



Forecasting & Time Series Analysis – Manufacturing Case Study Example (Part 1) (<http://ucanalytics.com/blogs/forecasting-time-series-analysis-manufacturing-case-study-example-part-1/>)

29th April, 2015 (<http://ucanalytics.com/blogs/forecasting-time-series-analysis-manufacturing-case-study-example-part-1/>) · Roopam Upadhyay (<http://ucanalytics.com/blogs/author/roopam/>) 15 Comments (<http://ucanalytics.com/blogs/forecasting-time-series-analysis-manufacturing-case-study-example-part-1/#comments>)

Today we are starting a new case study example series on YOU CANalytics involving forecasting and time series analysis. In this case study example, we will learn about time series analysis for a manufacturing operation. Time series analysis and modeling have many business and social applications. It is extensively used to forecast company sales, product demand, stock market trends, agricultural production etc. Before we learn more about forecasting let's evaluate our own lives on a time scale:

Life is a Sine Wave

I learnt a valuable lesson in life when I started my doctoral research in physics & nano-technology. I always loved physics, but during my doctoral studies, I was not enjoying the aspect of spending all my time in an isolated lab performing one experiment after another. Doing laboratory research could be extremely lonely. Additionally, I always enjoyed solving more applied and practical problems which I believed was missing in my research work. After getting frustrated for some time I decided to take some career advice from a trusted physicist friend. Before you read further, I must warn you that physicists as a community are usually mathematical, and occasionally philosophical. Physicists prefer to create a simple mathematical model about a complicated situation. They slowly add complexity to this simple model to make it fit with reality. The following is the key point I discovered during that conversation with my friend.

A simple model for life is a sine wave – where we go through ups and downs of moods and circumstances. Like a sine wave, we don't spend much of our time either on the peaks or the troughs but most of our time is spent climbing up or sliding down. Now keeping these moods and circumstances cycle in mind, a perfect choice of career is where one could enjoy both climbs and slides – as the up and down cycle is inevitable in life.

Keeping the above in mind I prepared a list of keywords that I associated with a job that I can truly love to absorb the up and down cycle of life. The following is my list of keywords:

Practical problem-solving	Mathematics	Creativity
Working with people on smart solutions	Scientific investigation	Learning every day
Seeing the fruits of my efforts reasonably fast	Producing quantifiable business benefits	Knowledge sharing

This prompted me to change my career from laboratory research to data science and business consulting. I am lucky that my career in data science and business analytics for over a decade has allowed me to check mark all these keywords.



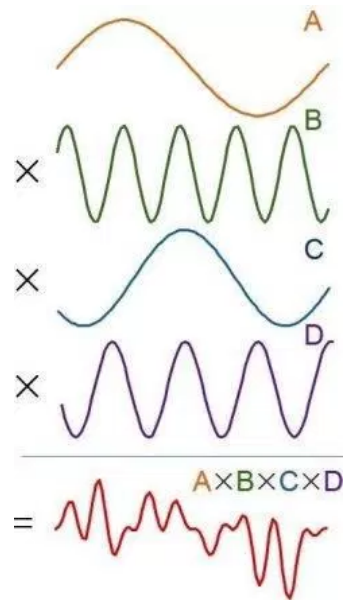
Your work is going to fill a large part of your life, and the only way to be truly satisfied is to do what you believe is great work. And the only way to do great work is to love what you do. If you haven't found it yet, keep looking. Don't settle. As with all matters of the heart, you'll know when you find it. And, like any great relationship, it just gets better and better as the year's roll on. So keep looking until you find it. Don't settle. – **Steve Jobs**

Interference of Other waves

Now in the true tradition of physics, let me add some complexity to the simple sine wave model for life. We live in a society and interact with many people. Everyone around us has a different shape to their lives' sine waves. The interference of different regular and predictable sine waves can produce patterns that are highly irregular and could at times be close to randomness.

This is what is displayed in the adjacent chart where the product of four harmonic sine waves is an irregular shape at the bottom. Eventually, our actual lives' function looks more like an irregular pattern produced through the interference of several sine waves.

On some level the above is the fundamental principle behind Fourier series and Fourier transforms; most engineering and physics students will get a cold chill of fear at the mention of Fourier series. However, the basic idea is simple that the linear combination of sine and cosine functions (similar to our lives' sine waves) can produce any complicated patterns including the irregular function we observed and much more complicated Fractals. I find it absolutely wonderful that a combination of harmonic motions can produce absolutely irregular patterns!



Time Series Analysis – Decomposition

(<https://i2.wp.com/ucanalytics.com/blogs/wp-content/uploads/2015/04/Time-Series-Analysis-Decomposition.jpg>)

Now, let me try to create a connection between what we discussed above with time series analysis and forecasting. The fundamental idea for time series analysis is to decompose the original time series (sales, stock market trends, etc.) into several independent components. Typically, business time series are divided into the following four components:

- **Trend** – overall direction of the series i.e. upwards, downwards etc.
- **Seasonality** – monthly or quarterly patterns
- **Cycle** – long-term business cycles
- **Irregular remainder** – random noise left after extraction of all the components



Interference of these components produces the final series.

Now the question is: why bother decomposing the original / actual time series into components? The answer: It is much easier to forecast the individual regular patterns produced through decomposition of time series than the actual series. This is similar to reproduction and forecasting the individual sine waves (A, B, C, and D) instead of the final irregular pattern produced through the product of these four sine waves.

Time Series Analysis – Manufacturing Case Study Example

(<https://i0.wp.com/ucanalytics.com/blogs/wp-content/uploads/2015/04/Tractor.jpg>) PowerHorse, a tractor and farm equipment manufacturing company, was established a few years after World War II. The company has shown a consistent growth in its revenue from tractor sales since its inception. However, over the years the company has struggled to keep its inventory and production cost down because of variability in sales and tractor demand. The management at PowerHorse is under enormous pressure from the shareholders and board to reduce the production cost. Additionally, they are also interested in understanding the impact of their marketing and farmer connect efforts towards overall sales. In the same effort, they have hired you as a data science and predictive analytics consultant.



