50 Shades of Stable Distributions

by Rakeen Tanvir, Bruno Kömel, and Derl Clausen

Financial markets are rich with properties that both delight and doom mathematical theorists and practitioners. This paper seeks to determine whether financial indices provide evidence for infinite variance in share values, in particular with a historical data set from the Dow Jones Industrial Average (DJI) from 1986 to 2020. Common characteristics of market behavior dealt with herein include volatility, randomness, and cycles. The findings suggest that there is strong reason to model financial data on several time horizons using stable distributions, which range from the familiar Gaussian to models that exhibit infinite variance such as the Cauchy distribution. The conclusions support further research into these models to better account for uncertainty and extreme market behaviors that have significant political and economic implications.

1 Background

During the 2020 COVID-19 pandemic-fueled recession, the presence of large sigma events, exhibited primarily by market crashes, once again raised the age-old question: Do market time series have infinite variance? If share values were normal, then large sigma events would be unlikely and infrequent. In addition, if market share values behave according to models with finite variance, then expected variance is well-defined.

The raw DJI.csv file contains data drawn from Yahoo! Finance from January 01, 1985 through March 20, 2020 for 8858 observations (rows) of the DJI index. It contains seven variables (columns): date, share price at open, high, low, close, and adjusted close, and the trade volume.

2 Methodology

The null hypothesis is that market index share values follow a probability model with finite, well-defined variance. The initial results of the exploratory data analysis indicated that first differences in the data have properties of both normal and heavy-tailed distributions. Given this, it is natural to leverage stable distributions parameter estimation and data.

The finite DJI data produces a finite variance since sample variance is always defined, so it is a challenge to infer infinite variance from a finite data set. There is a family of distributions where the choice of parameter values affect whether variance is well-defined: the stable distributions. A stable distribution with alpha parameter 1 is the Cauchy distribution and with alpha parameter 2 is the normal distribution. Mathematically, the integral of the second moment for a Cauchy random variable is divergent, i.e. it does not exist. Thus, we have infinite variance.

By the central limit theorem, would expect that as sample size increases, our standardized observations would have a standard normal distribution if the observations are independent and identically distributed random variables from a light-tailed distribution. However, due to undefined variance, this result should not hold for the Cauchy distribution. We should be able to see this in a simulation of empirical cumulative distributions.

3 Results

Initial data exploration reveals that the DJI open values resemble those of the other price variables, so analyzing only the open values is sufficient for our purposes. By visual inspection, open values seem to follow a random walk while their first differences seem to be a heavy-tailed distribution with a left skew and high density around the center. The remaining analyses focus on these first differences.

There were 32 days with greater than 5 sigma events. The least rare of these rare events are expected to occur every 12000 years by an assumed normal distribution.

The data is then explored further for properties of a random walk and infinite variance. The Fourier Analysis allows us to confirm that there are no cyclical patterns in the data to inhibit us from creating a sufficiently good random walk model.

While the normal distribution fits the first differences for a Gaussian random walk model, it is a poor fit for both first differences of open values as well as the first differences of the logarithm of open values, as demonstrated through visualizations of histograms and probability density curves.

Fractal analysis formulated by Benoit Mandelbrot allowed the estimation of wild randomness in the data using the Hurst exponent. This was used to demonstrate that both our first difference data as well as our Gaussian random walk model had Hurst exponents around 0.5, meaning they exhibit extreme and similar wild randomness. Thus, it seems our first differences follow a random walk, but do not follow a normal distribution.

The Cauchy distribution is overlaid on histograms of the first differences and of their logarithm. Visually, it seems to fit. If this assumption is true, then by integration we show that the expectation for the second moment of a Cauchy random variable has an undefined integral, indicating that its variance is divergent.

A partial variance analysis demonstrates that variance is directly proportional to elapsed time. However, while over log-time the variance converges for the normal distribution, it increases in infrequent but extreme jumps for the Cauchy distribution, demonstrating divergence. Interestingly, the variance for first differences and open values both increase indefinitely in a smoother fashion, also demonstrating divergence, but possibly less susceptible to extreme values due to finite samples.

In fact, it turns out that sample size also affects the size of variance and in turn how well a given model fits the data. Increasing the size of the samples drawn from our data and comparing them to normal and stable distributions shows that at lower sample sizes (500 elements), the data is well approximated by a Cauchy distribution. This good fit fades as sample sizes increase. In a particular simulation of running repeated Kolmogorov-Smirnov tests, 92 percent of p-values were above the 0.05 threshold for rejection of the null hypothesis. Contrastingly, even at small sample sizes the data is not well modeled by a normal distribution, as measured by Kolmogorov-Smirnov tests.

Interestingly, the bootstrap simulation visually shows that the first differences behave more similarly to the normal distribution than they do the Cauchy distribution in terms of empirical cumulative distribution behavior and the Kolmogorov-Smirnov test statistic. This is likely because the first differences come from a single finite sample that we draw bootstrap samples whereas the Cauchy sample is drawn from the underlying Cauchy distribution itself.

Stable Distribution

A chi-square goodness-of-fit test was run on several stable distributions including the Gaussian and Cauchy distributions. Each test rejected the null hypothesis that the data followed the reference distribution for the first differences of Open values averaged over days as well as over months. However, the first differences of the yearly average of Open values smoothed enough of the noise to produce a failure to reject the null hypothesis in the case of the normal distribution, meaning it passed the chi-square goodness-of-fit test.

