Final: risk assessment score for twitch users/influencers

This document is part two of this paper. It was originally only going to contain some real world examples on the implications of the theories discussed in part one. However, I came across some interesting findings, while I was testing out the theories, which made me reconsider the accuracy and methodology of the system.

Earlier parts of this document will contain those findings and how they affect the calculations. The later parts will contain my remarks on the same.

D5 rule on the TSLA stock

I started by fetching, from google sheets GOOGLEFINANCE function, Tesla's stock price from 2010 all the way to 2021. The accuracy of the API data isn't going to matter here because the D5 rule will only enable us to check against already given data points. Next i wrote an algorithm in python to parse through that data and generate the higher order derivatives respectively (by the way this document will also contain a DN.py file with it that will have the algorithm I'm talking about).

The way the algorithm works is by sampling a group of say 3 data points together. Let those P1, P2, and P3--note that these are prices at given days, the smallest unit of measurement for GOOGLEFINANCE is a day.

So,

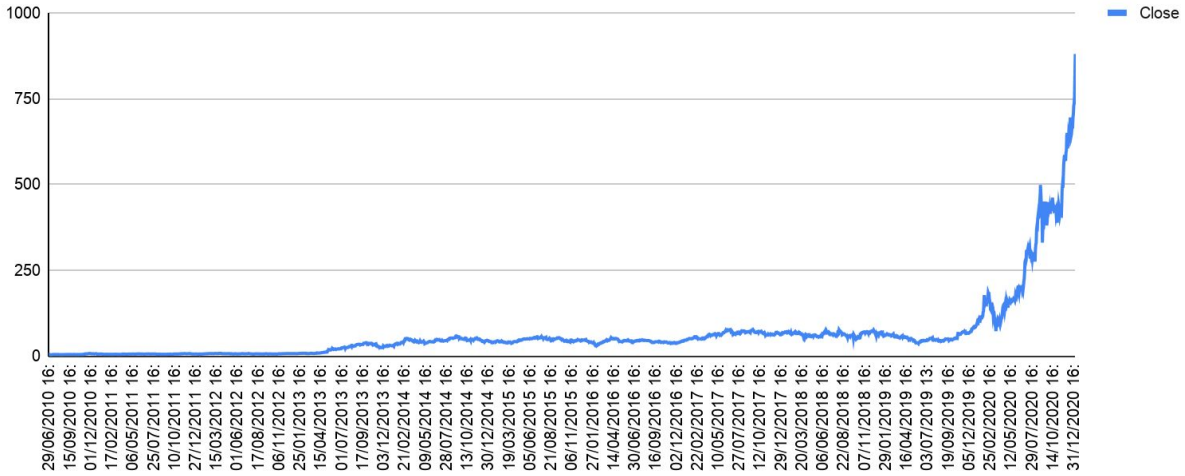
$$D_2 = (P_3 - P_1)/t$$

Where D_2 = 2nd derivative or rate of change of price on the day of P_2 (day 2).
P_1 and P_3 are data points on day 1 and 3.
t = interval between day 1 and day 3--which is 2 days so 2.