You will start your investigation of this problem in the next part of this series using the concept discussed in this article. Eventually, you will develop an ARIMA model to forecast sale / demand for next year. Additionally, you will also investigate the impact of marketing program on sales by using an exogenous variable ARIMA model.

Sign (Sine) off Note

Whether you like it or not, life inevitably goes through up and down cycle. A perfect career or relationship doesn't make the variability disappear from our lives but makes us appreciate the swings of life. They keep us going in the tough times. They make us realise that variability is beautiful!



Time Series Decomposition – Manufacturing Case Study Example (Part 2) (<http://ucanalytics.com/blogs/time-series-decomposition-manufacturing-case-study-example-part-2/>)

24th May, 2015 (<http://ucanalytics.com/blogs/time-series-decomposition-manufacturing-case-study-example-part-2/>) · Roopam Upadhyay
(<http://ucanalytics.com/blogs/author/roopam/>) 11 Comments (<http://ucanalytics.com/blogs/time-series-decomposition-manufacturing-case-study-example-part-2/#comments>)

In the previous article, we started a new case study on sales forecasting for a tractor and farm equipment manufacturing company called PowerHorse. Our final goal is to forecast tractor sales in the next 36 months. In this article, we will delve deeper into time series decomposition. As discussed earlier (<http://ucanalytics.com/blogs/forecasting-time-series-analysis-manufacturing-case-study-part-1/>), the idea behind time series decomposition is to extract different regular patterns embedded in the observed time series. But in order to understand why this is easier said than done we need to understand some fundamental properties of mathematics and nature as an answer to the question:

Why is Your Bank Password Safe?

Mix one part of blue and one part of yellow to make 2 parts of green: a primary school art teacher writes this on the blackboard during a painting class. The students in the class then curiously try this trick and *Voilà!* they see green colour emerging from nowhere out of blue and yellow. One of the students after exhausting all her supplies of blue and yellow curiously asks the teacher: how can I extract the original yellow and blue from my two parts of green? This is where things get interesting, it is easy to mix things however it is really difficult (sometimes impossible) to reverse the process of mixing. The underlining principle at work over here is entropy (read the article on decision trees and entropy (<http://ucanalytics.com/blogs/decision-tree-entropy-retail-case-part-6/>)); reducing entropy (read randomness) requires a lot of work. This is essentially the reason why time series are difficult to decipher, and also the reason why your bank password is safe.

Cryptography, the science of hiding communication, is used to hide secrets such as bank passwords or credit card numbers and relies heavily on the above property of mixing being easier than “un-mixing”. When you share your credit card information on the internet it is available on the public domain for anybody to access. However, what makes it difficult for anyone without the key to use this information is the hard to decipher encryption. These encryptions at the fundamental level are created by multiplying 2 really large prime numbers. By the way, a prime number (aka prime) is a natural number greater than 1 that has no positive divisors other than 1 and itself. Now multiplication of two numbers, no matter how large, is a fairly straight forward process like mixing colours. On the other hand, reversing this process i.e. factorizing a product of two large primes could take hundreds of years for the fastest computer available on the planet. This is similar to “un-mixing” blue and yellow from green. You could learn more about cryptography and encryption by reading a fascinating book by Simon Singh called ‘The Code Book’.

Time Series Decomposition – Manufacturing Case Study Example

Back to our case study example, you are helping PowerHorse Tractors with sales forecasting (read part 1 (<http://ucanalytics.com/blogs/forecasting-time-series-analysis-manufacturing-case-study-part-1/>)). As a part of this project, one of the production units you are analysing is based in South East Asia. This unit is completely independent and caters to neighbouring geographies. This unit is just a decade and a half old. In 2014, they captured 11% of the market share, a 14% increase from the previous year. However, being a new unit they have very little bargaining power with their suppliers to implement Just-in-Time (JiT)

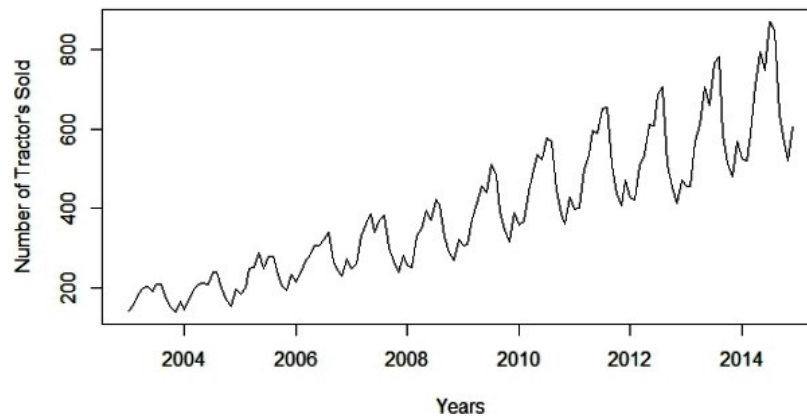


(<https://i0.wp.com/ucanalytics.com/blogs/wp-content/uploads/2015/05/Painter1.jpg>)
Painter and Time Series Decomposition – by Roopam



manufacturing principles that have worked really well in PowerHorse's base location. Hence, they want to be on top of their production planning to maintain healthy business margins. Monthly sales forecast is the first step you have suggested to this unit towards effective inventory management.

In the same effort, you asked the MIS team to share month on month (MoM) sales figures (number of tractors sold) for the last 12 years. The following is the time series plot for the same:



(<https://i1.wp.com/ucanalytics.com/blogs/wp-content/uploads/2015/05/Tractor-Sales.jpeg>)

Now you will start with time series decomposition of this data to understand underlying patterns for tractor sales. As discussed in the previous article, usually business time series are divided into the following four components:

- **Trend** – overall direction of the series i.e. upwards, downwards etc.
- **Seasonality** – monthly or quarterly patterns
- **Cycle** – long term business cycles
- **Irregular remainder** – random noise left after extraction biof all the components

