The annual retail trades survey (ARTS) produces a national estimate of total annual sales, e-commerce sales, end of year, inventories, inventory-to-sales ratios, purchases, total operating expenses, inventories held outside the United States, gross margins, and end-of-year accounts receivables for retail businesses and annual sales and e-commerce sales for accommodation and food service firms located in the U.S.

Figure 1 (Right Click to Zoom in)

.

Tables like the one in Figure 1 are created with data from the census in 2012 and estimated to 2014. This is an example of how we can use predictions and time-series analysis to predict 2 years into the future.

By looking at the per capita totals, we can generate a prediction about the time series data into the following two years using an autoregressive integrated moving average model.

This is the idea about why we use time series analysis, in the post I’ll take a look at some more financial data and then lightly touch on the theory behind time series.

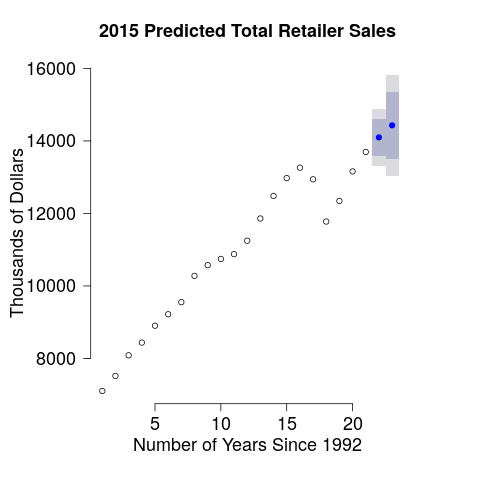


Figure 2 (Line 78 in per\_cap.R)

.

Timer Series Analysis’ is formulated by Y = a + bX, and it plots x and y, over some period. The analysis is interesting because in Figure 2 just before the 20th-month mark you can see where the 2009 US economic recession hit. It’s also useful for predictions and estimates. Time series is also helpful in stock trading, and spare parts are planning.

Time Series Analysis occurs in two stages, identify and model the structure of the time series, and forecast future values in the time series.

Identification is made through Box-Jenkins Methodology, (named after the two developers). The components we’re looking for are tended, seasonality, cyclic, and randomness to variations in the graph or data. The methodology consists of (1) conditioning the data and selecting a model, (a) identify the components, (b) re-examine. Then (2) estimate the model parameters, and (3) assesses the model and repeat until conditioned.

The ARIMA model is the one we use here, (that was mentioned before as autoregressive integrated moving average). To get the data predictions accurate we want to feed it a model that’s been conditioned to have a steady average, that doesn’t have a bunch of varying lines, and that the previous points look like they line up, so we create a stationary time series as in figure 3. It’s easier to think of number 2 looking like Figure 3 and that’s how we make the predictions because it’s just more steady

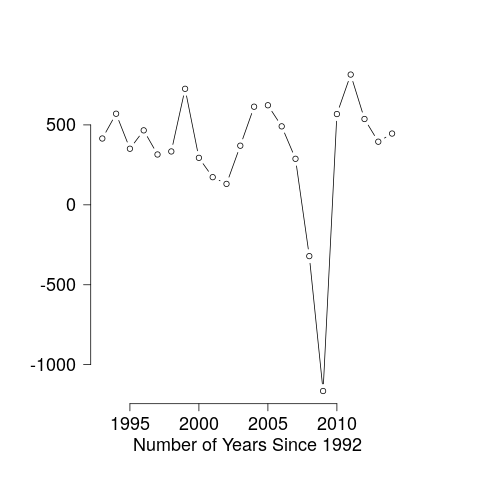


Figure 3 (Line 94 in per\_cap.R)

Let’s look at unemployment benefits in Australia for a moment to recap that idea.

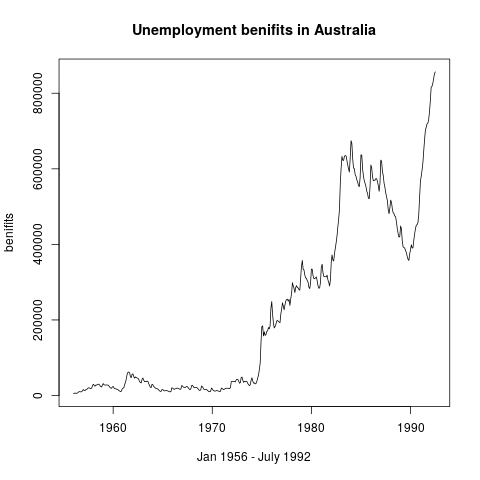


Figure 3 (From timeseries.R)

In this time series, we can note some interesting things; our initial time series shows an economic recession from 1974 to 1975 seen in Figure 4 with a 5% increase with long-term fluctuations between 1977 to 1979.

