Financial Market Randomness

Nathan Shepherd

Paper Outline

1 Introduction

Since the introduction of the stock market there have been individuals who have both gained and those who have lost income within it. Specific techniques of trading thus return a value to an investor after execution. However, there is also a controversy about this as some financial analysis literature suggests that value based trading follows a random distribution [1]. Implying that there exists no algorithmic process guaranteed to halt on a profitable return given any stock price histories besides a random one. The focus of this paper is an investigation to whether or not there exists a deterministic influence of these histories on future price movements.

1.2 Why Explore Trading Strategies?

A dichotomy between the wealthiest of the world's people and the poorest results in a need for financial independence. The American stock market could prove to be an important asset for saving given there exists a reasonable strategy to supplement income over time. There are many ways to lose money using these investments thus this paper attempts to shed light on simple techniques to produce reasonable investment growth over a moderate period of time.

1.3 Similar Work

Since the early statistical work on stock prices, up to 1975, a number of new and potentially important statistical models and techniques have been developed. Some arrive with a great flourish and then vanish, such as catastrophe theory, whereas others seem to have longer staying power. More recently there has been published work that suggests the change in stock prices follows a cumulative distribution 'white chaos'. These series have the physical appearance of a stochastic independent (i.i.d). process and also the linear properties of a white noise such as zero autocorrelations and a flat spectrum. [5]

The question arises of whether the series of prices that might appear as stochastic white noise are actually white chaos and are thus actually perfectly forecastable. At least in the short run and provided the actual generating mechanism is known exactly. Some research suggests that chaotic processes occur frequently. On the other hand, a clear case can be made that they do not occur in the real world, as opposed to in laboratory physics experiments. There is no statistical test that has chaos as the null hypothesis. There also appears to be no characterizing property of a chaotic process, that is a property that is true for chaos but not for any completely stochastic process. [1,5]

Published research that is easily accessible for the domain of financial market trading is difficult to find. Particularly, the author of such a paper has the incentive to not publish their best results or choose to make profit from the analysis contained therein. Resulting in many articles which take the form of "Lessons for forecasters" by Clive W.J. G-anger. Similarly, "Naive trading rules in financial markets" authored by Neftci. S.N. simply explain methods that should not work.

1.4 Overview

The research in this paper investigates the relationship between time and the expected return as a function of the cost associated with a portfolio of one or more stocks. The following pages include sections that denote multiple facets of analysis and explanation. An emphasis is made on leveraging the large amount of data available online compared to more historical research that use smaller samples. Also this paper takes advantage of faster computation time that has occurred since the introduction of the personal computer.

To the effect of addressing a degree of randomness within the distribution of stock price data several techniques are used. This includes an estimation of population distributions for the expected return on investments and price histories. As well as a comparison of metrics to measure magnitudes of information gain given by transformations of price level histories. Then conclusions are drawn ascertaining plausibility in the proposed trading strategies with coefficients

2 Data

As a starting point, the S&P 500 stock market index was collected from the records of Yahoo Finance. December 2020 tickers contained within the fund, maintained by S&P Dow Jones Indices, are 505 common stocks issued by 500 large-cap companies and traded on American stock exchanges. S&P® is a registered trademark of Standard & Poor’s Financial Services LLC (“S&P”), a part of McGraw Hill Financial. Dow Jones® is a registered trademark of Dow Jones Trademark Holdings LLC(“Dow Jones”).

Day level closing price information extracted from 12/1/2015 until 12/1/2020 provides the basis of analysis for this paper. Providing 1266 data points for each of the 505 companies.

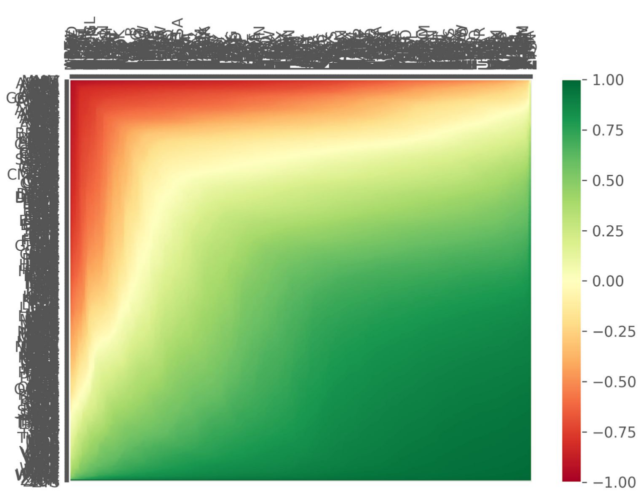


Figure 1: A correlation heatmap between the individual stocks of the S&P 500 show that the majority are positively correlated. Notable is the wide band of weakly correlated data points.