Our analysis concludes with the consideration of the impact of political regimes on the performance of the market. The unsurprising result is that the granularity of analysis is as significant as the data. By permutation tests, we observe that there is a statistically significant discrepancy in Dow Jones performance in favor of democratic administrations. However, as we control for more variables via regression analyses, we note that the significance of the impact of individual administrations diminishes completely. In fact, we only observe a significant ”White House” impact on the Dow Jones Industrial Average once we remove days where the first differences were more than five standard deviations away from the population mean. However, given our thesis that measuring market fluctuations in light of the standard normal distribution is unintelligible, as the market seems to exhibit infinite variance, the practice of ignoring ”rare” events appears unwarranted.

4 Discussion

The results indicate that the first differences follow different stable distributions depending on which time scale is used to model the data. Cauchy seems to work best with daily first differences whereas normal seems to work best with yearly first differences.

The financial implications could be enormous. In applications such as actuarial modeling and reinsurance where the risk of low-probability, extreme values could be disastrous, the ability of deployed models to capture this phenomenon is paramount to reliable forecasting. Additionally, an important question to be grappled with is whether we, as an economy, have appropriately valued the market as a valuation mechanism. For the issue ought to be addressed of whether it is wise to rely on an infinitely complex system to determine yearly bonuses, retirement plans, and policy decisions.

The economic implications are that market behavior is both difficult to predict, and also vulnerable to large sigma events that cast doubt on the validity of the best financial models to accurately capture real-world behavior. Further, our data seem to indicate that large sigma events are more prevalent in the negative direction. This has profound implications for the livelihood of the various stakeholders in capital markets. Most notoriously, large negative sigma events have the dreadful effect of draining savings and retirement plans. In addition to these first order effects, one cannot ignore the second and third order effects of large negative sigma events, as seen in the recession in 2008, which also seem inevitable in light of the current pandemic.

Perhaps the permutation tests on political regimes can undercut arguments that closely tie the specific presidential administration in power to stock market performance, hinting at a foundation for policy that can stand on its own merits, not relying on myopic market returns for validation.

Philosophically, the analysis speaks to the fallibility of deterministic models of the world with finite constraints. If infinite variance can exist within human-made systems like financial markets, then what are the limits of constrained models to capture real-world behavior, whether in nature or society? The analysis also poses a challenge to our presumed ability to perceive the essence of the reality from the imperfect, partial, and finite world that surrounds us. In terms of this project, how can we *know* that the essence of capital markets is to be infinitely variant, when we can only *see* finite transactions?

5 Conclusion

This analysis concludes that financial market index data of changes in share value are well-modeled by stable distributions, with the Cauchy distribution working better over shorter time scales such as daily differences and the normal distribution working better over longer time scales such as yearly differences. Assuming the Cauchy distribution or a similarly heavy-tailed distribution underlies market share data, then it has infinite variance.

Further research is needed to validate these conclusions. The data is a snapshot of the DJI market data and does not reflect the entire population. In addition, the DJI is a single market index and is one of many indices and other mechanisms to track financial market share value. A wider analysis of the available financial data is needed to understand whether our results generalize to reflect overall market behavior.

Acknowledgements

Special thanks go to our teachers Paul Bamberg, Michael ”Freshwater” Liotti, Grant ”Please don’t grade this” Murray, and Alek ”Monte Carlo” Westover for your patient instruction throughout the past year. MATH23 has been

an unforgettable time series of fantastic moment generating functions.

References

[1] Benoit Mandelbrot *The Variation of Certain Speculative Prices*, The Journal of Business, Vol. 36, No. 4 (Oct., 1963), pp. 394-419 The University of Chicago Press

[2] Benoit Mandelbrot *How Fractals Can Explain What's Wrong with Wall Street*, Scientific American, Feb. 1999 Spring Nature America, Inc.

[3] Benoit Mandelbrot & Richard L. Hudson *The (Mis)behavior of Markets*, 2006, Basic Books

[4] Benoit Mandelbrot & Howard M. Taylor *On the Distribution of Stock Price Differences*, Operations Research, Vol. 15, No. 6, November-December 1967, pp. 1057-1062 Institute for Operations Research and the Management Sciences

[5] Eugene F. Fama *The Behavior of Stock-Market Prices*, The Journal of Business, Vol. 38, No. 1. (Jan., 1965), pp. 34-105

[6] Michael Grabchak & Gennady Samorodnitsky *Do financial returns have finite or infinite variance? A paradox and an Explanation*, Quantitative Finance Volume 10, 2010 - Issue 8 pp. 883-893 Routledge

[7] J. Huston McCulloch *Simple Consistent Estimators of Stable Distribution Parameters*, Communications in Statistics - Simulation and Computation Volume 15, 1986 pp. 1109-36 Taylor Francis

[8] M. Mahdizadeh & Ehsan Zamanzade *Goodness-of-Fit Testing for the Cauchy Distribution with Application to Financial Modeling*, Journal of King Saud University - Science, Volume 31, Issue 4, 2019,pp. 1167-1174, King Saud University

[9] Vladik Kreinovich & Olga Kosheleva *Goodness-of-Fit Testing for the Cauchy Distribution with Application to Financial Modeling*, International Journal of Innovative Management, Information Production, 2012, Volume 3, No. 3, 2012,pp. 1-9, Waseda University

[10] Marco Tavora *How the Mathematics of Fractals Can Help Predict Stock Markets Shifts*, Towards Data Science June 16th, 2019 Medium

[11] Henry (https://stats.stackexchange.com/users/2958/henry), *Test for finite variance?*, https://stats.stackexchange.com/q/9143 (version: 2011-04-04) , Cross Validated

[12] Mur1lo (https://stats.stackexchange.com/users/120428/mur1lo), *Test for finite variance?*, https://stats.stackexchange.com/q/272557 (version: 2017-04-07), Cross Validated

[13] Kyle Siegrist, Phd, *Proof: The normal distribution is stable with index alpha = 2*, https://www.randomservices.org/random/special/Stable.html (accessed: 5/10/2020), Random Services