In the above data, a cyclic pattern seems to be non-existent since the unit we are analysing is a relatively new unit to notice business cycles. Also in theory, business cycles in traditional businesses are observed over a period of 7 or more years. Hence, you won't include business cycles in this time series decomposition exercise. We will build our model based on the following function:

$$Y_t = f(\text{Trend}_t, \text{Seasonality}_t, \text{Remainder}_t)$$

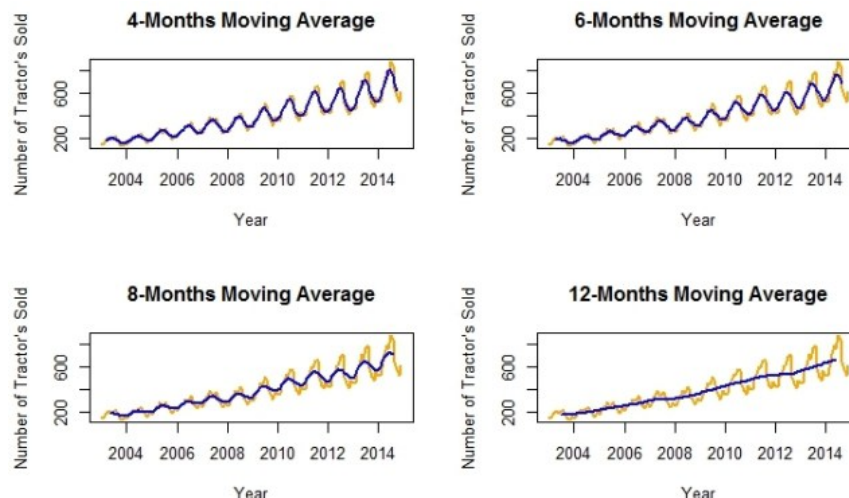
In the remaining article, we will study each of these components in some detail starting with trend.

Trend – Time Series Decomposition

Now, to begin with let's try to decipher trends embedded in the above tractor sales time series. One of the commonly used procedures to do so is moving averages. A good analogy for moving average is ironing clothes to remove wrinkles. The idea with moving average is to remove all the zigzag motion (wrinkles) from the time series to produce a steady trend through averaging adjacent values of a time period. Hence, the formula for moving average is:

$$\text{Moving Average} = \frac{\sum_{i=-m}^m Y_{t+i}}{2m+1}$$

Now, let's try to remove wrinkles from our time series using moving average. We will take moving average of different time periods i.e. 4,6,8, and 12 months as shown below. Here, moving average is shown in blue and actual series in orange.

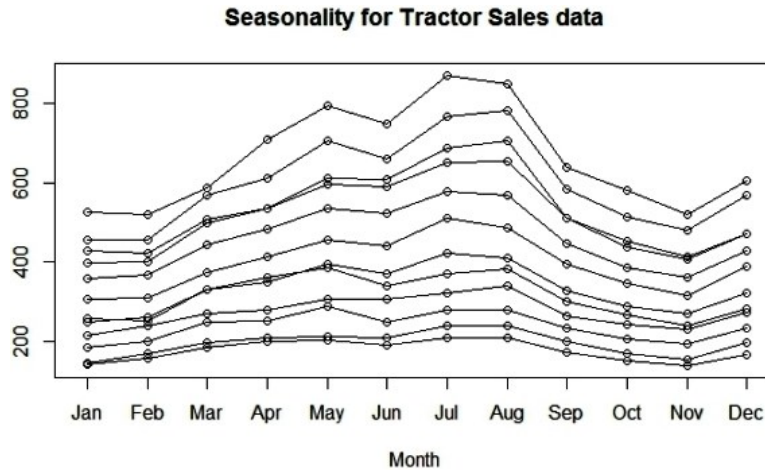


(<https://i2.wp.com/ucanalytics.com/blogs/wp-content/uploads/2015/05/Rplot.jpeg>)

As you could see in the above plots, 12-month moving average could produced a wrinkle free curve as desired. This on some level is expected since we are using month-wise data for our analysis and there is expected monthly-seasonal effect in our data. Now, let's decipher the seasonal component

Seasonality – Time Series Decomposition

The first thing to do is to see how number of tractors sold vary on a month on month basis. We will plot a stacked annual plot to observe seasonality in our data. As you could see there is a fairly consistent month on month variation with July and August as the peak months for tractor sales.



(<https://i1.wp.com/ucanalytics.com/blogs/wp-content/uploads/2015/05/Seasonal-Decomposition.jpeg>)

Irregular Remainder – Time Series Decomposition

To decipher underlying patterns in tractor sales, you build a multiplicative time series decomposition model with the following equation

Instead of multiplicative model you could have chosen additive model as well. However, it would have made very little difference in terms of conclusion you will draw from this time series decomposition exercise. Additionally, you are also aware that plain vanilla decomposition models like these are rarely used for forecasting. Their primary purpose is to understand underlying patterns in temporal data to use in more sophisticated analysis like Holt-Winters seasonal method or ARIMA.

(<https://i1.wp.com/ucanalytics.com/blogs/wp-content/uploads/2015/05/Time-Series-Decomposition-Plot.jpeg>)

The following are some of your key observations from this analysis:

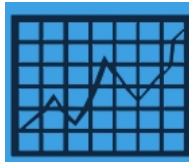
- 1) Trend:** 12-months moving average looks quite similar to a straight line hence you could have easily used linear regression to estimate the trend in this data.
- 2) Seasonality:** as discussed, seasonal plot displays a fairly consistent month-on-month pattern. The monthly seasonal components are average values for a month after removal of trend. Trend is removed from the time series using the following formula:

- 3) Irregular Remainder (random):** is the residual left in the series after removal of trend and seasonal components. Remainder is calculated using the following formula:

The expectations from remainder component is that it should look like a white noise i.e. displays no pattern at all. However, for our series residual display some pattern with high variation on the edges of data i.e. near the beginning (2004-07) and the end (2013-14) of the series.

White noise (randomness) has an important significance in time series modelling. In the later parts of this manufacturing case study, you will use ARIMA models to forecasts sales value. ARIMA modelling is an effort to make the remainder series display white noise patterns.