More data is collected from Yahoo Finance that includes an aggregate of the top performing companies' stocks for 16 global indices is collected. This information will be used in a later model to simulate and analyze the effect of different markets within an optimal portfolio of stocks. From a heatmap of the cross correlation between all global indices there is some indication of various directions of change for the price of a stock in different markets. However, compared to the S&P500 this index is much more neutrally correlated.

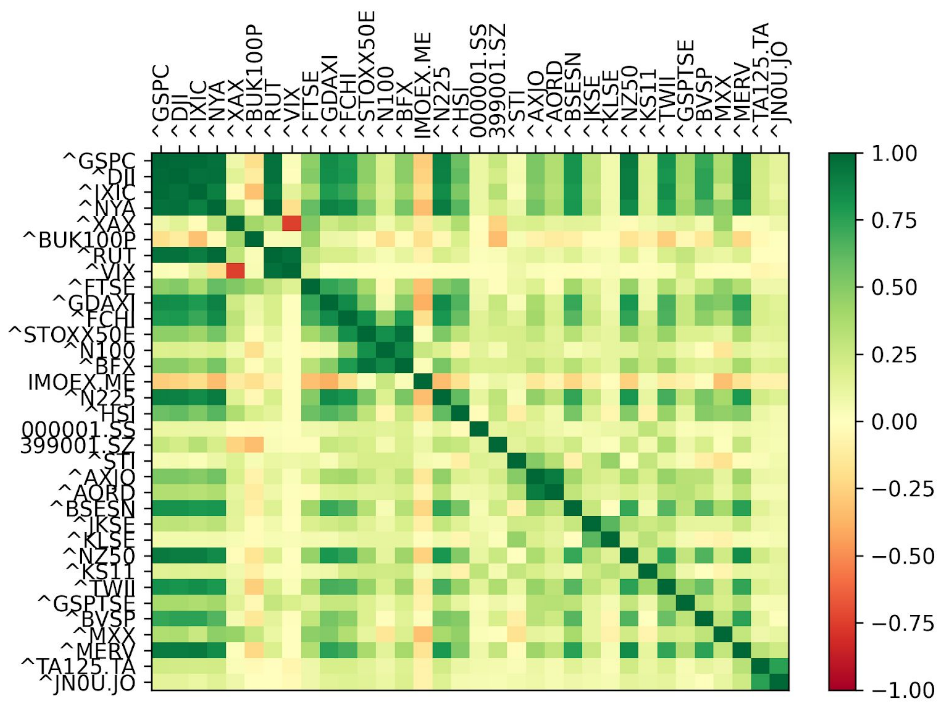


Figure 2: Heatmap of cross correlation for the selected portfolios of Global indices.

3 Methods

For reference some initial notation and assumptions are given. [1]

h	= duration of time in days ≥ 0
t	= an interval of time such that t precedes $t+1$
S_n	= a portfolio of n stock investments
P_t	= daily stock closing price, at time t
X_t	= any transformation of P_{t-h} from S_n
D_t	= dividend for period t such that $D_t = \frac{d}{dt}log(P_t) = P_t^{-1}$
R_t	= investment value growth $= (P_t + D_t - P_{t-1}) / P_{t-1}$

The risk is measured using the capital asset pricing model (CAPM). Where $r_{S_n}(t)$ is defined over the larger market from which the portfolio is selected. This aggregates multiple stocks from a domain in order to estimate the risk of a portfolio within that domain. [1, 5]

CAPM$\Leftarrow$	$R_t - r_t \approx$ market excess return
$r_{S_n}(t)$	= return on a ‘risk free’ investment
$\xi_{S_n}(t)$	= excess return (varies per market) = $R_t - r_t$
$\Omega(S_n)$	= risk level of a portfolio
A_t	= risk adjusted profits = $\xi_{S_n}(t) * \Omega(S_n) + r_t$

