Lab4

Cantor Alexandru Horatiu

2 November 2023

1 Data modelling

2 Modeling the Experimental Part

In this section, we discuss the modeling of the experimental part, focusing on key aspects such as the choice of data, selection of experiments, validation procedures, and methods for comparing results.

2.1 Data Selection

The data utilized in this paper comprises historical data for SPX (S&P 500), spanning from the year 1927 to November 2023. The data is collected on a daily basis and sourced from Yahoo Finance. Specifically, the algorithm is applied using the daily closing prices, which have been prefiltered to exclusively include this closing price data. It's noteworthy that the algorithm demonstrated in this paper is versatile and can be applied to any stock or index.

2.2 Experiments and Validation

The experiments conducted involved the analysis of daily data from the S&P 500 and other prominent indexes. Validation procedures were executed by predicting the price for the subsequent day based on a subset of these indexes and subsequently comparing the predictions with the actual results. Notably, the accuracy of the tests ranged between 52% and 57%, surpassing the probability of a mere coin flip. This outcome suggests a small but noteworthy performance, warranting careful consideration in future applications.

2.3 Advantages Over Existing Approaches

When comparing our approach to existing methodologies, distinct advantages emerge. In the current landscape, prevalent strategies for entropy calculation often hinge on the entire dataset, giving precedence to recent prices with augmented probabilities. In contrast, our approach introduces a fundamental shift in perspective.

In our methodology, every individual price point is afforded the same probability weight. This deliberate departure from assigning varying probabilities based on recency fundamentally alters the conventional linearity considerations associated with price sequences. Rather than treating prices as a linear progression, our approach emphasizes their collective nature as a set.

By eschewing the conventional weighting schemes favoring recent prices, our model captures a more egalitarian representation of the entire dataset. This not only introduces a nuanced understanding of the price dynamics but also opens avenues for uncovering patterns and insights that may be overshadowed by approaches emphasizing recent trends.

This nuanced handling of probabilities and a departure from the linearity-centric mindset contribute to the distinctive strengths of our approach, presenting a compelling case for its adoption in the realm of entropy-based stock analyses.

2.4 Mathematical Models

The algorithm underpinning this study draws inspiration from Sinai entropy [12], employing entropy calculations in two distinctive ways, each contributing to a nuanced understanding of price dynamics.

1. $h(T) = \sup_\xi h(T, \xi)$: This expression represents the supremum of entropy over all possible partitions, encapsulating the highest achievable entropy value within the system.

2.

$$h(T, \xi) = -\lim_{n\to\infty} \frac{1}{n} \sum_{i_1,\ldots,i_n} \mu(T^{-1}C_{i_1} \cap \ldots \cap T^{-n}C_{i_n}) \log \mu(T^{-1}C_{i_1} \cap \ldots \cap T^{-n}C_{i_n}).$$

The entropy theorem posits the equality of these two formulas when ξ serves as a generating partition. Consequently, the algorithm assesses market dynamics by evaluating whether prices are likely to ascend or descend, emphasizing the significance of a generating partition in this determination.

Recognizing the potential limitations in representing the limit exactly, the algorithm introduces a pragmatic approach. Instead of relying on a singular value, the mean average of six values from the entropy subset is considered. A pivotal decision is then made: if the square of the supremum exceeds the mean entropy subset value, it signifies a probable increase in the price.

This intricate interplay of mathematical formulations and practical considerations underscores the algorithm's sophistication in forecasting price movements and contributes to its robust analytical capabilities.

3 Initial Case Study

In this phase of our research, we embark on a series of experiments centered around a smaller, artificially generated dataset. The primary objective is to vividly showcase the methodology and potential inherent in the proposed approach.

3.1 Experimental Design

The chosen dataset for this initial case study is deliberately compact, crafted or selected with precision to offer controlled conditions for rigorous experimentation. This deliberate choice enables us to delve deeply into the intricacies of the proposed approach, highlighting its functionality in a more constrained environment.

3.2 Dataset Overview

The experimentation involves subjecting the approach to a carefully curated dataset, denoted as $[1, 2, 3, 4, \ldots, 100]$.

3.3 Performance Metrics

The accuracy of these predictions is a key performance metric. Impressively, the approach achieves a commendable accuracy rate of 60%, attesting to its reliability in discerning patterns and trends within the dataset.

3.4 Code Reusability

Noteworthy is the fact that the same codebase utilized for experiments on the S&P 500 dataset is seamlessly applied to this smaller dataset. This not only underscores the versatility of the approach but also highlights the consistency in its performance across datasets of varying sizes and complexities.