Sign-off Note



ARIMA Models – Manufacturing Case Study Example (Part 3) (<http://ucanalytics.com/blogs/arima-models-manufacturing-case-study-example-part-3/>)

7th June, 2015 (<http://ucanalytics.com/blogs/arima-models-manufacturing-case-study-example-part-3/>) · Roopam Upadhyay (<http://ucanalytics.com/blogs/author/roopam/>)
11 Comments (<http://ucanalytics.com/blogs/arima-models-manufacturing-case-study-example-part-3/#comments>)

For the last couple of articles, we are working on a manufacturing case study to forecast tractor sales for a company called PowerHorse. You can find the previous articles on the links **Part 1** (<http://ucanalytics.com/blogs/forecasting-time-series-analysis-manufacturing-case-study-example-part-1/>) and **Part 2** (<http://ucanalytics.com/blogs/time-series-decomposition-manufacturing-case-study-example-part-2/>). In this part, we will start with ARIMA modeling for forecasting. ARIMA is an abbreviation for *Auto-Regressive Integrated Moving Average*. However, before we learn more about ARIMA let's create a link between...

ARIMA and Sugar Cane Juice

May and June are the peak summer months in India. Indian summers are extremely hot and draining. Summers are followed by monsoon rains. It's no wonder that during summers everyone in India has the habit of looking up towards the sky in the hope to see clouds as an indicator of the arrival of monsoons. While waiting for the monsoons, Indians have a few drinks that keep them hydrated. Sugar cane juice, or *ganne-ka-ras*, is by far my favorite drink to beat the heat. The process of making sugar cane juice is fascinating and has similarities with ARIMA modeling.

(<https://i2.wp.com/ucanalytics.com/blogs/wp-content/uploads/2015/05/Sugarcane-Juicer.jpg>) Sugar cane juice is prepared by crushing a long piece of sugar cane through the juicer with two large cylindrical rollers as shown in the adjacent picture. However, it is difficult to extract all the juice from a tough sugar cane in one go hence the process is repeated multiple times. In the first go, a fresh sugar cane is passed through the juicer and then the residual of the sugar cane that still contains juice is again passed through the juicer many times till there is no more juice left in the residual. This is precisely how ARIMA models work. Consider your time series data as a sugar cane, and ARIMA models as sugar cane juicers. The idea with ARIMA models is that the final residual should look like white noise otherwise, there is juice or information available in the data to extract.



(<https://i0.wp.com/ucanalytics.com/blogs/wp-content/uploads/2015/05/Sugarcane-Juice.jpg>)

Sugar Cane Juice and ARIMA – by Roopam

We will come back to white noise (juice-less residual) in the latter sections of this article. However, before that let's explore more about ARIMA modeling.

ARIMA Modeling

ARIMA is a combination of 3 parts i.e. AR (*AutoRegressive*), I (*Integrated*), and MA (*Moving Average*). A convenient notation for ARIMA model is ARIMA(p,d,q). Here p,d, and q are the levels for each of the AR, I, and MA parts. Each of these three parts is an effort to make the final residuals display a white noise pattern (or no pattern at all). In each step of ARIMA modeling, time series data is passed through these 3 parts like a sugar cane through a sugar cane juicer to produce juice-less residual. The sequence of three passes for ARIMA analysis is as following:

1st Pass of ARIMA to Extract Juice / Information

Integrated (I) – subtract time series with its lagged series to extract trends from the data

In this pass of ARIMA juicer, we extract trend(s) from the original time series data. Differencing is one of the most commonly used mechanisms for extraction of trends. Here, the original series is subtracted with it's lagged series e.g. November's sales values are subtracted with October's values to produce trend-less residual series. The formulae for different orders of differencing are as follow:

No Differencing (d=0)	$Y'_t = Y_t$
1st Differencing (d=1)	$Y'_t = Y_t - Y_{t-1}$
2nd Differencing (d=2)	$Y'_t = Y_t - Y_{t-1} - (Y_{t-1} - Y_{t-2}) = Y_t - 2 \times Y_{t-1} + Y_{t-2}$

For example, in the adjacent plot a time series data with a linearly upward trend is displayed. Just below this plot is the 1st order differenced plot for the same data. As you can notice after 1st order differencing, trend part of the series is extracted and the difference data (residual) does not display any trend.

The residual data of most time series usually become trend-less after the first order differencing which is represented as ARIMA(0,1,0). Notice, AR (p), and MA (q) values in this notation are 0 and the *integrated* (I) value has order one. If the residual series still has a trend it is further differenced and is called 2nd order differencing. This trend-less series is called stationary on mean series i.e. mean or average value for series does not change over time. We will come back to stationarity and discuss it in detail when we will create an ARIMA model for our tractor sales data in the next article.

2nd Pass of ARIMA to Extract Juice / Information

AutoRegressive (AR) – extract the influence of the previous periods' values on the current period

After the time series data is made stationary through the *integrated* (I) pass, the AR part of the ARIMA juicer gets activated. As the name auto-regression suggests, here we try to extract the influence of the values of previous periods on the current period e.g. the influence of the September and October's sales value on the November's sales. This is done through developing a regression model with the time lagged period values as independent or predictor variables. The general form of the equation for this regression model is shown below. You may want to read the following articles on regression modeling Article 1 (<http://ucanalytics.com/blogs/regression-mother-models-retail-case-part-9/>) and Article 2 (<http://ucanalytics.com/blogs/regression-model-retail-case-study-part-10/>).

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + e_t$$

AR model of order 1 i.e. p=1 or ARIMA(1,0,0) is represented by the following regression equation

3rd Pass of ARIMA to Extract Juice / Information

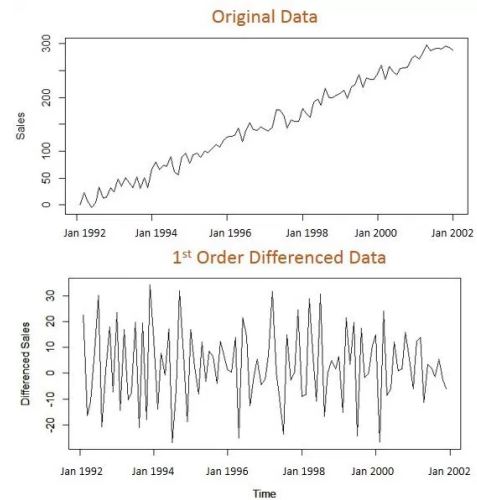
Moving Average (MA) – extract the influence of the previous period's error terms on the current period's error

Finally, the last component of ARIMA juicer i.e. MA involves finding relationships between the previous periods' error terms on the current period's error term. Keep in mind, this *moving average* (MA) has nothing to do with moving average we learned about in the previous article (<http://ucanalytics.com/blogs/time-series-decomposition-manufacturing-case-study-part-2/>) on time series decomposition. *Moving Average* (MA) part of ARIMA is developed with the following simple multiple linear regression values with the lagged error values as independent or predictor variables.

