Characterization of Polynomial Interpolation on Jollibee Food Corporation, Pancake House Inc, and AGI Stock prices

Brandon Chan[1*], Kim Angelee Choa[1], Eldrin Lee[1], Al Rey Villagracia[1]
[1]Physics Department, De La Salle University, Taft Ave., 1004 Manila, Philippines

* corresponding author: brandon_chan@dlsu.edu.ph

Abstract Previous studies have shown that there is some degree of predictability in the stock price behavior. This paper aims to capture the said predictability of the stock prices of Jollibee Food Corporation, Pancake House Inc, and AGI via interpolation. In this study, the researchers ran different kinds of interpolating polynomials and analyzed their accuracy in predicting the stock prices for the next five days. Three characteristics were monitored: data distribution behavior (linear, quadratic, and exponential), data distribution distance, and number of data points. These parameters were varied and were run through random points and calculating their absolute errors for each day. Results showed that for data distribution behavior, the accuracy of the quadratic distribution decays the least as the number of data points increases. Along with that, there is a consistent increase in accuracy when the data distribution distance reaches a certain point. However, the best number of data points is two, which renders the question of the best data distribution behavior moot.

Key Words: stocks, stock market, interpolation

1. INTRODUCTION

The world of stocks is generally known to be stochastic, where investors may gain huge amount of money in just one day, or lose it all in a second. That, however is not the case, as some studies have shown that stock price behavior can be predicted with a certain amount of accuracy. To be able to further support this statement, this experiment aims to be able to use interpolating polynomials to better predict the price of the stock market, using different characteristics of the interpolating polynomial such as the number of data points used and how the data points are distributed among available data points. This experiment seeks to find the best characteristics used for interpolating polynomials extrapolating future values of the stock price, and determine whether or not such a method is reliable for stock price prediction.

As it may seem that stock prices are completely random, some studies and experiments have seen correlation between present and past data, suggesting that stock prices are not completely independent from the company's past stock prices. Given that insight, a well-constructed model, even as simple as an interpolating polynomial, may be created to predict future stock prices to some degree of accuracy. However, a well-constructed model is not merely putting in the most number of data points as one can and constructing the interpolating polynomial. Runge's phenomenon has shown that too many interpolation points can lead to disastrous and chaotic patterns in the interpolating polynomial that departs from the original function. Such a problem may be answered, fortunately, by changing the distribution of the interpolation points.

2. METHODOLOGY

For the required software, the researchers used the FORTRAN language to carry out the long, numerous, and tedious calculations, which resulted in the calculation of the absolute errors of the predictions. After which, Microsoft Excel was used to look for patterns in the accuracy, along with generation of graphs. For the data points used in this experiment, the researchers have gathered the history of stock prices of three companies with similar service, with data points dating back to the last 250 days starting from June 29, 2014. These companies are namely: Jollibee Food Corporation, Pancake House Inc, and AGI. These data points have been gathered from the website of the Philippine Stock Exchange. The researchers have also gathered the consumer price index data, with plans of integrating them into the interpolation values.

After gathering, the researchers put the values of the stock price, the date, and the consumer price index into a text file, so it is readable by FORTRAN. The number of data points gathered is 250, wherein a set of predictions used at most 100. After which, the program was written such that it carried out numerous interpolations and extrapolations of data points given three parameters: the number of data points, the type of distribution (linear, quadratic, and exponential), and the characteristic of the said distribution (i.e. the distribution distance). After the extrapolation, the program will compute for absolute error for each day that the program tries to predict.

The linear, quadratic, and exponential distribution methods are used by researchers to determine the data points used in the interpolation method. They are found by the following equations:

$$X_i = a_0 + \lfloor a_i * (i-1) \rfloor \quad \text{(Eq. 1)}$$

$$Y_i = a_0 + \left\lfloor a_i * \frac{(i^2 - i)}{2} \right\rfloor \quad \text{(Eq. 2)}$$

$$Z_i = a_0 + \lfloor a_1^{i-1} - 1 \rfloor \quad \text{(Eq. 3)}$$

where:

X_i, Y_i, Z_i = the data point order the interpolation is going to use for data point

a_0 = a randomly generated number from 1 to 151.

a_1 = the distribution characteristic.

i = integers ranging from 1 to n, where n is the number of data points beyond the first.

