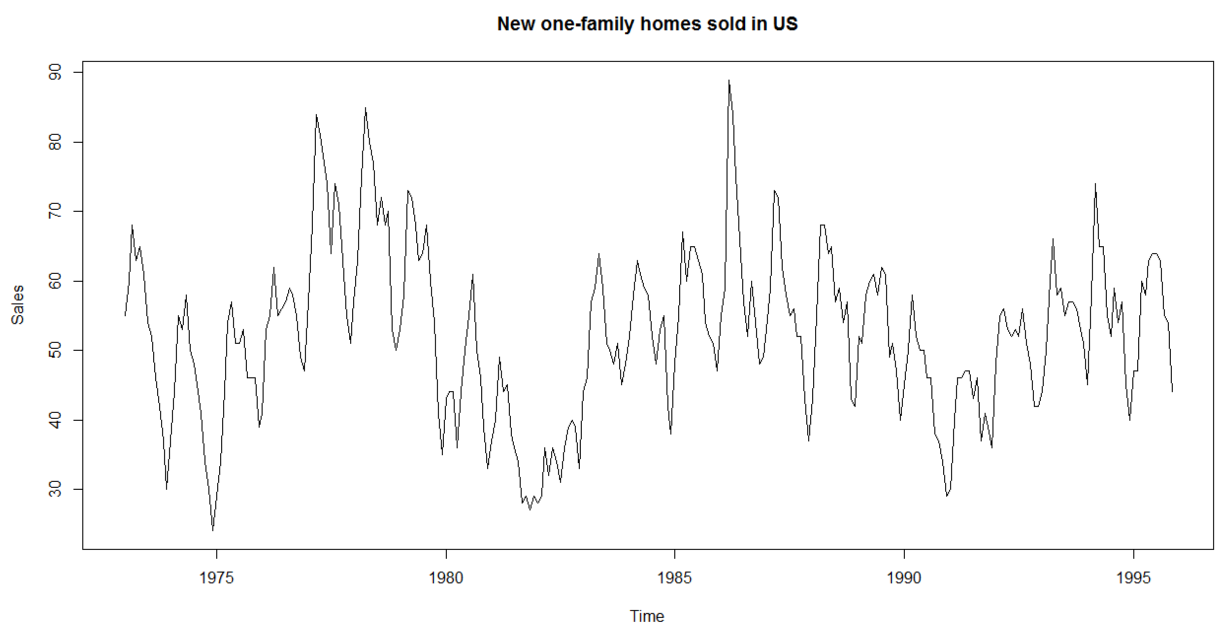

Figure 1

Examining Periodic Effects. The first step to finding any seasonal effects for home sales would be to locate patterns that stand out and determine if they are significant or not. This time series is composed of several different periodic effects, as shown by its spectral density in *Figure 2a*. Specifically, there are three cycles present by observing the ordinates computed exactly as an 8-year cycle, a 1-year cycle, and a 6-month cycle. Returning to *Figure 1*, starting at 1975 and following the graph to around 1983 and again to 1991 does show an 8-year cyclical pattern. Similarly, the aforementioned rising and falling of sales indicates that there is a 1-year cycle occurring. The 6-month pattern indicates an effect that happens somewhere during the year. Each of these periodic components were tested for significance, shown in *Figure 2b-d*. A 95% confidence interval was constructed for each of the peaks to find if most of the remaining frequencies fall below it, and the blue line represents the lower bound for each frequency. Both the 8 and 1-year cycle show obvious significance, however the 6-month effect is somewhat difficult to tell just by looking at the graph. Approximately 94% of the other frequencies fall below this line, so it reasonable to say that the 6-month cycle is also significant and should be considered in the overall assessment.