MA model of order 1 i.e. q=1 or ARIMA(0,0,1) is represented by the following regression equation

White Noise & ARIMA

Oh, how I miss the good old days when television was not on 24x7. For the good part of the day the TV used to look like the one shown in the picture – no signals just plain white noise. As a kid, it was a good pass time for my friends and me to keep looking at the TV with no signal to find patterns. White noise is a funny thing, if you look at it for long you will start seeing some false patterns. This is because the human brain is wired to find patterns, and at times confuses noises with signals. The biggest proof of this is how people lose money every day on the stock market. This is precisely the reason why we need a mathematical or logical process to distinguish between a white noise and a signal (juice / information). For example, consider the following simulated white noise:



(<https://i0.wp.com/ucanalytics.com/blogs/wp-content/uploads/2015/06/White-Nosie.jpeg>)

If you stare at the above graph for a reasonably long time you may start seeing some false patterns. A good way to distinguish between signal and noise is ACF (*AutoCorrelation Function*). This is developed by finding the correlation between a series of its lagged values. In the following ACF plot, you could see that for lag = 0 the ACF plot has the perfect correlation i.e. $\rho=1$. This makes sense because any data with itself will always have the perfect correlation. However as expected, our white noise doesn't have a significant correlation with its historic values ($\text{lag} \geq 1$). The dotted horizontal lines in the plot show the threshold for the insignificant region i.e. for a significant correlation the vertical bars should fall outside the horizontal dotted lines.

(<https://i0.wp.com/ucanalytics.com/blogs/wp-content/uploads/2015/06/ACF.jpeg>)

There is another measure *Partial AutoCorrelation Function* (PACF) that plays a crucial role in ARIMA modeling. We will discuss this in the next article when we will return to our manufacturing case study example.

Sign-off Note

In this article, you have spent your time learning things you will use in the next article while playing your role as a data science consultant to PowerHorse to forecast their tractor sales.


In the meantime, let me quickly check out of my window to see if there are any clouds out there..... Nope! I think there is still time before we will get our first monsoon showers in Bombay for this year – need to keep my glass of sugar cane juice handy to fight this summer.

Share this:

 LinkedIn (<http://ucanalytics.com/blogs/arima-models-manufacturing-case-study-example-part-3/?share=linkedin&nb=1>)

 Google (<http://ucanalytics.com/blogs/arima-models-manufacturing-case-study-example-part-3/?share=google-plus-1&nb=1>)

 Twitter (<http://ucanalytics.com/blogs/arima-models-manufacturing-case-study-example-part-3/?share=twitter&nb=1>)

 Facebook 6 (<http://ucanalytics.com/blogs/arima-models-manufacturing-case-study-example-part-3/?share=facebook&nb=1>)

 Email (<http://ucanalytics.com/blogs/arima-models-manufacturing-case-study-example-part-3/?share=email&nb=1>)

Related

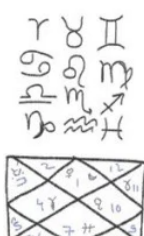


(<http://ucanalytics.com/blogs/how-effective-is-my-marketing-budget-regression-with-arima-errors-arimax-case-study-example-part-5/>)

How Effective is My Marketing Budget? - Regression with ARIMA Errors, Case Study Example (Part 5)

(<http://ucanalytics.com/blogs/how-effective-is-my-marketing-budget-regression-with-arima-errors-arimax-case-study-example-part-5/>)
July 12, 2015

In "Manufacturing Case Study Example"



(<http://ucanalytics.com/blogs/step-by-step-graphic-guide-to-forecasting-through-arima-modeling-in-r-manufacturing-case-study-example/>)

Step-by-Step Graphic Guide to Forecasting through ARIMA Modeling using R - Manufacturing Case Study Example (Part 4)
(<http://ucanalytics.com/blogs/step-by-step-graphic-guide-to-forecasting-through-arima-modeling-in-r-manufacturing-case-study-example/>)

June 28, 2015

In "Manufacturing Case Study Example"



(<http://ucanalytics.com/blogs/time-series-decomposition-manufacturing-case-study-example-part-2/>)

Time Series Decomposition - Manufacturing Case Study Example (Part 2)

(<http://ucanalytics.com/blogs/time-series-decomposition-manufacturing-case-study-example-part-2/>)

May 24, 2015

In "Manufacturing Case Study Example"

Posted in Manufacturing Case Study Example (<http://ucanalytics.com/blogs/category/manufacturing-case-study-example/>) |

« Time Series Decomposition – Manufacturing Case Study Example (Part 2) (<http://ucanalytics.com/blogs/time-series-decomposition-manufacturing-case-study-example-part-2/>)

Step-by-Step Graphic Guide to Forecasting through ARIMA Modeling using R – Manufacturing Case Study Example (Part 4) (<http://ucanalytics.com/blogs/step-by-step-graphic-guide-to-forecasting-through-arima-modeling-in-r-manufacturing-case-study-example/>) »



Step-by-Step Graphic Guide to Forecasting through ARIMA Modeling using R – Manufacturing Case Study Example (Part 4) (<http://ucanalytics.com/blogs/step-by-step-graphic-guide-to-forecasting-through-arima-modeling-in-r-manufacturing-case-study-example/>)

28th June, 2015 (<http://ucanalytics.com/blogs/step-by-step-graphic-guide-to-forecasting-through-arima-modeling-in-r-manufacturing-case-study-example/>) · Roopam Upadhyay (<http://ucanalytics.com/blogs/author/roopam/>) 116 Comments (<http://ucanalytics.com/blogs/step-by-step-graphic-guide-to-forecasting-through-arima-modeling-in-r-manufacturing-case-study-example/#comments>)

This article is a continuation of our manufacturing case study example to forecast tractor sales through time series and ARIMA models. You can find the previous parts at the following links:

Part 1 (<http://ucanalytics.com/blogs/forecasting-time-series-analysis-manufacturing-case-study-example-part-1/>): Introduction to time series modeling & forecasting

Part 2 (<http://ucanalytics.com/blogs/time-series-decomposition-manufacturing-case-study-example-part-2/>): Time series decomposition to decipher patterns and trends before forecasting

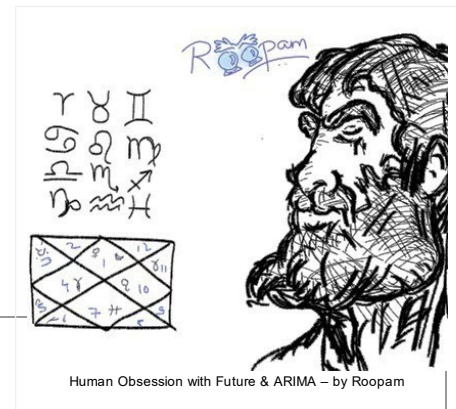
Part 3 (<http://ucanalytics.com/blogs/arima-models-manufacturing-case-study-example-part-3/>): Introduction to ARIMA models for forecasting

In this part, we will use plots and graphs to forecast tractor sales for PowerHorse tractors through ARIMA. We will use ARIMA modeling concepts learned in the previous article for our case study example. But before we start our analysis, let's have a quick discussion on forecasting:

Trouble with Nostradamus

Humans are obsessed about their future – so much so that they worry more about their future than enjoying the present. This is precisely the reason why horoscopists, soothsayers, and fortune tellers are always in high-demand. Michel de Nostredame (a.k.a Nostradamus) was a French soothsayer who lived in the 16th century. In his book *Les Propheties* (The Prophecies) he made predictions about important events to follow till the end of time. Nostradamus' followers believe that his predictions are irrevocably accurate about major events including the World Wars and the end of the world. For instance in one of the prophecies in his book, which later became one of his most debated and popular prophecies, he wrote the following

*"Beasts ferocious with hunger will cross the rivers
The greater part of the battlefield will be against Hister.
Into a cage of iron will the great one be drawn,
When the child of Germany observes nothing."*



His followers claim that *Hister* is an allusion to *Adolf Hitler* where Nostradamus misspelled Hitler's name. One of the conspicuous thing about Nostradamus' prophecies is that he never tagged these events to any date or time period. Detractors of Nostradamus believe that his book is full of cryptic pros (like the one above) and his followers try to force fit events to his writing. To dissuade detractors, one of his avid followers (based on his writing) predicted the month and the year for the end of the world as July 1999 – quite dramatic, isn't it? Ok so of course nothing earth-shattering happened in that month of 1999 otherwise you would not be reading this article. However, Nostradamus will continue to be a topic of discussion because of the eternal human obsession to predict the future.

Time series modelling and ARIMA forecasting are scientific ways to predict the future. However, you must keep in mind that these scientific techniques are also not immune to force fitting and human biases. On this note let us return to our manufacturing case study example.

ARIMA Model – Manufacturing Case Study Example

Back to our manufacturing case study example where you are helping PowerHorse Tractors with sales forecasting for them to manage their inventories and suppliers. The following sections in this article represent your analysis in the form of a graphic guide.

You could find the data shared by PowerHorse's MIS team at the following link Tractor Sales (<http://ucanalytics.com/blogs/wp-content/uploads/2015/06/Tractor-Sales.csv>). You may want to analyze this data to revalidate the analysis you will carry-out in the following sections.

Now you are ready to start with your analysis to forecast tractors sales for the next 3 years.

Step 1: Plot tractor sales data as time series

To begin with you have prepared a time series plot for the data. The following is the R code you have used to read the data in R and plot a time series chart.

```
1 data = read.csv('http://ucanalytics.com/blogs/wp-content/uploads/2015/06/Tractor-Sales.csv')
2 (http://ucanalytics.com/blogs/wp-content/uploads/2015/06/Tractor-Sales.csv')
3 data = ts(data[,2],start = c(2003,1),frequency = 12)
  plot(data, xlab='Years', ylab = 'Tractor Sales')
```

Clearly the above chart has an upward trend for tractors sales and there is also a seasonal component that we have already analyzed in an earlier article on time series decomposition (<http://ucanalytics.com/blogs/time-series-decomposition-manufacturing-case-study-example-part-2/>).

(<https://i1.wp.com/ucanalytics.com/blogs/wp-content/uploads/2015/06/1.-Tractor-Sales.jpeg>)

Step 2: Difference data to make data stationary on mean (remove trend)

The next thing to do is to make the series stationary as learned in the previous article (<http://ucanalytics.com/blogs/arima-models-manufacturing-case-study-example-part-3/>). This to remove the upward trend through 1st order differencing the series using the following formula:

1st Differencing ($d=1$)

The R code and output for plotting the differenced series are displayed below:

```
4 plot(diff(data),ylab='Differenced Tractor Sales')
```

Okay so the above series is not stationary on variance i.e. variation in the plot is increasing as we move towards the right of the chart. We need to make the series stationary on variance to produce reliable forecasts through ARIMA models.

(<https://i2.wp.com/ucanalytics.com/blogs/wp-content/uploads/2015/06/2-Differenced-Series.jpeg>)

Step 3: log transform data to make data stationary on variance

One of the best ways to make a series stationary on variance is through transforming the original series through log transform. We will go back to our original tractor sales series and log transform it to make it stationary on variance. The following equation represents the process of log transformation mathematically:

Log of sales

The following is the R code for the same with the output plot. Notice, this series is not stationary on mean since we are using the original data without differencing.

```
5 plot(log10(data),ylab='Log (Tractor Sales)')
```

Now the series looks stationary on variance.

(<https://i1.wp.com/ucanalytics.com/blogs/wp-content/uploads/2015/06/2.5-Log-Tractor-Sales.jpeg>)

Step 4: Difference log transform data to make data stationary on both mean and variance

Let us look at the differenced plot for log transformed series to reconfirm if the series is actually stationary on both mean and variance.