The method of interpolation that the researchers used is called the Lagrange Method, which uses the following formulas for finding the interpolating polynomial:[2]

$$A_x(y) = \prod_{j \neq x} \frac{(y - y_j)}{(y_x - y_j)} \quad \text{(Eq. 4)}$$

$$P(y) = \sum_{i=1} f(i) A_i \quad \text{(Eq. 5)}$$

Where:

$P(y)$ = the estimated stock price days after the starting day.

$f(i)$ = the stock price for day i, where takes on the values of and

y_i = the stock price of day

$A_i(y)$ = the Lagrangian polynomials

While the absolute error is given by this equation

$$Abs.Err. = \frac{|p_0 - p|}{p_0} \quad \text{(Eq. 6)}$$

where:

p_0 = the actual price for that day

p = the estimated price for that day

The methods were repeated through the three different companies, and with different data point sets and the absolute error were determined and tabulated with each trial, with its corresponding characteristics. The researchers ran five random tests each on three different companies, garnering a total of fifteen tests.

After acquiring average absolute error values corresponding to the different characteristics they were tabulated and analyzed.

3. RESULTS AND DISCUSSION

Here are the results of running the program:

Table 1. The error averaged over all fifteen trials and distribution characteristics

DT	DP	A.E.1	A.E.2	A.E.3	A.E.4	A.E.5
1	2	2.270946	3.08192	3.790279	4.24211	4.447554
1	3	2.584545	3.889141	5.336527	6.721371	8.021457
1	4	3.283009	6.039494	9.962042	15.00775	21.40239
1	5	4.789967	11.52148	23.44428	42.85205	72.4685
1	6	10	7.918758	24.66709	60.89275	131.1858
1	7	14.15898	54.62643	159.9153	398.3752	881.4164
2	2	2.270946	3.08192	3.790279	4.24211	4.447554
2	3	2.574797	3.843155	5.221145	6.523815	7.695264
2	4	3.036019	5.273306	8.172744	11.65206	15.79503
2	5	3.678244	7.554889	13.51078	21.85007	33.25967
2	6	4.431983	10.52828	21.18807	37.75173	62.62675
2	7	5.512275	15.07765	33.51248	64.58185	114.456
3	2	2.270946	3.08192	3.790279	4.24211	4.447554
3	3	2.903733	4.784489	7.024989	9.294411	11.60289
3	4	4.0602	8.6068	15.4677	24.29037	35.58697
3	5	4.9621	11.92865	23.8866	41.107	65.50865
3	6	6.7848	19.0466	42.0752	78.8491	134.461
3	7	6.9664	20.0333	45.0013	85.9081	148.6779

From table 1, the average error (from the fifteen trials) was computed with its corresponding data distribution type (1 for Linear, 2 for Quadratic, 3 for Exponential). It is apparent that Distribution Type 2 (Quadratic)'s accuracy decays the slowest among the three, making it the best data distribution type among them.

It is also apparent that as data point number increases, the accuracy of the interpolating polynomials strictly decay. From this, it can be concluded that the best number of data points is two.The absolute error of the predictions of interpolation using different data distribution methods were plotted given the day of prediction and the number of data points used. The results were the following:

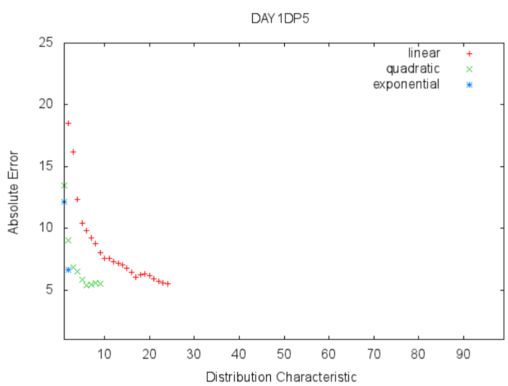

Figure 1A: Graph of absolute error of the interpolation of PCH stock prices with different data distribution on the first day, with 5 data points.