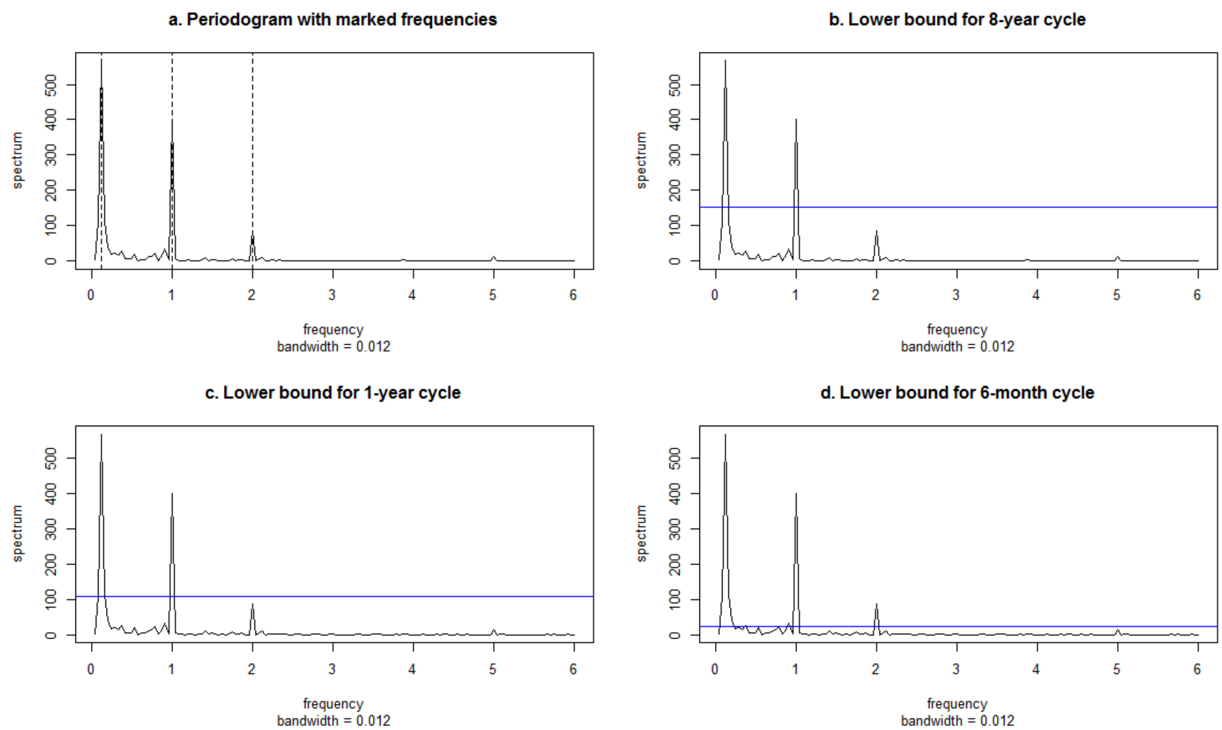

Figure 2 a. The dashed lines indicate the frequencies with an effect on sales. b. From a 95% C.I., the lower bound for the 8-year cycle. c. From a 95% C.I., the lower bound for a 1-year cycle. d. From a 95% C.I., the lower bound for a 6-month cycle. About 94% of the ordinates fell below this line.

Modelling. The best fit model was determined to be ARIMA(1,0,0) X $(4,1,1)_{12}$. The diagnostics in *Figure 3* show that it fits very well, has only a few outliers in the data, and there are no significant lags. In addition, these plots indicate that after the model is fitted, the only remaining variability in the time series is due to white noise.

It is interesting to note that there is a strong seasonal association between the previous four time points and a current time point for a 12-month period. There is also a first-order seasonal moving average component, indicating that the white noise is smoothed out by a previous time point.

Since causality for the model is upheld, and since it was a good fit for the time series, reasonable forecasting can now be enacted using the data.