1st Differencing ($d=1$) of log of sales

The following is the R code to plot the above mathematical equation.

```
6 plot(diff(log10(data)),ylab='Differenced Log (Tractor Sales)')
```

Yes, now this series looks stationary on both mean and variance. This also gives us the clue that I or integrated part of our ARIMA model will be equal to 1 as 1st difference is making the series stationary.

(<https://i0.wp.com/ucanalytics.com/blogs/wp-content/uploads/2015/06/3-Diff-Log-Series.jpeg>)

Step 5: Plot ACF and PACF to identify potential AR and MA model

Now, let us create autocorrelation factor (ACF) and partial autocorrelation factor (PACF) plots to identify patterns in the above data which is stationary on both mean and variance. The idea is to identify presence of AR and MA components in the residuals. The following is the R code to produce ACF and PACF plots.

```

7 | par(mfrow = c(1,2))
8 | acf(ts(diff(log10(data))),main='ACF Tractor Sales')
9 | pacf(ts(diff(log10(data))),main='PACF Tractor Sales')

```

(<https://i0.wp.com/ucanalytics.com/blogs/wp-content/uploads/2015/06/5.-ACF-PACF-diff-logts.jpeg>)

Since, there are enough spikes in the plots outside the insignificant zone (dotted horizontal lines) we can conclude that the residuals are not random. This implies that there is juice or information available in residuals to be extracted by AR and MA models. Also, there is a seasonal component available in the residuals at the lag 12 (represented by spikes at lag 12). This makes sense since we are analyzing monthly data that tends to have seasonality of 12 months because of patterns in tractor sales.

Step 6: Identification of best fit ARIMA model

Auto arima function in forecast package in R helps us identify the best fit ARIMA model on the fly. The following is the code for the same. Please install the required 'forecast' package in R before executing this code.

```

10 | require(forecast)
11 | ARIMAfit = auto.arima(log10(data), approximation=FALSE,trace=FALSE)
12 | summary(ARIMAfit)

```

Time series:	log ₁₀ (Tractor Sales)	
Best fit Model:	ARIMA(0,1,1)(0,1,1)[12]	
	mal	sma1
Coefficients:	-0.4047	-0.5529
s.e.	0.0885	0.0734
log likelihood=354.4		
AIC=-702.79	AICc=-702.6	BIC=-694.17

The best fit model is selected based on Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC) values. The idea is to choose a model with minimum AIC and BIC values. We will explore more about AIC and BIC in the next article. The values of AIC and BIC for our best fit model developed in R are displayed at the bottom of the following results:

As expected, our model has I (or integrated) component equal to 1. This represents differencing of order 1. There is additional differencing of lag 12 in the above best fit model. Moreover, the best fit model has MA value of order 1. Also, there is seasonal MA with lag 12 of order 1.

Step 6: Forecast sales using the best fit ARIMA model

The next step is to predict tractor sales for next 3 years i.e. for 2015, 2016, and 2017 through the above model. The following R code does this job for us.

```

13 | par(mfrow = c(1,1))
14 | pred = predict(ARIMAfit, n.ahead = 36)
15 | pred
16 | plot(data,type='l',xlim=c(2004,2018),ylim=c(1,1600),xlab = 'Year',ylab = 'Tractor Sales')
17 | lines(10^(pred$pred),col='blue')
18 | lines(10^(pred$pred+2*pred$se),col='orange')
19 | lines(10^(pred$pred-2*pred$se),col='orange')

```

The following is the output with forecasted values of tractor sales in blue. Also, the range of expected error (i.e. 2 times standard deviation) is displayed with orange lines on either side of predicted blue line.

(<https://i2.wp.com/ucanalytics.com/blogs/wp-content/uploads/2015/06/7-ACF-PACF-residual.jpeg>)

Now, forecasts for a long period of 3 years is an ambitious task. The major assumption here is that the underlining patterns in the time series will continue to stay the same as predicted in the model. A short-term forecasting model, say a couple of business quarters or a year, is usually a good idea to forecast with reasonable accuracy. A long-term model like the one above needs to be evaluated on a regular interval of time (say 6 months). The idea is to incorporate the new information available with the passage of time in the model.

Step 7: Plot ACF and PACF for residuals of ARIMA model to ensure no more information is left for extraction

Finally, let's create an ACF and PACF plot of the residuals of our best fit ARIMA model i.e. ARIMA(0,1,1)(0,1,1)[12]. The following is the R code for the same.

```

20 | par(mfrow=c(1,2))
21 | acf(ts(ARIMAfit$residuals),main='ACF Residual')
22 | pacf(ts(ARIMAfit$residuals),main='PACF Residual')

```

(<http://ucanalytics.com/blogs/wp-content/uploads/2015/06/7-ACF-PACF-residual.jpeg>)

Since there are no spikes outside the insignificant zone for both ACF and PACF plots we can conclude that residuals are random with no information or juice in them. Hence our ARIMA model is working fine.

However, I must warn you before concluding this article that randomness is a funny thing and can be extremely confusing. We will discover this aspect about randomness and patterns in the epilogue of this forecasting case study example.

Sign-off Note


I must say Nostradamus was extremely clever since he had not tagged his prophecies to any time period. So he left the world with a book containing some cryptic sets of words to be analysed by the human imagination. This is where randomness becomes interesting. A prophesy written in cryptic words without a defined time-period is almost 100% likely to come true since humans are the perfect machine to make patterns out of randomness.

Let me put my own prophesy for a major event in the future. If someone will track this for the next 1000 years I am sure this will make me go in the books next to Nostradamus.