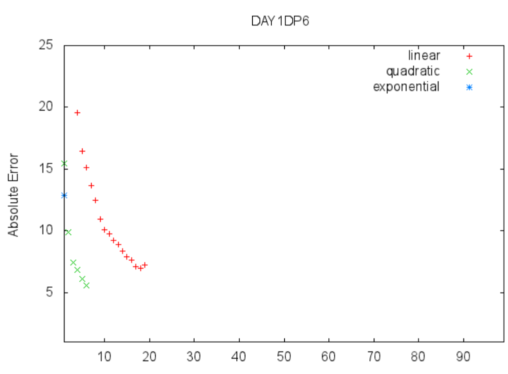

Figure 1B: Graph of absolute error of the interpolation of PCH stock prices with different data distribution on the first day, with 6 data points.

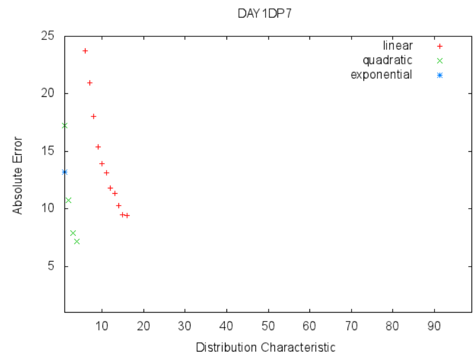

Figure 1C: Graph of absolute error of the interpolation of PCH stock prices with different data distribution on the first day, with 7 data points.

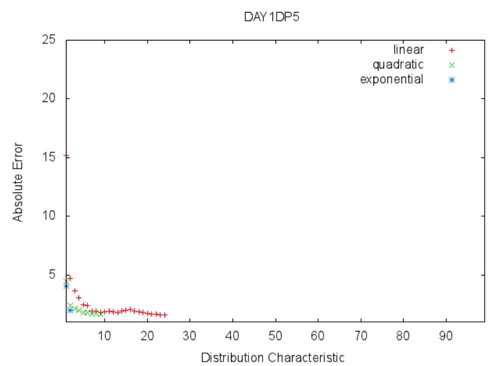

Figure 2A: Graph of absolute error of the interpolation of JFC stock prices with different data distribution on the first day, with 5 data points.

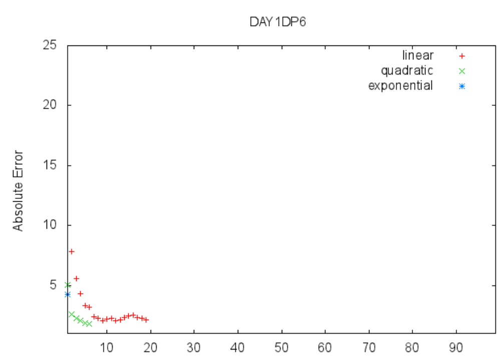

Figure 2B: Graph of absolute error of the interpolation of JFC stock prices with different data distribution on the first day, with 6 data points.

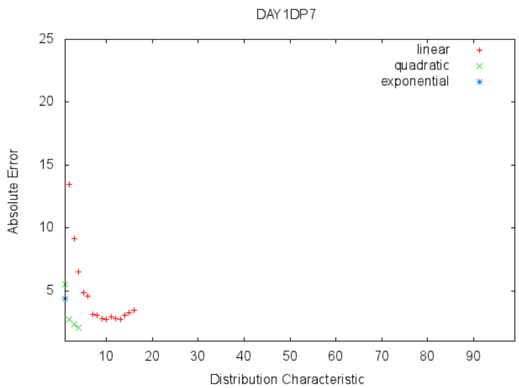

Figure 2C: Graph of absolute error of the interpolation of JFC stock prices with different data distribution on the first day, with 7 data points.

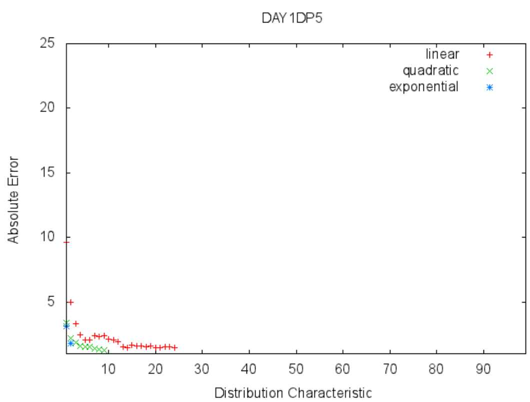

Figure 3A: Graph of absolute error of the interpolation of AGI stock prices with different data distribution on the first day, with 5 data points.