In order to actually make predictions, one must carefully consider the combination of above parameters and assumptions within the financial market. The method used to preprocess the stock information includes a calculation of the above parameters. Research suggests that prediction of a stock price is clearly possible and there exists many statistical techniques that can be applied to maximize returns. Most notably there is improved success in identifying future prices from investing over long periods of time. Also, carefully choosing which portfolios are closely related to the larger market or a single company. [1,2,3] With respect to outliers this paper makes the assumption that every stock within S_n has been combined into a matrix of size (t, n). The matrix computed as an outer join on all S_1 and missing values are interpolated locally by a linear regression between the two border points of the missing value. Ultimately, this suggests that given current research for predicting the return on investment can take the form:

β = beta coefficients $< \beta_0, \ldots, \beta_n >$ of Ordinary Least Squares

X_{t-h} = Matrix of assets such as $\{R_t, A_t\}$, that coincide with all of S_n

$$\mathrm{E}[R_t \mid X_{t-h}] = \psi(w_{t-h})) * \beta * X_{t-h}$$

Where $\psi(w)$ is a cumulative distribution function of a continuous random variable defined on the interval $0 \le \psi(w) \le 1$. This function has parameter *w* which is a switching random variable. Thus, $\mathrm{E}[R_t \mid X_{t-1}]$ utilizes the assumption that at some point time, *t*, there exists a likely signal to buy or sell depending on the value and of ψ and distribution of *w,* that is present *h* days before the investment. [5]

Furthermore, the question is approached of whether or not an investor can have information about the current stock price, P_t, that would allow one to gain some return on investment. A naive approach considers $E[P_{t+1} \mid X_t] = P_t$. However, this expectation is false since the process of receiving profit from the raw data is too simple for consistent returns for the average investor. [5] Instead, this research investigates E[$R_t \mid X_{t-h}$] by utilizing a function for the distribution P_t that captures the distribution of a larger portfolio. This function, $\psi(w_{t-h})$, depends on a signal which is assumed to be a random distribution. The distributions tested include: $w_{t-h} \sim (Poisson, Normal, Uniform)$. Out of the these possible parameters for ψ the performance is evaluated using a Cross Validation technique and loss function that is the residual sum of squares and variance between the predicted output of E[$R_t \mid X_{t-1}$] and the actual R_t for the portfolios given in section 2 Data. The best random distributions are selected and then evaluated similarly to different historic value based trading strategies. Each of the strategies are evaluated on different timescales (weekly, monthly, 3 mo.).

For the simulations R and Python were used concurrently in order to produce many sampled distributions for further analysis. This included sampling the size S_n which is the same as selecting a subset of *n* stock from the SP500 index. As well as computing statistics for transformations of the price. In example, the expected return of many bootstrap samples representing the mean of return over a certain period within a simulated portfolio S_n. These distributions are evaluated using confidence intervals assuming a specific distribution.

4 Simulations

A common hypothesis in market trading research is that a stock price can be modelled as a random walk. A random walk is the "instantaneous adjustment" property of an efficient market which implies successive price changes in individual securities will be independent. Most simply the theory of random walks implies that a series of stock price changes has no memory; the past history of such a series cannot be used to predict the future in any meaningful way. The future path of the price level of a security is no more predictable than the path of a series of cumulated random numbers. [1] Therefore, the assumption that individual prices are independent of past prices meets the condition needed for Monte Carlo simulation. Whether or not stock prices are random an investor would seek to understand what the expected value of growth for their investments is likely to be. Hence, an investigation into the confidence intervals of several statistics of various strategies follows in this section.

If the assumption of independent variables is actually met then we would expect for a large sample of prices Sn over some period of time to vary by some random distribution. First to be simulated was the expected value of growth from an initial investment into the SP500 index as a whole. This is referred to as the “Buy and Hold” trading strategy. For this the value of $E[R_t \mid X_{t-h}]$ is calculated as follows:

$start$ = initial investment of \$1000

Inv_t = the value in dollars of start distributed as a margin trade over S_n

$$E[R_t \mid X_{t-h}] = (start / P_t) * (P_t - P_{t-h})$$

Furthermore, compared to a random trading strategy, this Buy and Hold performed best when simulating a large number of portfolios and randomly sampling mean equity over different time periods. Notably, the longer a portfolio is held onto, the greater the expected return of the initial investment regardless how the market was performing at the time. Similarly, random distributions for the parameter of $\psi(w)$ where evaluated using a cross validation where the residual sum of squares is minimized between the starting distribution as a predictor of expected return. If we were to relax the condition that future prices are independent of past prices, we would evaluate the parameters of a distribution which has a good “fit” with historical data. Some other methods to select these parameters are given in section 6 Works Cited.