*A boy of strength will rise from the home of the poor
Will rule the world and have both strong friends and enemies
His presence will divide the world into half
The man of God will be the key figure in resolution of this conflict*

Share this:

 LinkedIn 360 (<http://ucanalytics.com/blogs/step-by-step-graphic-guide-to-forecasting-through-arima-modeling-in-r-manufacturing-case-study-example/>)
share=linkedin&nb=1)

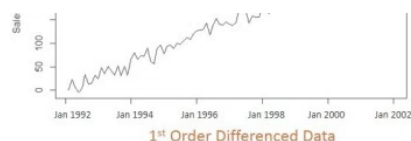
 Google (<http://ucanalytics.com/blogs/step-by-step-graphic-guide-to-forecasting-through-arima-modeling-in-r-manufacturing-case-study-example/?share=google-plus-1&nb=1>)

 Twitter (<http://ucanalytics.com/blogs/step-by-step-graphic-guide-to-forecasting-through-arima-modeling-in-r-manufacturing-case-study-example/?share=twitter&nb=1>)

 Facebook 9 (<http://ucanalytics.com/blogs/step-by-step-graphic-guide-to-forecasting-through-arima-modeling-in-r-manufacturing-case-study-example/>)
share=facebook&nb=1)

 Email (<http://ucanalytics.com/blogs/step-by-step-graphic-guide-to-forecasting-through-arima-modeling-in-r-manufacturing-case-study-example/?share=email&nb=1>)

Related



(<http://ucanalytics.com/blogs/arima-models-manufacturing-case-study-example-part-3/>)
ARIMA Models - Manufacturing Case Study Example (Part 3)
(<http://ucanalytics.com/blogs/arima-models-manufacturing-case-study-example-part-3/>)
June 7, 2015
In "Manufacturing Case Study Example"



(<http://ucanalytics.com/blogs/how-effective-is-my-marketing-budget-regression-with-arima-errors-arimax-case-study-example-part-5/>)
How Effective is My Marketing Budget? - Regression with ARIMA Errors, Case Study Example (Part 5)
(<http://ucanalytics.com/blogs/how-effective-is-my-marketing-budget-regression-with-arima-errors-arimax-case-study-example-part-5/>)
July 12, 2015
In "Manufacturing Case Study Example"



(<http://ucanalytics.com/blogs/time-series-decomposition-manufacturing-case-study-example-part-2/>)
Time Series Decomposition - Manufacturing Case Study Example (Part 2)
(<http://ucanalytics.com/blogs/time-series-decomposition-manufacturing-case-study-example-part-2/>)
May 24, 2015
In "Manufacturing Case Study Example"

Posted in Manufacturing Case Study Example (<http://ucanalytics.com/blogs/category/manufacturing-case-study-example/>) |

« ARIMA Models – Manufacturing Case Study Example (Part 3) (<http://ucanalytics.com/blogs/arima-models-manufacturing-case-study-example-part-3/>)

How Effective is My Marketing Budget? – Regression with ARIMA Errors, Case Study Example (Part 5)
(<http://ucanalytics.com/blogs/how-effective-is-my-marketing-budget-regression-with-arima-errors-arimax-case-study-example-part-5/>) »

Ram says:

June 28, 2015 at 9:43 pm (<http://ucanalytics.com/blogs/step-by-step-graphic-guide-to-forecasting-through-arma-modeling-in-r-manufacturing-case-study-example/#comment-1154>)

Hi Roopam,

Great article, very good explanation. Thank you.

Here is a suggestion to help the readers:.

As the data is not in a time-series format, it will help if we convert it to a time-series when plotting it.

Change the code in the following steps:

Step 1: `plot(ts(data[,2]), xlab="Years", ylab = "Tractor Sales")`

Step 2: `plot(diff(ts(data[,2])), xlab="Years", ylab = "Diff Tractor Sales")`

Step 3: `plot(log(ts(data[,2])), xlab="Years", ylab = "Log Tractor Sales")`

Keep up the great work.

Regards

Ram

Reply (<http://ucanalytics.com/blogs/step-by-step-graphic-guide-to-forecasting-through-arma-modeling-in-r-manufacturing-case-study-example/?replytocom=1154#respond>)

Roopam Upadhyay says:

June 29, 2015 at 11:49 am (<http://ucanalytics.com/blogs/step-by-step-graphic-guide-to-forecasting-through-arma-modeling-in-r-manufacturing-case-study-example/#comment-1155>)

Thanks Ram for sharing your code to convert data into time series data format. I did miss out on this, now have included an extra step in the beginning for the same. Appreciate your help!

Cheers

Roopam

Reply (<http://ucanalytics.com/blogs/step-by-step-graphic-guide-to-forecasting-through-arma-modeling-in-r-manufacturing-case-study-example/?replytocom=1155#respond>)

Ram says:

June 29, 2015 at 7:06 pm (<http://ucanalytics.com/blogs/step-by-step-graphic-guide-to-forecasting-through-arma-modeling-in-r-manufacturing-case-study-example/#comment-1156>)

Cool, Thank you bro.

Reply (<http://ucanalytics.com/blogs/step-by-step-graphic-guide-to-forecasting-through-arma-modeling-in-r-manufacturing-case-study-example/?replytocom=1156#respond>)

amber hanif says:

April 24, 2016 at 8:15 pm (<http://ucanalytics.com/blogs/step-by-step-graphic-guide-to-forecasting-through-arma-modeling-in-r-manufacturing-case-study-example/#comment-1478>)

hello im doing my thesis using arima(autoregressive integrated moving average) modelling techniqe but i dont know what should i do to study the arima . as it needed a practice study bt i just read it as theoraticle . please guide me in this case i'll be gratefull to u !!!!

Reply (<http://ucanalytics.com/blogs/step-by-step-graphic-guide-to-forecasting-through-arma-modeling-in-r-manufacturing-case-study-example/?replytocom=1478#respond>)

Roopam Upadhyay says:

March 7, 2017 at 8:44 am (<http://ucanalytics.com/blogs/step-by-step-graphic-guide-to-forecasting-through-arma-modeling-in-r-manufacturing-case-study-example/#comment-1680>)

Play around with your data – plot them and make naive predictions in the beginning and then figure out how to improve those predictions using other techniques (like ARIMA, neural networks etc.). I think that's the best way to practice and get a practical exposure – play with your data. All the best.

Reply (<http://ucanalytics.com/blogs/step-by-step-graphic-guide-to-forecasting-through-arma-modeling-in-r-manufacturing-case-study-example/?replytocom=1680#respond>)

Saurabh says:

July 29, 2016 at 5:43 pm (<http://ucanalytics.com/blogs/step-by-step-graphic-guide-to-forecasting-through-arma-modeling-in-r-manufacturing-case-study-example/#comment-1526>)

Roopam,

you have already converted into time series in your first step:

`data<-read.csv("[location of data]")`