Digging Deeper with Blast Motion Data: Autocorrelation

Wareham Gatemen Baseball

Zachariah Zanger M.S.

Motivation

This series of analysis (Digging Deeper with Blast Motion Data) is meant to apply methods and statistical frameworks used in Quantitative Finance and Statistics to interpret and analyze raw Blast Motion Data. The writing will be straight to the point with the intention of using primarily words, not numbers, in the text for explanation. I will bounce back and forth from Finance to Baseball. If a reader feels like the message is not getting across well, please tell me and I will adjust the writing style (and caffeine consumption while writing). I have Python and R code to backup all analysis, as well as spreadsheets to store information. Given the phenomenal time referenced indexing of Blast data, it would be foolish not to explore Time Series Analysis frameworks first. The first framework for analysis will be on the topic of Autocorrelation

Definition and Finance Application

For a given vector/list of a variable's values in a time series format, Autocorrelation measures the correlation of values with a lagged copy of the same values moving into the future. Intuitively speaking, *Autocorrelation measures the correlation of today's value to a future value for the same variable in n-lag time*. N-lag refers to the lag in time the user wants to examine correlation for within the vector of values. N-lag time follows a custom measure of time, for Blast data purposes, we would measure in swings. For financial data we measure in days, weeks, months, quarters, etc… .

Examples of N-Lag Time

- The quarterly Microsoft stock return in 2 quarter lag time = The quarterly Microsoft stock return 2 quarters from now
- The 1 swing plane score in 3 swing lag time = The swing plane score 3 swings from now
- The average 3 swing rotation score in 1 lag time = The average 3 swing rotation score in the next 3 swings

Autocorrelation is measured in between -1 and 1, where 1 would indicate a perfect correlation of a value today to the value in n-lag time. A -1 Autocorrelation would indicate a perfectly negative correlation of a value today to the value in n-lag time. Let's break down the concept of n-lag time with returns for Microsoft on a quarterly basis:

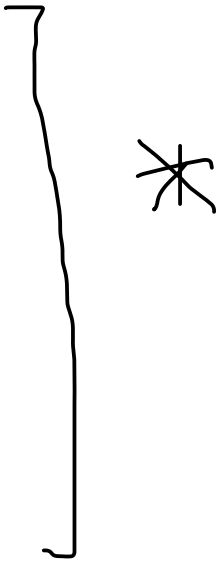
Vector of Microsoft Monthly Returns from 9/30/2012 – 10/31/2013

9/30/2012	-0.03439
10/31/2012	-0.04099
11/30/2012	-0.05961
12/31/2012	0.003381
1/31/2013	0.027705
2/28/2013	0.021135
3/31/2013	0.029137
4/30/2013	0.156938
5/31/2013	0.061774
6/30/2013	-0.01032
7/31/2013	-0.07817
8/31/2013	0.056387
9/30/2013	-0.00359

10/31/2013	0.064003

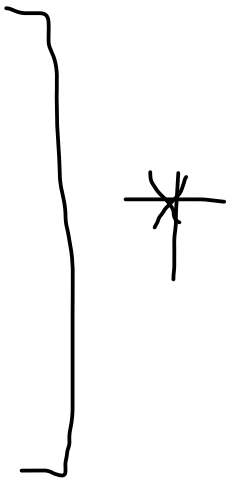
In the left column is the date of a monthly return. In the right column is the closing return for the corresponding date. We could say, "*I want to find the Autocorrelation of Microsoft quarterly prices in 1-lag time"*. We take the existing vector and shift it down by 1 lag.

1-Lag Time	
-0.03439	
-0.04099	-0.03439
-0.05961	-0.04099
0.003381	-0.05961
0.027705	0.003381
0.021135	0.027705
0.029137	0.021135
0.156938	0.029137
0.061774	0.156938
-0.01032	0.061774
-0.07817	-0.01032
0.056387	-0.07817
-0.00359	0.056387
0.064003	-0.00359
0.084981	0.064003