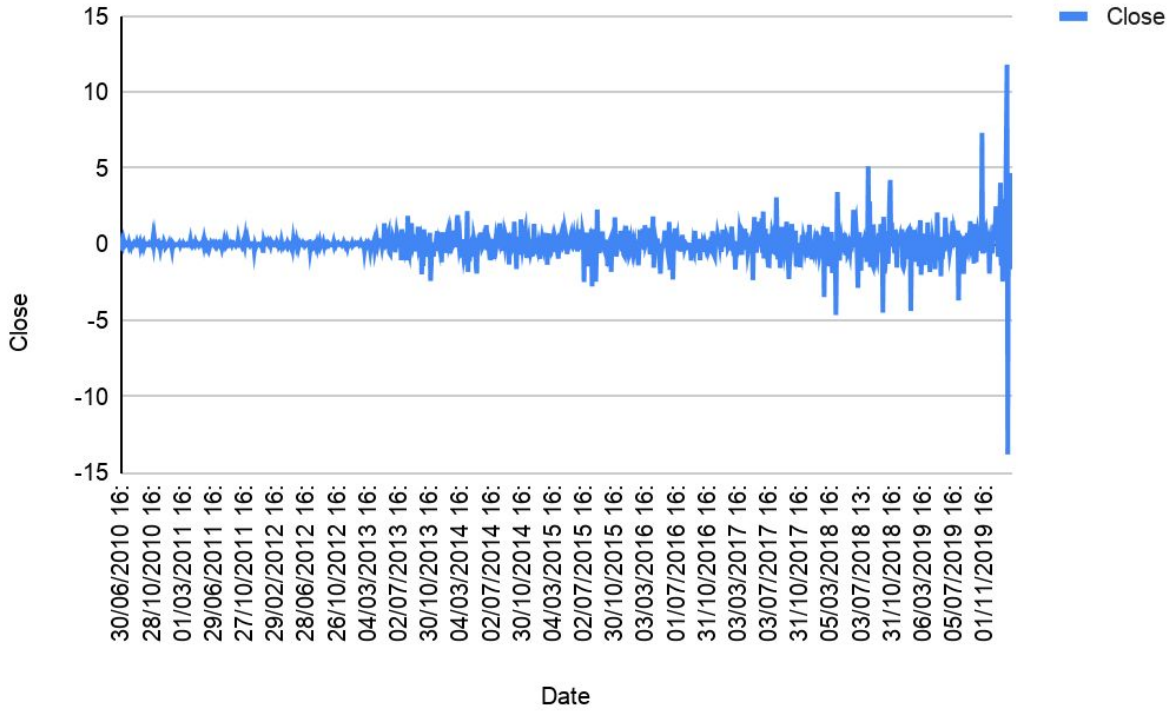
You can notice that by using the above formula, this algorithm does not create any additional data points, thus preserving the initial data space.

The graphs for the respective derivatives are given below.

