

## False Seasonality Data Challenge. Accepted

Posted on Apr 08, Posted by [Tom Reilly](#) Category [Forecasting](#)

It's been 6 months since our last BLOG. We have been very busy.

We engaged in a debate on a [linkedin discussion group](#) over the need to pre-screen your data so that your forecasting algorithm can either apply seasonal models or not consider seasonal models. A set of GUARANTEED random data was generated and given to us as a challenge four years ago. This time we looked a little closer at the data and found something interesting. 1) you don't need to pre-screen your data 2) be careful how you generate random data

### Here is my first response:

As for your random data, we still have it when you sent it 4 years ago. I am not sure what you and Dave looked at, but if you download run the 30 day trial now and we always have kept improving the software you will get a different answer and the results posted here on dropbox.com. [https://www.dropbox.com/s/s63kxrkquzc6e00/output\\_miket.zip](https://www.dropbox.com/s/s63kxrkquzc6e00/output_miket.zip)

I have provided your data(xls file), our model equation (equ), forecasts(pro), graph(png) and audit of the model building process(htm).

Out of the 18 examples, Autobox found 6 with a flat forecast, 7 with 1 monthly seasonal pulse or a 1 month fixed effect, 4 with 2 months that had a mix of either a seasonal pulse or a 1 month fixed effect, 2 with 3 months that had a mix of either a seasonal pulses or a 1 month fixed effect.

Note that no model was found with Seasonal Differencing, AR12, with all 11 seasonal dummies.

Now, in a perfect world, Autobox would have found 19 flat lines based on this theoretical data. If you look at the data, you will see that there were patterns found where Autobox found them that make sense. There are sometimes seasonality that is not persistent and just a couple of months through the year.

If we review the 12 series where Autobox detected seasonality, it is very clear that in the 11 of the 12 cases that it was justified in doing so. That would make 17 of the 18 properly modeled and forecasted.

Series 1 - Autobox found feb to be low. A All three years this was the case. Let's call this a win.

Series 2 - Autobox found apr to be low. All three years were low. Let's that call this a win.

Series 3- Autobox found sep and oct to be low. 4 of the 6 were low and the four most recent were all low supporting a change in the seasonality. Let's call this a win.

Series 4- Autobox found nov to be low. All three years were low. Let's call this a win.

Series 5- Autobox found mar, may and aug to be low. All three years were low. Let's call that a win.

Series 7- Autobox found jun low and aug high. All three years matched the pattern. Let's call that a win.

Series 10 - Autobox found apr and jun to be high. 5 of the 6 data points were high. Let's call this a win.

Series 12 - Autobox found oct to be high and dec to be low. All three years this was the case. Let's call this a win.

Series 13 - Autobox found aug to be high. Two of the three years were very very high. Let's call this a win.

Series 14 - Autobox found feb and apr to be high. All three years this was the case. Let's call this a win.

Series 15 - Autobox found may jun to be high and oct low. 8 of the 9 historical data points support this, Let's call this a win.

Series 16 - Autobox found jan to below. It was very low for two, but one was quite high and Autobox called that an outlier. Let's call this a loss.

**A little sleep and then I posted this response:**

After sleeping on that very fun exercise, there was something that still wasn't sitting right with me. The "guaranteed" no seasonality statement didn't match with the graph of the datasets. They didn't seem to have randomness and seemed more to have some pattern.

I generated 4 example datasets from the link below. I used the defaults and graphed them. They exhibited randomness. I ran them through Autobox and all had zero seasonality and flat forecasts.

<http://www.random.org/sequences/>

Tags: Tagged in: [arima](#) [box-jenkins](#) [acf](#) [pacf](#) [cran](#) [ets](#) [auto.arima](#) [gretl](#) [rats](#) [SAS](#)  
[forecasting](#) [arimax](#) [outliers](#) [sas](#) [transfer function](#) [hyndman](#)  
[time series forecasting](#) [trends](#) [level shifts](#) [seasonality](#) [outliers](#)