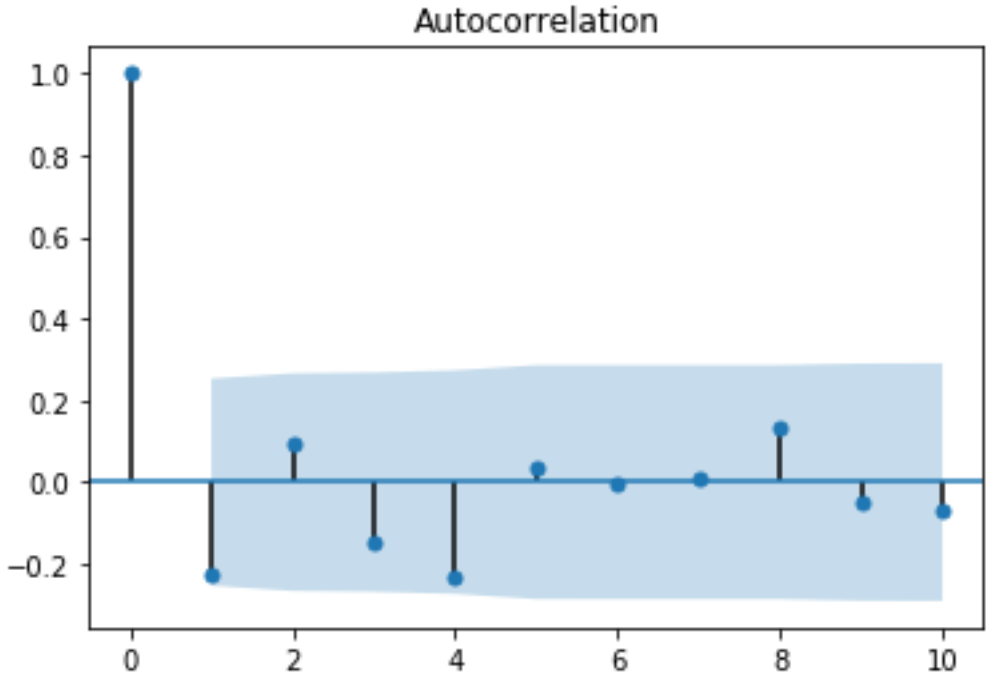
The computer calculates the correlation of observed values for different periods of time with the same values. If Microsoft had a true 95% Autocorrelation in 1-month lag time, the likelihood of getting a 1% monthly return 1 month from now after observing a 1% monthly return, are very good. This is what 3-lag time looks like:

3-Lag Time	
-0.03439	
-0.04099	
-0.05961	
0.003381	-0.03439
0.027705	-0.04099
0.021135	-0.05961
0.029137	0.003381
0.156938	0.027705
0.061774	0.021135
-0.01032	0.029137
-0.07817	0.156938
0.056387	0.061774
-0.00359	-0.01032
0.064003	-0.07817



0.084981	0.056387

Let's visually examine the Autocorrelations for Microsoft returns on a monthly basis in n-lag time in lags from 1 to 10:



What does this series of Autocorrelations tell a hedge fund manager? Well not much, the Autocorrelation terms are relatively weak. This would be expected, as Microsoft is a poster child for slow and steady wins the race. But there are still some patterns to look at. The up and down nature of Autocorrelation terms here represent the up and down nature of stocks in general. There is anticipated up and down in equities, and the Autocorrelation terms show that. Examining 1 to 4-month lag time, the fund manager says:

- 1 Lag Time: After observing a monthly return today, the 1 month return in 1 month is autocorrelated by -20%
- 2 Lag Time: After observing a monthly return today, the 1 month return in 2 months is autocorrelated by 9%
- 3 Lag Time: After observing a monthly return today, the 1 month return in 3 months is autocorrelated by -14%
- 4 Lag Time: After observing a monthly return today, the 1 month return in 4 months is autocorrelated by -23%

Application for Blast Motion Data

Tora Otsuka has logged 908 swings with a Blast Motion senser on his bat. Per Blast Motion, the three main scores to examine are: Plane Score, Rotation Score, and Connection Score. This writing is not a breakdown of the 3 scores, but instead an application of Autocorrelation to the data that is set up nicely in Time Series format. We can perform the same analysis with Autocorrelation for a number of other measurables coming from the Blast sensor such as Bat Speed, Time to Contact, or Rotational

Acceleration. As a reminder, the goal of Autocorrelation is to measure the correlation of a variable's value today to a variable's value in future n-lag time.