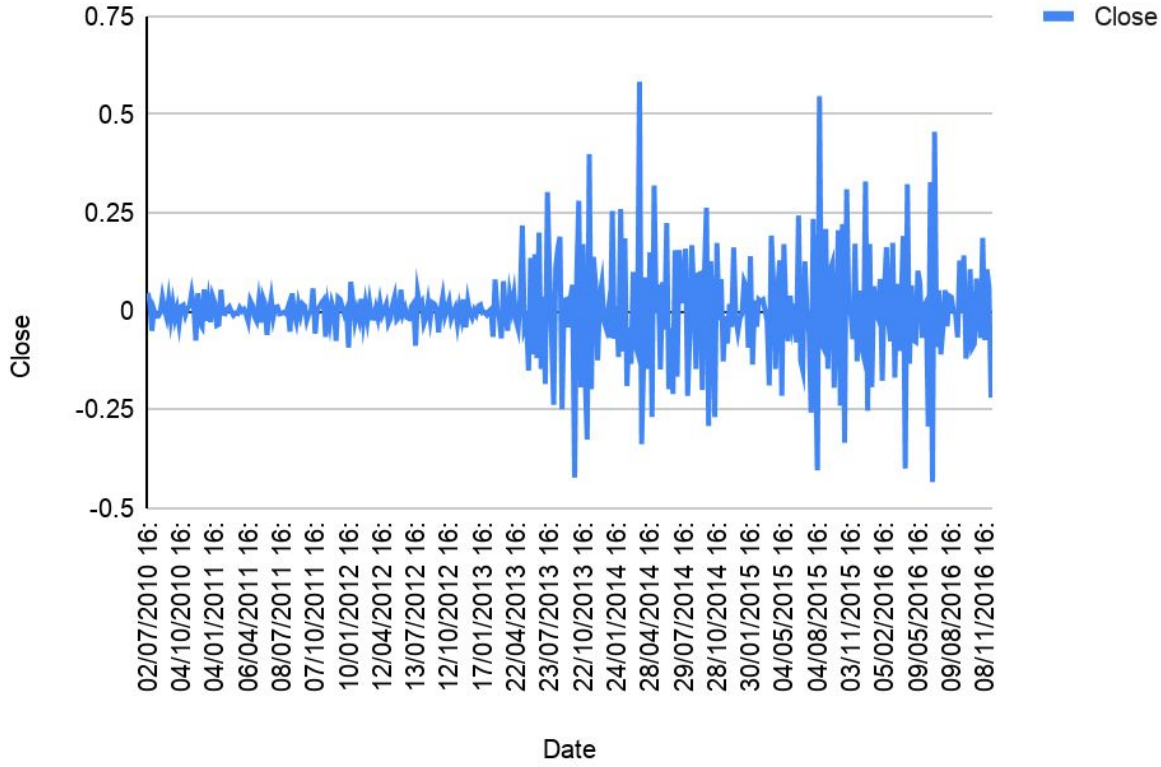
D0


Close
1000
750
500
250
0
29/06/2010 16:
15/09/2010 16:
01/12/2010 16:
17/02/2011 16:
06/05/2011 16:
25/07/2011 16:
10/10/2011 16:
27/12/2011 16:
15/03/2012 16:
01/06/2012 16:
17/08/2012 16:
06/11/2012 16:
25/01/2013 16:
15/04/2013 16:
01/07/2013 16:
17/09/2013 16:
03/12/2013 16:
21/02/2014 16:
09/05/2014 16:
28/07/2014 16:
13/10/2014 16:
30/12/2014 16:
19/03/2015 16:
05/06/2015 16:
21/08/2015 16:
06/11/2015 16:
27/01/2016 16:
14/04/2016 16:
30/06/2016 16:
16/09/2016 16:
02/12/2016 16:
22/02/2017 16:
10/05/2017 16:
27/07/2017 16:
12/10/2017 16:
29/12/2017 16:
20/03/2018 16:
06/06/2018 16:
22/08/2018 16:
07/11/2018 16:
29/01/2019 16:
16/04/2019 16:
03/07/2019 13:
19/09/2019 16:
05/12/2019 16:
25/02/2020 16:
12/05/2020 16:
29/07/2020 16:
14/10/2020 16:
31/12/2020 16:

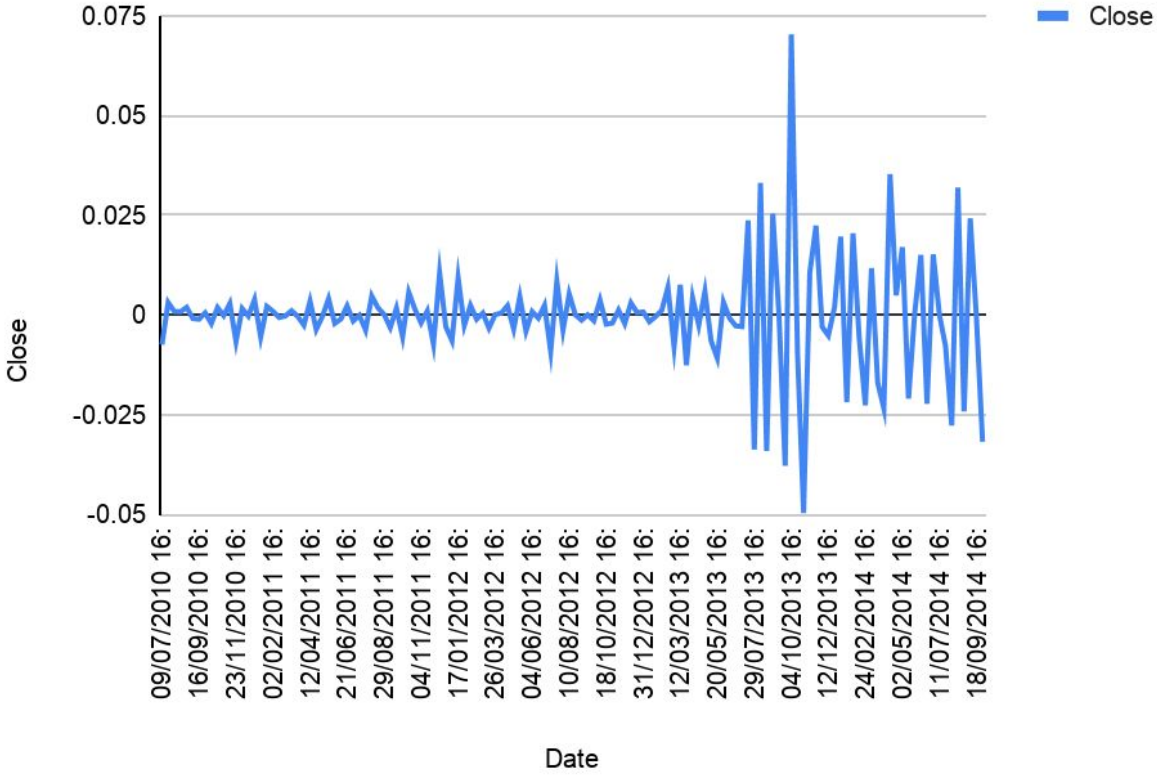
D1


Close
15
10
5
0
-5
-10
-15
Close
30/06/2010 16:
28/10/2010 16:
01/03/2011 16:
29/06/2011 16:
27/10/2011 16:
29/02/2012 16:
28/06/2012 16:
26/10/2012 16:
04/03/2013 16:
02/07/2013 16:
30/10/2013 16:
04/03/2014 16:
02/07/2014 16:
30/10/2014 16:
04/03/2015 16:
02/07/2015 16:
30/10/2015 16:
03/03/2016 16:
01/07/2016 16:
31/10/2016 16:
03/03/2017 16:
03/07/2017 16:
31/10/2017 16:
05/03/2018 16:
03/07/2018 13:
31/10/2018 16:
06/03/2019 16:
05/07/2019 16:
01/11/2019 16:
Date