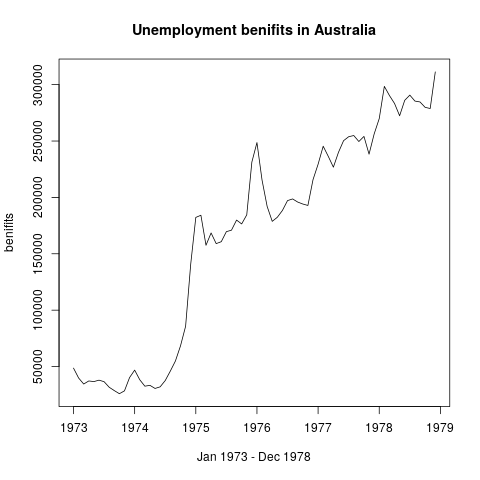


Figure 4 (From timeseries.R)

Figure 4 shows this range as in Figure 5.

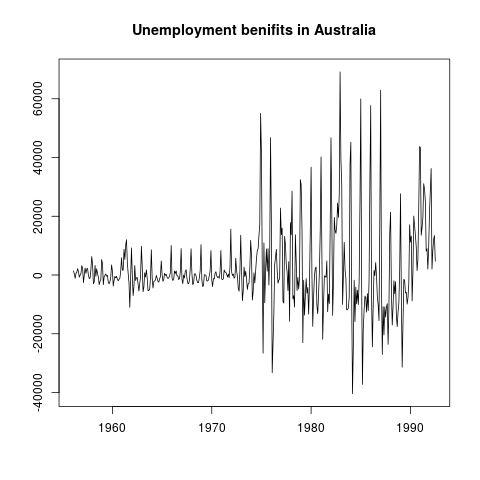


Figure 5 (From timeseries.R)

The stationary version of the graph looks just the same but with a steady average, as seen in Figure 5, notice how the same 1975 mark and 1979 mark appear alike Figure 4.

The next step is to condition the data a few more times, iteratively transforming the data set using the autocorrelation function and partial autocorrelation function we can detect insignificant changes throughout the time series and smooth them out.

While there are several ways to do this the easiest way to visualize it for this explanation is the HoltWinters method as shown a red line. Notice how the line is on average constant throughout, with less variability, and each point previous is close to the next point in Figure 6.

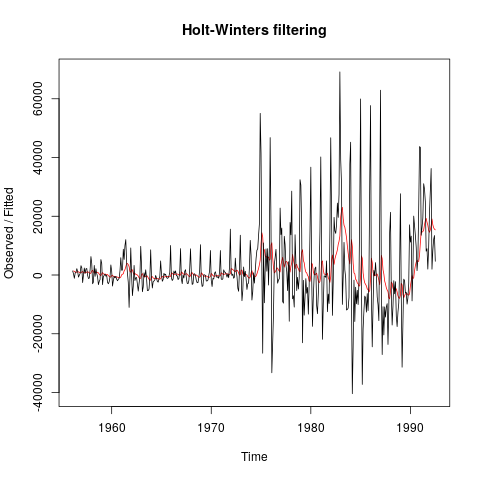


Figure 6 (From timeseries.R)

Since the red line is stationary, (it looks less crumpled than), we can use it to predict a linear trend. The final step is forecasting in which we take the data and create a prediction with a lower bound and an upper bound of where we might see expected levels to go.

For the forecasting, we can quickly look at any data set on the Quandl website, and I’ll be looking at Retailer Sales again from the Federal Reserve for the Jan 1992 to Sep 2017 and predict the next two years.

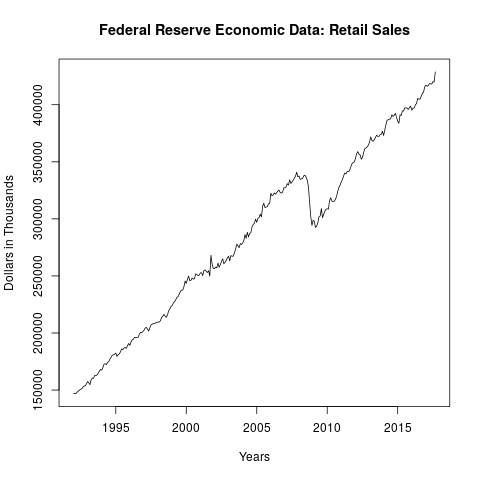


Figure 7 (Line 8 in timeseries2.R)

We can take a look at the federal reserve time series data; it should look familiar to Figure 2.

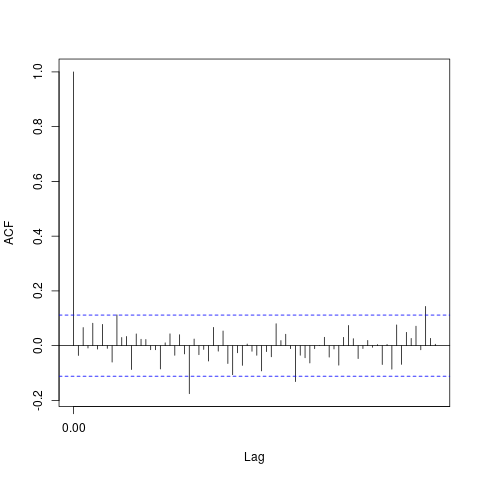


Figure 8

Figure 8 shows an example of lag points that cross over the significant level, which require smoothing, but would skew any forecasting and predictions we would make with the model we’re using.

