Using Entropy to Measure Risk in US Stock Prices

BMCC Foundation Fund for Undergraduate Research

Latif Fetahaj, Borough of Manhattan Community College

Mentor , Glenn Miller, Mathematics

1 Introduction

In this project we use the concept of entropy, specifically as defined in information theory, to rate various securities in terms of their risk. Entropy quantifies how much is not known in an event, uncertainty. A difference between entropy and information is that entropy measures the unknown, and information is the known part. Information allows us to make a more accurate prediction about a future event, in this case the stock market. We use two approaches to test the data. First, we look for patterns in when the security increased in price (coded as 1) or decreased (coded as 0). In the second approach we tested the data in 5 levels, considering changes in the stock price as a dramatic decrease (coded as 0), slight decrease (1), relatively unchanged (2), slight increase (3), and dramatic increase (4). The first approach did not significantly reduce the entropy compared to a random sequence, whereas the second approach gives us more information making the stock price for certain companies more predictable, that is less risky, than others.

2 Background

2.1 Entropy as Uncertainty in Information Theory

Shannon developed information theory to measure the amount of information produced by a source of binary digits. For example, if a black-and-white video is converted to a string 1's and 0's, there are many long sequences of 1's and 0's and that bit stream could be compressed considerably for transmission. So, in the information theoretical sense, each bit of that stream provides less information than bits from a source which is composed of a random stream of 1's and 0's. Entropy provides a way of measuring the amount of information in a "message."

The basic formula for entropy is

$$H(x_i)=-\sum_{i=0}^{n} p(x_i)\log(p(x_i))$$

where x_i's represent various states (0 and 1 for the binary example above and various intervals of possible changes in securities prices for our project.)

2.2 Conditional Entropy

The conditional entropy of a random variable given the outcome of another random variable is written H(X|Y) and can be used to measure the amount of information (in bits!) that the outcome of Y provides about the possible outcome of X. This can be computed directly, but it is computed more easily using the relationship between the entropy of the joint distribution of X and Y given by Shannon ([2]):

$H(X,Y)=H(Y)+H(X|Y)$ or $H(X|Y)=H(Y)-H(X,Y)$

The first of these makes sense since the information given by X and Y jointly can be partitioned into the information that Y provides and the additional information that X provides given that Y has occurred.

2.3 Entropy of time series data

Market data is in the form of a time series. The entropy in a time series is based on more than just the relative frequencies of the characters, as it is also affected by each observation's dependence on prior observations. Markets are notoriously streaky, so it is important to consider conditional entropy when assessing time series data. There is no standard approach to do this as various authors use slightly different methods for seeking patterns of varying lengths which help to reduce the information provided by each new observation giving some idea of how well the prior performance patterns

predict future changes in the time series.

3 Method

3.1 Algorithms and programs(Python) for Entropy and Information of Time Series Data

When we consider L levels of price change, then we can consider a sequence of prior n days returns as a number expressed in base L. Then convert that base L number to decimal form and increment the element in a matrix with L rows and 2^n columns to keep count of the current day's outcome for each possible three prior day pattern. For example, when we consider a stock going up or down only (two levels) and look at three prior days, the matrix has two rows (the current day's change as 0 or 1) and 7 columns for the possible prior three day patterns that correspond to counting from 0 to 7 in binary: {000, 001, 010, . . . , 111}. Python example below.

```
def pattern_counter(returns,n):
    base = 2
    columns = (base)**(n)

    #create a matrix of L rows and (L^n) columns depending on the prior days
    #n = 1 up to 8 prior days
    pattern_matrix  = np.zeros((base,columns), dtype = int)
    indices = np.arange(n)

    for i in range(0,(returns.size-n)):
        n_array = np.take(returns,indices)

        #convert from the given base to decimal
        num = n_array.dot((base)**np.arange(n_array.size)[::-1])

        #increment the matrix's cell to keep the count for the current day's outcome
        pattern_matrix[returns[n]][num]+=1
        indices = np.add(indices,1)
        n=n+1

    return pattern_matrix
```