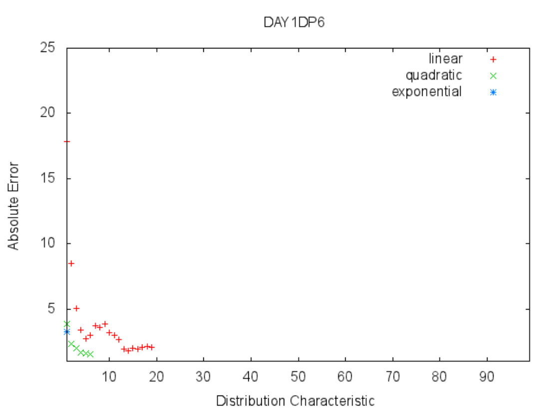

Figure 3B: Graph of absolute error of the interpolation of AGI stock prices with different data distribution on the first day, with 6 data points.

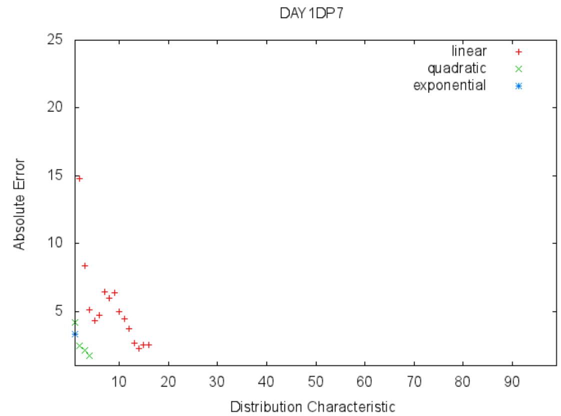

Figure 3C: Graph of absolute error of the interpolation of AGI stock prices with different data distribution on the first day, with 7 data points.

As seen in Figures 1A - 3C, it is evident that the absolute error goes down usually when the distribution characteristic is increased. However, using the 100 data points, the quadratic distribution reaches the lowest absolute error when compared to the other distribution methods.

Given that the best number of data points used is 2, and since 2 data points will yield the same distribution for the linear, quadratic, and exponential methods, the result becomes insignificant for the purposes of prediction.

4.CONCLUSION AND RECOMMENDATION

Among the three types of data distribution tested (Linear, Quadratic, and Exponential), the quadratic distribution decays the slowest, and maintains to be the most accurate of the three. However, an increase in the number of data points used in interpolation ensues the strict decrease in accuracy for the extrapolation, therefore the best number of data points is the minimum number of two. Lastly, a decreasing exponential pattern was observed from the plot of the distribution characteristic vs. the absolute error, which implies that the farther the data points reach through past data, the better the interpolating polynomial's prediction seems to be.

5. REFERENCES

Abramowitz, M., & Stegun, I. A. (1964). *Handbook of mathematical functions with formulas, graphs, and mathematical tables*. Washington: U.S. Govt. Print. Off.

Hayter, A. (2013). *Probability and statistics for engineers and scientists*. Australia: Cengage Learning.

Harder, D. (n.d.). *Polynomial Wiggle*. Retri'eved June 24, 2014, from https://ece.uwaterloo.ca/~dwharder/NumericalAnalysis/05Interpolation/wiggle/

PSE (n.d.). *THE PHILIPPINE STOCK EXCHANGE, INC*. Retrieved June 24, 2014, from http://www.pse.com.ph/stockMarket/companyInfo.html?id=135&security=163&tab=0

Snee, T. (2014, January 31). *Predicting Stock Prices*. Retrieved June 24, 2014, from http://now.uiowa.edu/2014/01/predicting-stock-prices

Wible, B. (2009, June 1). *Picking Up Pennies for Profit*. Retrieved June 24, 2014, from http://insight.kellogg.northwestern.edu/article/picking_up_pennies_for_profit