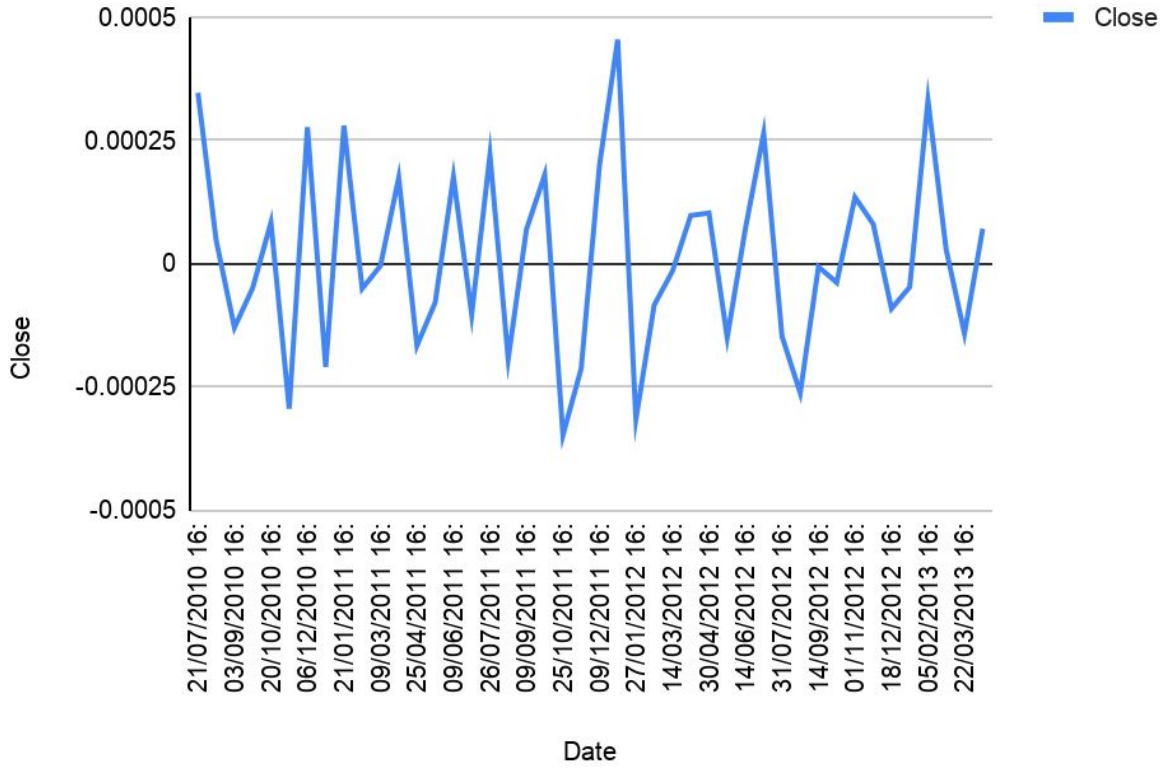
D2

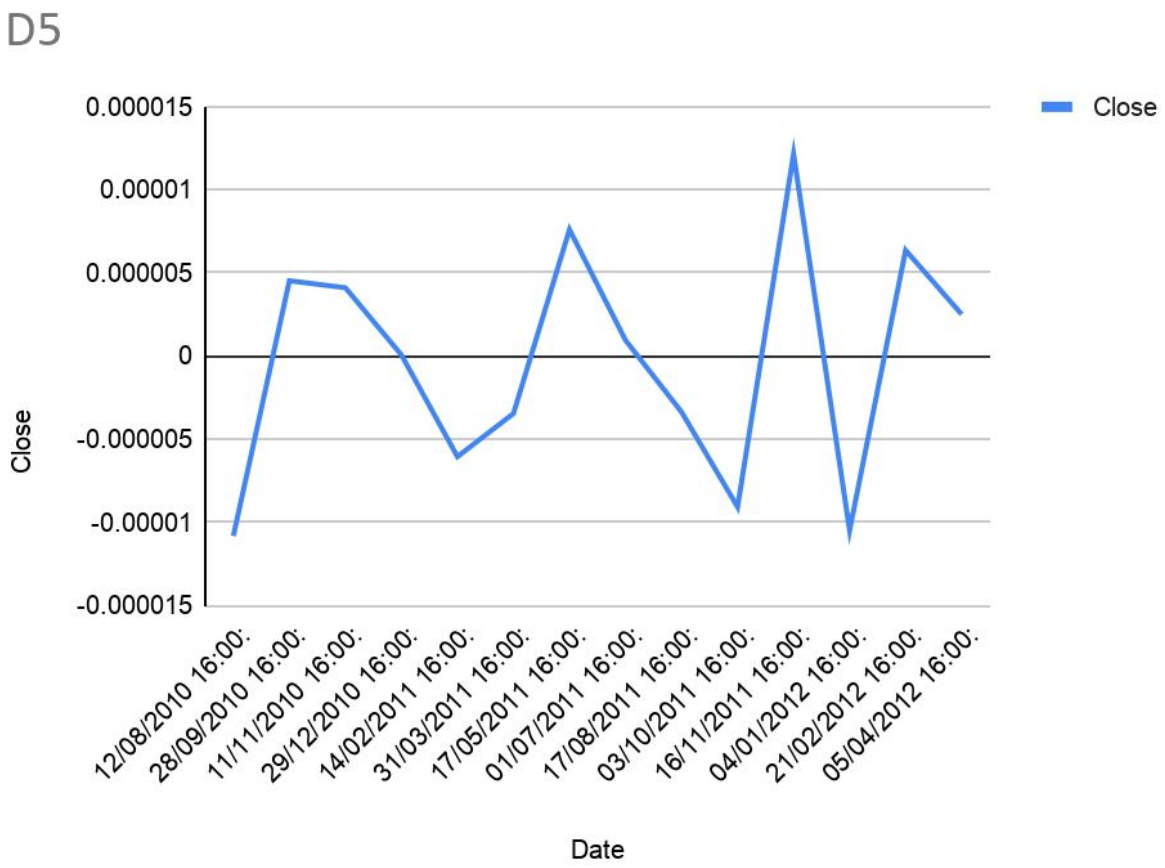

Close
0.75
0.5
0.25
0
-0.25
-0.5
Close
02/07/2010 16:
04/10/2010 16:
04/01/2011 16:
06/04/2011 16:
08/07/2011 16:
07/10/2011 16:
10/01/2012 16:
12/04/2012 16:
13/07/2012 16:
12/10/2012 16:
17/01/2013 16:
22/04/2013 16:
23/07/2013 16:
22/10/2013 16:
24/01/2014 16:
28/04/2014 16:
29/07/2014 16:
28/10/2014 16:
30/01/2015 16:
04/05/2015 16:
04/08/2015 16:
03/11/2015 16:
05/02/2016 16:
09/05/2016 16:
09/08/2016 16:
08/11/2016 16:
Date