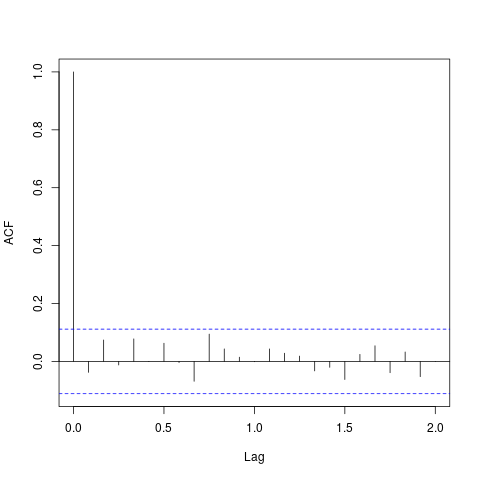


Figure 9 (Lines 34 in timeseries2.R)

So far we haven’t looked at the autocorrelation function (ACF) or the partial autocorrelation function (PACF), and they aren’t relevant for merely looking at the forecasting, but to touch on them briefly. To compare Figure 8 and Figure 9, you’ll notice that the two blue lines are crossed at specific lag lines. The blue lines represent significant levels of where the time series data needs to be smoothed for a proper prediction. In Figure 9 you’ll notice that they appear within the blue lines meaning that the model it would be good to use for forecasting.

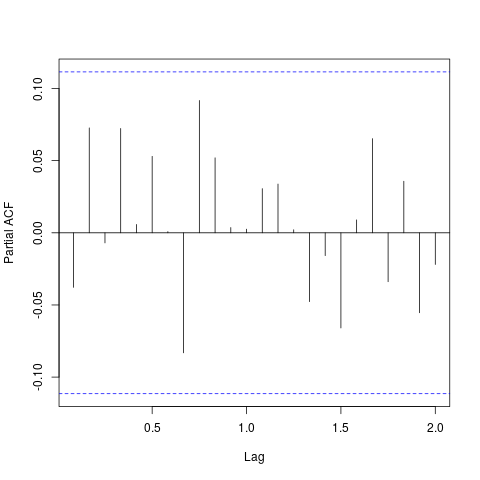


Figure 10 (Lines 34 in timeseries2.R)

The PACF is much like the ACF but degrades quicker (degradation not shown in Figure 10), and in Figure 10 we can see that the model is at significant levels.

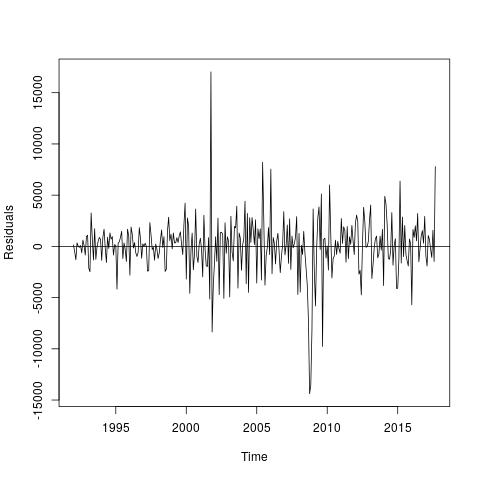


Figure 11 (Lines 43 in timeseries2.R)

Notice that in near 2002 there was a steady increase and at 2009 a steady decline, this is representative of an economic rise and an economic downturn. We can note that around 2002 the US economy experienced a spike and also in 2009 a recession.

Figure 11 shows the stationary graph for the time series in Figure 7, and at 0 you may notice a line, this represents a constant average, the chart also displays a reasonably constant variance, and nearly all points correlate with the previous scores, (this is called covariance).

This is perfect; we now know this will provide a steady prediction.

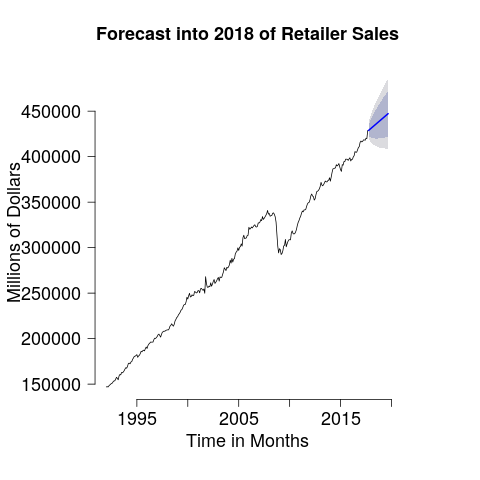


Figure 12 (Lines 62 in timeseries2.R)

So for a reason, we all came to the show; the forecast predicts with 95% confidence that there will be an increase between 4.6% and 13.3%.

That’s time series forecasting, take a look at the links below for the R scripts used.