3.2 Technical considerations for implementing the algorithm for stock price data : GL Entropy

1 Use log change. $N_i=\log(M_i/M_{i-1})$ (i= 1, 2, 3….M-1)

2 Set standard interval for all stocks (used the minimum and maximum of INTC to set the range

3 Compute conditional entropy looking at 1 prior day, 2 prior days, . . .up to 8 prior days.

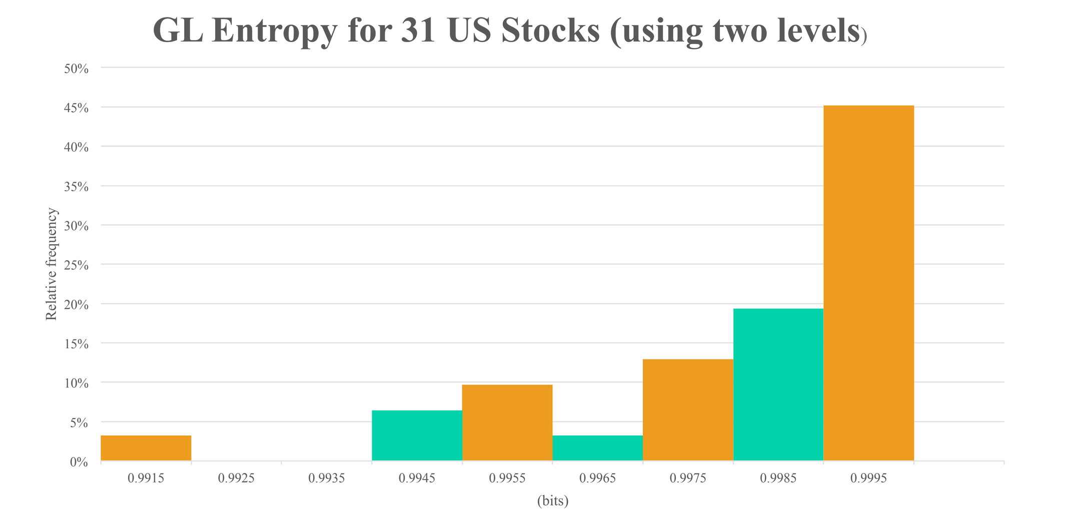
4 Compare those (subtract from) the corresponding conditional entropies for a randomized uniform distribution. Allow for a minimum difference of 0.

5 Compute a weighted average for those using the greatest weight of 8 for the most recent time period (1 prior day) down to a weight of 1 for 8 day prior periods.

6 The result of step 5 is the information GAINED by looking for patterns of varying lengths in the data, so subtract that weighted average from the entropy per daily data unit (LOG base 2 of the number of intervals) to get the entropy or uncertainty.

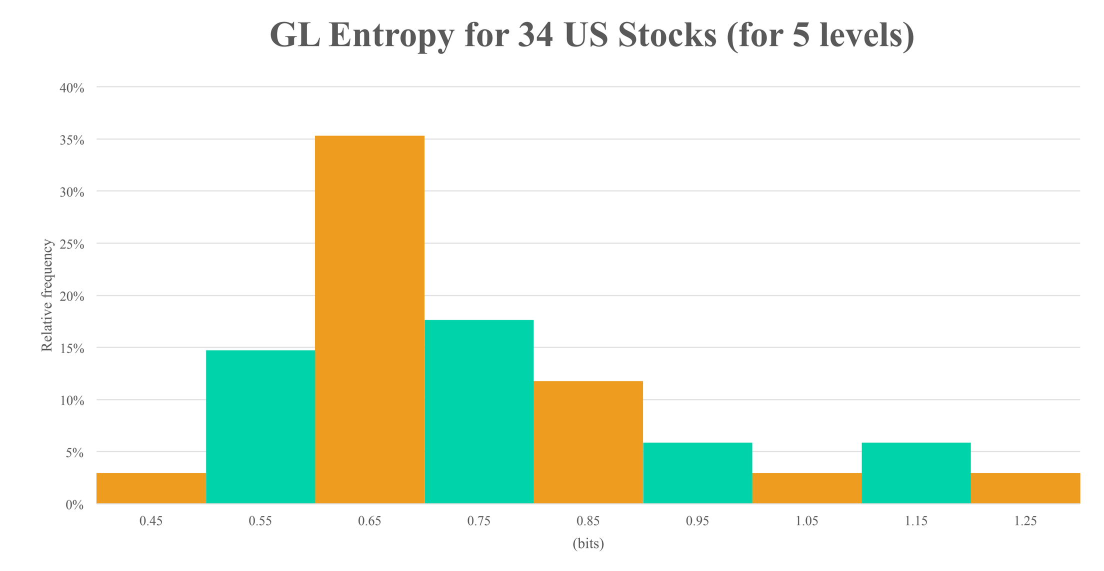
4 Results

Using only information about whether the stock was "up" or "down" yielded a mean GL entropy of 0.9981 bits with a standard deviation of 0.0004 bits. So, there was not much information gained by studying patterns based on just streaks of up and down days and not much variation between the various companies.



However, using the levels in the log change of the stock prices provided more information with more differentiation between companies. The mean GL entropy and standard deviation are 0 .75 and 0.19 bits respectively…The GL entropies computed did rank many companies as one might expect: stodgy industrial companies such as MMM and IBM were assessed as low risk.

The S&P 500 was one of the least risky as one would expect for its diversified portfolio. And younger tech companies did generally have among the highest values for GL Entropy (NFLX and TSLA) There were some surprises such as an airline company (DAL) being among the most risky investments.



5 Conclusion

Using conditional entropy to find the information in time series data is not a standardized field. Various authors have made a variety of attempts to quantify the risk (or "predictability") of financial products using entropy. Many of them do not test against random ordering of the data set to ensure that their methods are capturing the information about the patterns in the time series that they are purporting to find. Our contribution is to incorporate that test inside the algorithm for computing entropy (as information gained over a random ordering.)

Acknowledgements

This research was supported by BMCC Foundation Fund. I thank my mentor, Dr. Miller Glenn, who has guiding me through this research

References

[1] Adami C. What is information?†. *Philos Trans A Math Phys Eng Sci*. 2016;374(2063):20150230. doi:10.1098/rsta.2015.0230

[2] Shannon, CE–Weaver, and Warren Weaver. "W.:(1949) The Mathematical Theory of Communication." *Press Uol, editor* (1948).

[3] Stone, James V. *Information Theory: A Tutorial Introduction.* Sebtel Press, 2015