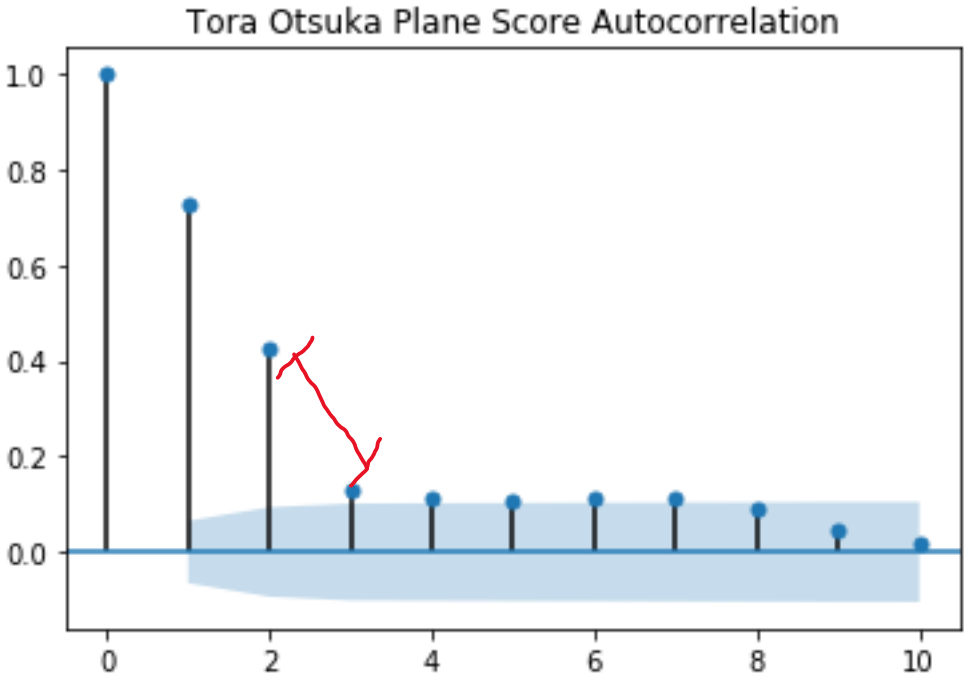
For the Autocorrelation of Tora Otsuka Plane Score, Rotation Score, and Connection score, time domain will be defined as 3 swings. The unit of measurement will be the moving average score for 3 swings. Therefore, our N-Lag time intuition is as follows:

1 Lag Time: For every observation of 3 swings with an average Connection Score, the Autocorrelation for the average connection score for the next 3 swings

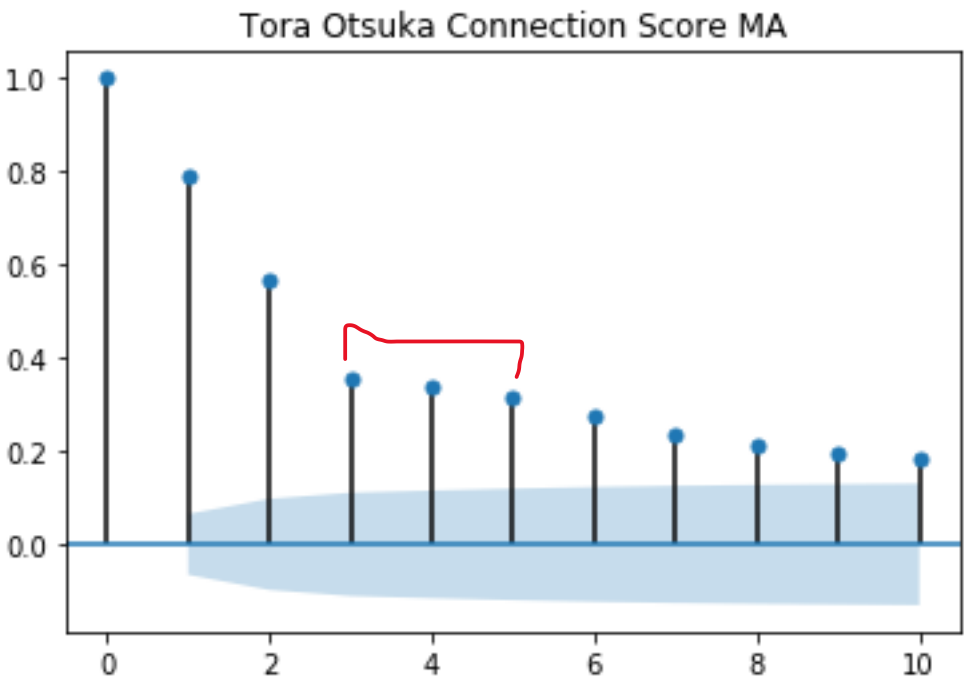
2 Lag Time: For every observation of 3 swings with an average Connection Score, the Autocorrelation for the average Connection Score for the 2nd batch of 3 swings