Buy and Hold Statistics

BOOTSTRAP CONFIDENCE INTERVALS	Based on 10000 bootstrap replicates conf.level = 0.95
Basic (Normal)	(46.59, 50.43)
Percentile	(46.57, 50.41)

Table 1: Simulated difference **E[R_t | X_{t-h}] for 90 days minus 7 days.** With zero not being in this interval one can expect on average to have greater expected returns over longer periods of time.

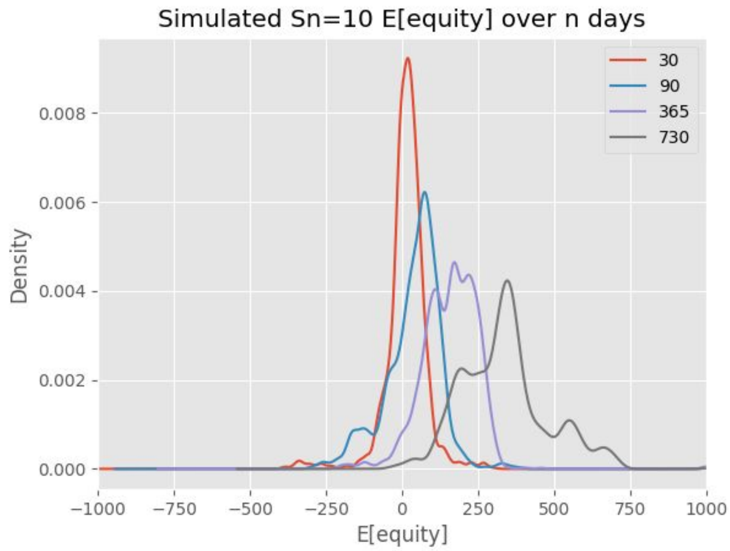
T TEST CONFIDENCE INTERVALS	conf.level = 0.999
Simulated $E[R_t \mid X_{t-h}]$ for 7 days	[4.708226, 4.910179]
Simulated $E[R_t \mid X_{t-h}]$ for 1 year	[207.3741, 208.1462]
Simulated $E[R_t \mid X_{t-h}]$. for 2 years	[425.5643, 426.4563]

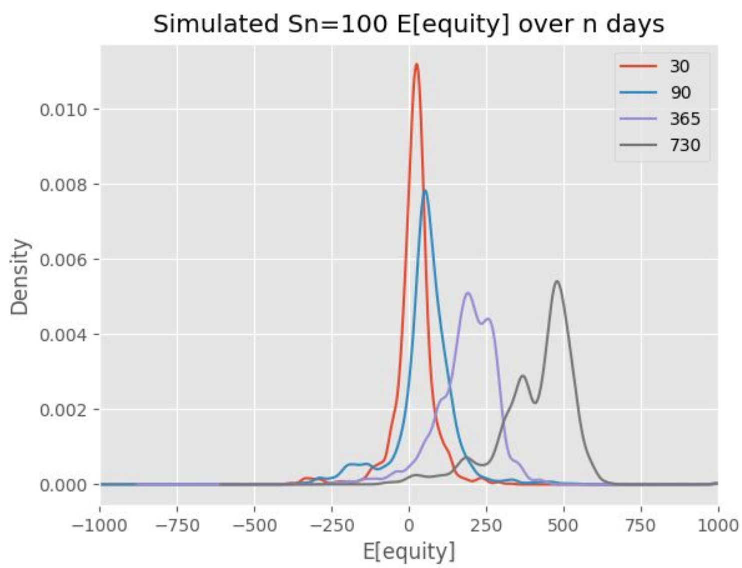
Table 2: Estimates of confidence interval for the value in dollars surplus of initial investment over different periods. Assumes a t distribution.

Random Strategy Statistics

BOOTSTRAP CONFIDENCE INTERVALS	Based on 10000 bootstrap replicates conf.level = 0.95
w ~ Normal	Basic (-1.2723, 1.2704)
w ~ Normal	Percentile (-1.2704, 1.2723)
w ~ Poisson	Basic (-1.2750, 1.2561)
w ~ Poisson	Percentile (-1.2561, 1.2750)

Table 3. Different estimations for $E[R_t \mid X_{t-h}] = \psi(w_{t-h})) * \beta * X_{t-h}$ generated via random distributions of *w*. Observe that the intervals do include zero. So on average, with a large number of these intervals the expected return on investment will include zero in the interval 95% of the time.