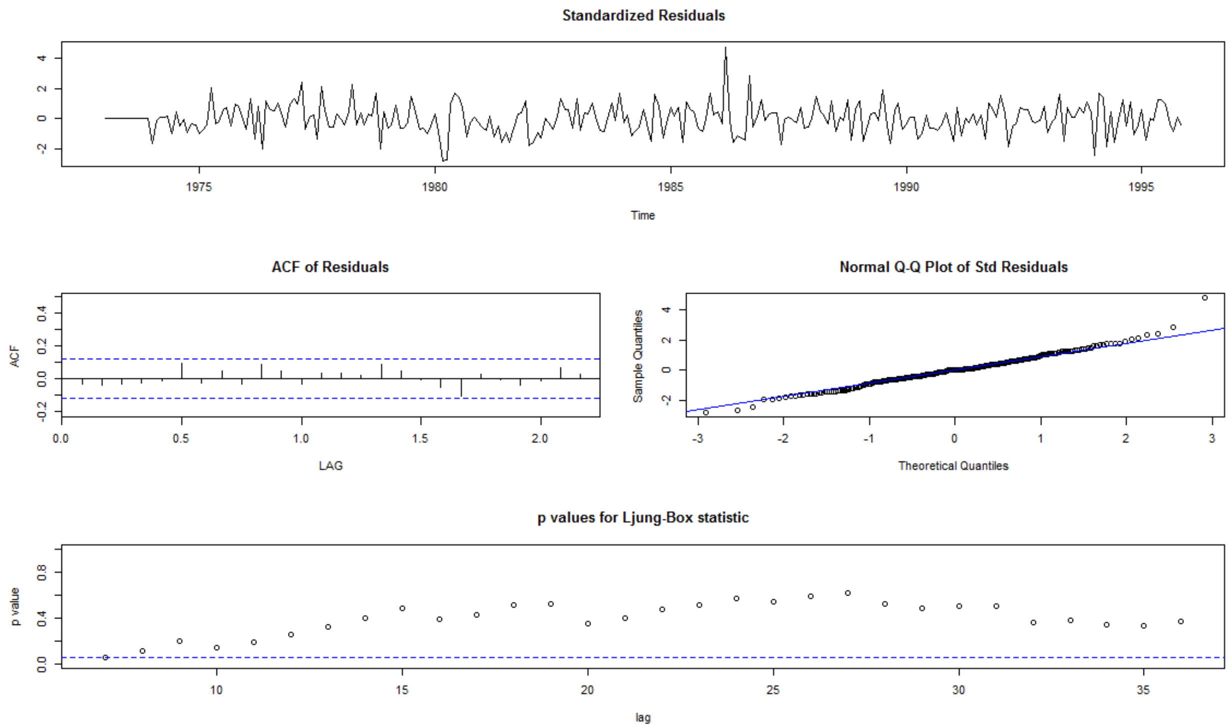

Figure 3

Forecasting. Using the fitted ARIMA model, a 1-year forecast was calculated from December of 1995 to November of 1996. Since with each additional time point they will tend more and more to the average home sales, projecting only a year seemed most reasonable. It also is most pertinent due to the fact that the immediate proceeding year is the market that would matter most. The following table shows the forecasted values for each month while also including their 95% prediction intervals. For brevity, home sales were rounded to the nearest sale.

Month	Sales	95% P.I	Month	Sales	95% P.I
12 (1995)	40	(36,44)	6	58	(50,67)
1 (1996)	45	(40,51)	7	55	(47,64)
2	50	(44,57)	8	56	(47,64)
3	62	(55,69)	9	51	(42.60)
4	60	(53,68)	10	50	(41,59)
5	61	(53,69)	11	40	(34,52)

These values are included in *Figure 4*, which displays a portion of the end of the time series followed by the forecasts, in blue. The red lines indicate the prediction interval which grows wider in time, a result of higher variability with further time points.

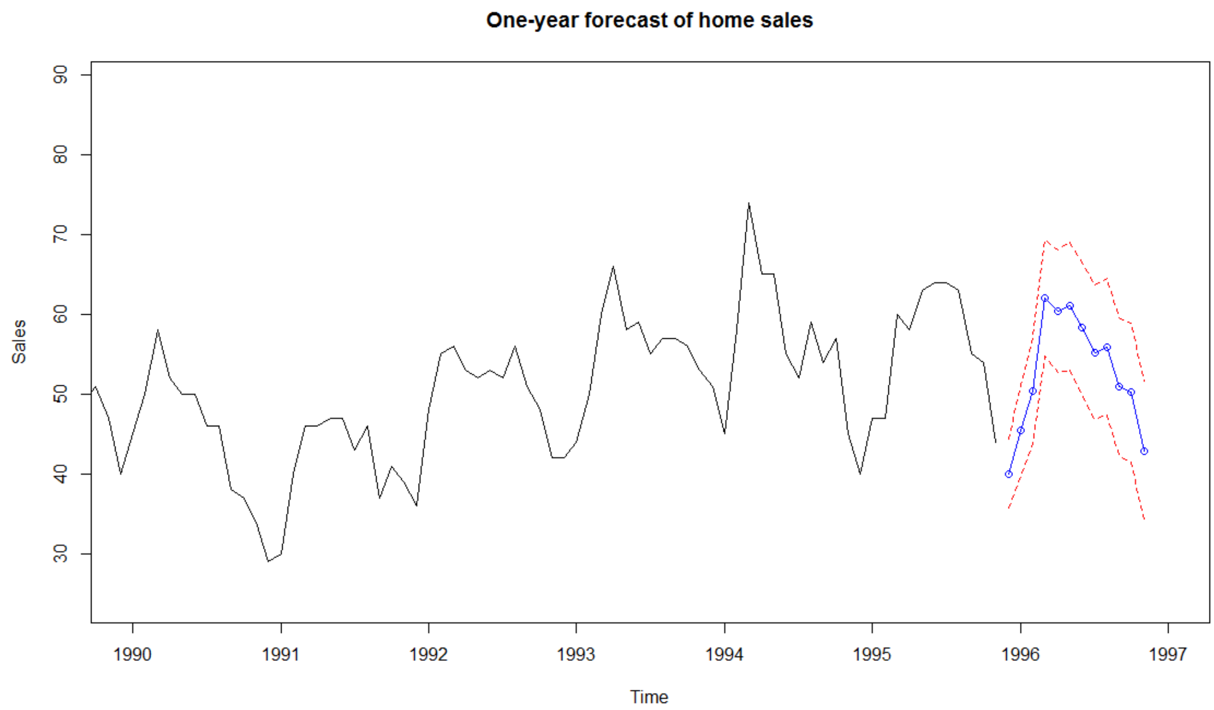

Figure 4

Conclusions. With the three periodic components (found to be significant) present in the time series, they can help describe where in the cycle the time series at any given point is, as well as what to expect for the following year. The 8-year effect shows that following a local minimum, a peak is eventually reached, only to be followed by another local minimum returning to the cycle. In the current cycle, that peak should have already occurred leading to a downward trend that is expected to continue until 1999.