3.5 Insights for Research Report

The outcomes of this initial case study give us hope that the proposed algorithm might give valuable insights into a real-world example.

This comprehensive case study sets the stage for subsequent phases of the research, laying a solid foundation for further exploration and refinement of the proposed approach.

4 Related Work

In the realm of stock market forecasting, our paper presents a pioneering approach by introducing Sinai entropy to the domain. This novel methodology leverages the disparity between two entropy values to predict the movement of a given dataset, exemplified by a test subset and rigorously tested on the entire S&P 500 dataset, achieving a commendable 55% accuracy. This integration of Sinai entropy adds a unique dimension to our predictive model. In contrast, Liu[9] work delves into well-established statistical learning models, such as Logistic Regression, Gaussian Discriminant Analysis, Naive Bayes, and Support Vector Machine (SVM), specifically optimizing the SVM model with a Radial Basis Function (RBF) kernel for forecasting the S&P 500 index. Their study encompasses a narower dataset from 01/01/2004 to 12/31/2014, concluding that the SVM model with an RBF kernel attains an impressive 62.51% accuracy for predicting the future market trend of the S&P 500 index. Both papers contribute valuable insights, with our approach offering a distinctive perspective rooted in entropy, introducing the novel application of Sinai entropy, while Liu's work provides a comprehensive analysis of well-established machine learning models.

5 Github

`https://github.com/x64alex/Research`

6 Citations

[1] [12] [6] [5] [3]
[13] [10] [15] [14] [4]
[11] [7] [16] [2] [8]

References

[1] Sonia R Bentes and Rui Menezes. Entropy: A new measure of stock market volatility? *Journal of Physics: Conference Series*, 394(1):012033, nov 2012.

[2] John M. Cozzolino and Michael J. Zahner. The maximum-entropy distribution of the future market price of a stock. *Operations Research*, 21(6):1200–1211, 1973.

[3] Aram Ebtekar. The algorithmic second law of thermodynamics, 2023.

[4] Rongbao Gu. Multiscale shannon entropy and its application in the stock market. *Physica A: Statistical Mechanics and its Applications*, 484:215–224, 2017.

[5] Martin Hilbert and David Darmon. How complexity and uncertainty grew with algorithmic trading. *Entropy*, 22(5), 2020.

[6] Aleksander Jakimowicz. The role of entropy in the development of economics. *Entropy*, 22(4), 2020.

[7] S. Sohn K. Ahn, D. Lee and B. Yang. Stock market uncertainty and economic fundamentals: an entropy-based approach. *Quantitative Finance*, 19(7):1151–1163, 2019.

[8] Ladislav Kristoufek and Miloslav Vosvrda. Measuring capital market efficiency: Long-term memory, fractal dimension and approximate entropy. *The European Physical Journal B*, 87:art. 162, 07 2013.

[9] Chongda Liu, Jihua Wang, Di Xiao, and Qi Liang. Forecasting sp 500 stock index using statistical learning models. *Open Journal of Statistics*, 06:1067–1075, 01 2016.

[10] Joanna Olbryś and Krzysztof Ostrowski. An entropy-based approach to measurement of stock market depth. *Entropy*, 23(5), 2021.

[11] Joanna Olbryś and Krzysztof Ostrowski. An entropy-based approach to measurement of stock market depth. *Entropy*, 23, 2021.

[12] Y. Sinai. Kolmogorov-Sinai entropy. *Scholarpedia*, 4(3):2034, 2009. revision #91407.

[13] Gaurang Sonkavde, Deepak Sudhakar Dharrao, Anupkumar M. Bongale, Sarika T. Deokate, Deepak Doreswamy, and Subraya Krishna Bhat. Forecasting stock market prices using machine learning and deep learning models: A systematic review, performance analysis and discussion of implications. *International Journal of Financial Studies*, 11(3), 2023.

[14] Claudiu Vinte and Marcel Ausloos. The Cross-Sectional Intrinsic Entropy. A Comprehensive Stock Market Volatility Estimator. Papers 2205.00104, arXiv.org, April 2022.

[15] Claudiu Vințe and Marcel Ausloos. The cross-sectional intrinsic entropy—a comprehensive stock market volatility estimator. *Entropy*, 24(5):623, apr 2022.

[16] Zang Yeze and Wang Yiying. Stock price prediction based on information entropy and artificial neural network. In *2019 5th International Conference on Information Management (ICIM)*, pages 248–251, March 2019.