D3


Close
0.075
0.05
0.025
0
-0.025
-0.05
Close
09/07/2010 16:
16/09/2010 16:
23/11/2010 16:
02/02/2011 16:
12/04/2011 16:
21/06/2011 16:
29/08/2011 16:
04/11/2011 16:
17/01/2012 16:
26/03/2012 16:
04/06/2012 16:
10/08/2012 16:
18/10/2012 16:
31/12/2012 16:
12/03/2013 16:
20/05/2013 16:
29/07/2013 16:
04/10/2013 16:
12/12/2013 16:
24/02/2014 16:
02/05/2014 16:
11/07/2014 16:
18/09/2014 16:
Date

D4


Close
0.0005
0.00025
0
-0.00025
-0.0005
Close
21/07/2010 16:
03/09/2010 16:
20/10/2010 16:
06/12/2010 16:
21/01/2011 16:
09/03/2011 16:
25/04/2011 16:
09/06/2011 16:
26/07/2011 16:
09/09/2011 16:
25/10/2011 16:
09/12/2011 16:
27/01/2012 16:
14/03/2012 16:
30/04/2012 16:
14/06/2012 16:
31/07/2012 16:
14/09/2012 16:
01/11/2012 16:
18/12/2012 16:
05/02/2013 16:
22/03/2013 16:
Date



Findings(problems) with the D5 rule

The first thing that I want your attention on is the observation that the data space of the D5 graph is a lot smaller than the D0 graph-- D5 graph spans from 2010 to 2012 while the D0 graph spans from 2010 to 2020. This is not a calculator or graphing error, but in fact a property of the algorithm itself--rather a property of derivatives.

When we calculate derivatives on discrete data spaces (which is very unconventional), to come up with a result for one point, we need some surrounding points. This means that we're, although, not expanding the data space, we are shrinking it in order to calculate higher order derivatives. A continuous data space (a continuous function) has infinitely many points and a change in any direction is infinitesimally small. Those two facts are the epitome of calculus and in this particular example we're limited by the amount of information, thus we can not accurately predict anything beyond 2012 with this method.

This algorithm started with over 2000 data points and in every iteration it cut the data space by a third, which over 5 iterations resulted in about 20 data points.

What does this have anything to do with accuracy? Well, we start with D5 at 05/04/2012 and if we continue the trend line for 15 days, at 20/04/2012, the D5 will be about -0.000005.

At 05/04/2012, the original price was $6.9,

So, At 20/04/2012 D0 will be

D0 + D0' + D1*(t) + 0.5*D2*(t^2) + (1/6)*D3*(t^3) + (1/24)*D4*(t^4) + (1/120)*D5*(t^5)

D0 = 6.9 + 0.105 - 0.26115 + 0.5189 + 0.2046 + 0.159 = 7.62635

So, the price at 20/04/2012, according to the D5 rule, should be about $7.62635

The big picture suggests that TSLA's stock has risen during that period, but in reality it has actually fallen to about $5.2. I didn't take into account the averaging method for each graph, and any method will be way off because of the shrunken data space. In reality, smaller order derivatives will be of larger value than higher order ones(and I suspect that D1' is actually of negative value) so adjusted for all those situations, D0 should've been something less than 6.9 and greater than 5.2. Needless to say, $7.62 and $5.2 lie in a narrow margin.

However, there's no point in doing this calculation because this only predicts something that has already happened (and with terrible accuracy).

Final remarks on the D5 rule

The current D5 rule is highly inaccurate in discrete dataspaces, and something fundamental about this rule doesn't quite fit when I process and goal of the calculation. I'd recommend not to use this version of the rule to make any kind of calculations in this score index.