3 Lag Time: For every observation of 3 swings with an average Plane Score, the Autocorrelation for the average Connection Score for the 3rd batch of 3 swings

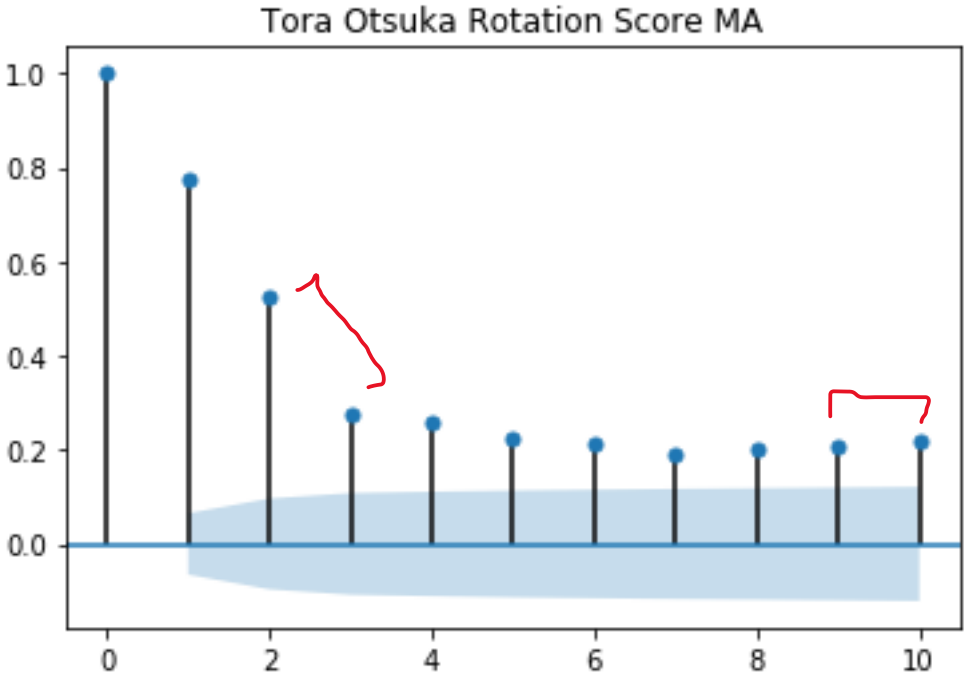
It is vital to continue to reinforce Autocorrelation as a measure of a value today and its correlation to the future value in n-lag time. Therefore, in the context of Blast Motion Data, Autocorrelation coefficients should tell us that given an observation of 3 swings and an average score, how strong the correlation is to future average scores in n-lag time. Here are the visuals:



Plane Score Autocorrelation Interpretation: Of the three scores, Plane Scores for Tora Otsuka loose predictive value the fastest. The drop-off from 2-lag time to 3-lag time is substantial, and the score eventually loses almost all predictive value. After the computer completes the computations for correlation for the score, there was likely just a lot of noise and not much signal. For now, in a batch of 3 swings with a better average Plane Score, do not expect Tora to sustain it after batch 2.



Connection Score Autocorrelation Interpretation: Connection Score maintained Autocorrelation and therefore an observation in batch 1 has greater predictive value. The computer found that the lagged pairings were more similar and correlated. Correlation of 30% is not a significant number, but when interpreted in the context of observing 3 swings, knowing that there is some predictive value for the 4th and 5th batch of swings is helpful for analysis.



Rotation Score Autocorrelation Interpretation: For Tora, Rotation Score Autocorrelation experiences the similar drop off in correlation as Plane Score in between 2 and 3 lag time. The increase in Autocorrelation from 9 to 10 lag time is an odd one.

Takeaway

Correlation always gets a bunch of hype, but Autocorrelation has a lot to explain. These analysis frameworks only work if they make their way on to a field. It could look like engaging in light conversation with the batter about their swing, adjustments they have made recently, anything that could explain the underlying data. It could also look like going to a player development coach, stating that in batches of 3 swings, Connection Score and Rotation Score hold greater predictive value to future batches of swings than Plane Score. The coach could have an idea of the player's swing deficiencies and this kind of analysis helps reinforce their intuition. This is only one framework, there are a lot more to utilize and apply.