Figures: Showing the density plot of simulated portfolio's expected return as equity over different timescales (30 days, 90days, etc). Smaller portfolios resulted in a noisier estimate and the ranges of quartile distributions were greater than larger portfolios.

5 Analysis / Results

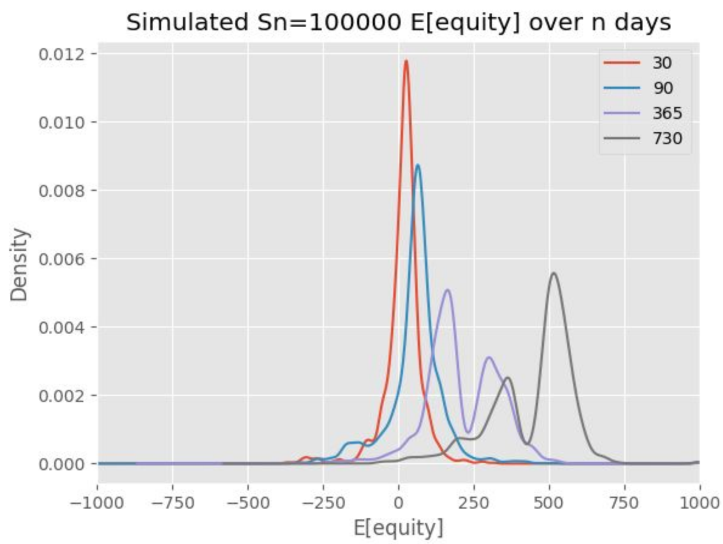


Figure 3: The expected return given a simulated portfolio with 100k randomly selected stocks securities expressed as a density graph. These were randomly "purchased" randomly 10k times at a time interval of n days over 5years of data. This graph shows that holding into a large portfolio more than 3 months, regardless of starting price, will yield returns on investment on average.

Given the risks associated with market trading we can evaluate the expected variance and magnitude of returns given distributions of $\psi(w_{t-h})$ with parameters selected via Section 3 Simulations. We found that by buying and selling given a random signal is similar to gambling. There is only a small chance of making money and the limit of expectation for large samples could approach zero. Among the choices of random strategies and long term holding, Buy and Hold exhibits the greatest expected returns over any time scale, E[$R_t \mid X_{t-h}$]. It was found that if one invested \$1000 they can earn on average, in the percentile range of (46.57, 50.41) more by holding for 90 days as opposed to 7 days and then randomly selling. When holding for 2 years or more, the returns could be as high as $\$425.5 \pm 0.5$.

6 Discussion / Conclusion

In comparison, the idea of random trading given distributions of $\psi(w_{t-h})$ is similar to trading without being properly informed of what might be driving that market price. Given this research, a safer approach for the average investor without access to non-linear algorithms would be to buy as much individual stock of the companies that comprise a large index such as SP500. This large sample resulted in a cumulative distribution that gave returns for the testing in sections 3 and 4 above. Hence, the stock that gave negative return might have been in balance by a majority that grew in price. This initial investment should be scaled to how much the investor can reasonably expect to not spend over a duration of time greater than 90 days. This would be an ideal strategy for anyone that does not know about the underlying properties of the stock market.

7. Works Cited

[1] Fama, Eugene F. "Random Walks in Stock Market Price." *Financial Analysts Journal*, 1995, pg. 75–80., bit.ly/3nwBkyI.

[2] Neftci. S.N. "Naive trading rules in financial markets and Wiener-Kolmogorov prediction theory", *Journal of Business*, 1991, s49-571.

[3] Eugene F. Fama, Marshal E. Blume.. *Journal of Business*, 1996, Issue 1, pg. 226-241

[4] Nelson, C.R. and M.J. Kim "Predictable stock returns: reality or statistical illusion?" Working paper, *Economics Department*, University of Washington, Seattle.

[5] Clive W.J. G-anger "Forecasting stock market prices: Lessons for forecasters" *International Journal of Forecasting*, 1992, p3-13

8 Appendix I: Implementation Details and Data Collection

The code used to generate these results can be found on my GitHub at:

github.com/nathanShepherd/market_determinism