The Alternative

You might recall from the earlier section, about the D5 rule on Tesla's stock, that derivatives only work on continuous data spaces, and all we have are limited data points. The natural line of reasoning would then be to convert those to a continuous function aka guessing the curve of best fit given the D0 data space.

Through finance and statistics, curve of best fit (or non linear interpolation) is a common tool to approximate the best continuous function ti fit the data, and is versatile for every kind of analysis.

Looking back at the D0 graph for Tesla's stock, I was quite amazed that this company followed an exponential curve beautifully closely. I then decided to take a look at the graphs of many of top companies in the S&P 500, and to my surprise almost every company traced an exponential curve of some kind. If you really about how these markets work, for obvious reasons an exponential curve is the ideal curve.

Nevertheless, I was able to come up with an equation to describe the tesla's curve. Note that I didn't actually interpolate, I just eyeballed and came up with a curve. I cannot say anything about the method's result without interpolating but my point is that interpolation can be useful, and combined with the D5 rule they can help with the predictive part of this system.

Implication of the system on twitch users

After the final nail in the coffin for the D5 rule, I took a look at the top 30 most watched channels on Twitch from sullygnome.com. Surprisingly enough, twitch users have a very stable curve (no anomalies or sudden up or down ticks). Social media is a relatively less volatile platform than the NYSE. This could be a good sign; If there are no surprises, we might not need the D5 rule at all. In that case, identifying the preliminary index would be enough.

Revision of the score index assessment

I got asked a question, a suggestion rather, of using a severity vs likelihood matrix to classify users based on their performance. Upon further pondering, I realized that this would fit perfectly with a volatility vs growth matrix. Influencers aren't that volatile but that doesn't mean anomalies and surges don't happen. On the other hand, the primary classifier is growth.

The new system would measure growth much the same way as it did for the primary index--with an addition of growth nature.
Nature of growth indicates how closely the growth resembles an exponential curve. A multiplier of -1 would mean an inverse exponential curve, a multiplier of +1 would mean perfect exponential curve.

Volatility of a curve is the measure of how frequent it's rate of change oscillates, and how large are those oscillations. The D1 curve is a clear example of the volatility of TSLA's stock increasing exponentially with time.

According to sullygnome.com, TheGrefg is currently the most watched channel on Twitch. They currently have about 6 mil followers and more than 102 mil views. Based on this data, I'll keep a range of about 4,000,000 as the minimum threshold for the preliminary index. Most followed channel on Twitch is Ninja with about 16 mil followers. A 4 to 16 scale sounds pretty great to begin with. Given below is a table that contains the a rough outline of how the risk matrix is going to look for TheGrefg.

Growth Index ~ 50th - 75th percentile
Volatility Index ~ 0.2

Volatility/Growth	25th	50th	75th	100th
> 0 < 0.25		X	X	
>= 0.25 < 0.5				
>= 0.5 < 0.75				
>= 0.75 < 1				

TheGrefg is one of the least volatile, but just cause of their follower count, they're quite average growth.

Footnote: The above calculations are not based on a rigid algorithm, but rather just eyeballing the current graphs and coming up with an outline of how they might look. For a rigid algorithm, numerically defining certain parameters and tweaking some scales are important but are beyond the scope of this paper.

Final remarks

Due to some inherent flaws in the earlier version the D5 rule, I would not recommend it for any calculation or analysis. The interpolation of curves could prove to be a valuable method, but given the behaviour of social media charts and curves, there's a possibility that neither D5 nor interpolation would be required to be included in the score assessment.

Finally, some revisions to the system itself have actually given some clarity and insight to the indices and numbers used in this assessment. Moving forward, I think the revised version of the system would be highly intelligible and accurate to use plainly because of the fact that it provides a spectrum of information rather than a single number to measure a score.

Looking forward to the interpolation method and implementation of this system, I think a thorough analysis of Twitch's public data is required to cater this system to the individual's needs, and to decide on the credibility of interpolation.

Given that, I hope that I've clearly lined out the basic structure of the system, cleared some doubts about different methods, and ultimately solidified the grounds on which to make measurements.