The 1-year effect sees that there is always a rise and fall of home sales cycle that is exhibited annually. When this is coupled with the 6-month effect, there appears to be a steep rise until about the month of March, then gradually falls until January of the next year. The time series leaves off on November, and so the subsequent forecasted point is not surprisingly the lowest point of sales in 1995.

Going into 1996, the months of January through May seem to be a reasonable representation of what to expect sales to be in that time. However there is a large degree of uncertainty as to what the peak sales will be in March. This same notion holds as the sales decrease later in the year. The general shape is the same, and at the very least the high points and low points of the year can be easily anticipated. March is the peak month, and indeed the absolute maximum in the time series occurred in this month. From there, similarly high sales are expected which will then drop off after June.

The prediction intervals may give some hope as to having higher sales in 1996 than in 1995. However, considering the 8-year cycle, the highest yearly peak has already been reached in the current one, and all following yearly peaks should be moving downward until 1999. Therefore, the chance of this happening is unlikely. This shows the importance of including all periodic effects during assessment, as the 1-year cycle indicates that there will always be a high point in the year, but the 8-year cycle helps show where that peak will be. If anything, the highest yearly peak in the current 8-year cycle is well below that of the previous two which may mean there could be a slight leveling off period about to occur. Forecasting that far would not give accurate results, so it is better to consider only the next year.

Appendix:

```
## Time Series Final ##

#monthly data of new one family home sales
homes <- read.csv("monthly-sales-of-new-onefamily-h.csv")

#the last missing value was a result of some junk at the end and can
#be filtered out
homes <- homes[-nrow(homes),]

#time series begins the first month of 1973
homesales <- ts(homes[,2],freq=12,start=c(1973,1))

#some summary statistics
summary(homesales);min(time(homesales));max(time(homesales))

#output
plot.ts(homesales,ylab="Sales",main="New one-family homes sold in US")

##SARIMA model
homesales.fit <- arima(homesales,order=c(1,0,0),seasonal=list(order=c(4,1,1),
                       period=12))

#diagnostics
sarima(homesales,1,0,0,4,1,1,12)

##forecasting next year
homesales.predicted <- predict(homesales.fit,n.ahead=12)
U <- homesales.predicted$pred + homesales.predicted$se
L <- homesales.predicted$pred - homesales.predicted$se

#output
plot.ts(homesales,xlim=c(1990,1997),ylab="Sales",
        main="One-year forecast of home sales")
lines(homesales.predicted$pred,type="p",col="blue")
lines(homesales.predicted$pred,type="l",col="blue")
lines(U,lty="dashed",col="red")
lines(L,lty="dashed",col="red")

##observing frequencies
homesales.pgram <- spec.pgram(homesales,taper=0,log="no")
spec.desc <- sort(homesales.pgram$spec,decreasing=T)
homesales.pgram$freq[homesales.pgram$spec==spec.desc[1]]
homesales.pgram$freq[homesales.pgram$spec==spec.desc[2]]
homesales.pgram$freq[homesales.pgram$spec==spec.desc[5]]

U <- qchisq(0.025,2)
L <- qchisq(0.975,2)
#8-year cycle
c(2*spec.desc[1]/L,2*spec.desc[1]/U)
#1-year cycle
```

```
c(2*spec.desc[2]/L,2*spec.desc[2]/U)
#6-month cycle
c(2*spec.desc[5]/L,2*spec.desc[5]/U)
#proportion of frequencies falling below lower bound of 6-month cycle
sum(2*spec.desc[5]/L > spec.desc)/length(spec.desc)
#output
par(mfrow=c(2,2))
spec.pgram(homesales,taper=0,log="no",
           main="a. Periodogram with marked frequencies")
abline(v=0.125,lty="dashed");abline(v=1,lty="dashed");abline(v=2,lty="dashed"
)
spec.pgram(homesales,taper=0,log="no",
           main="b. Lower bound for 8-year cycle")
abline(h=2*spec.desc[1]/L,col="blue")
spec.pgram(homesales,taper=0,log="no",
           main="c. Lower bound for 1-year cycle")
abline(h=2*spec.desc[2]/L,col="blue")
spec.pgram(homesales,taper=0,log="no",
           main="d. Lower bound for 6-month cycle")
abline(h=2*spec.desc[5]/L,col="blue")
```