Using Grover's Algorithm to Find Anomalies and Outliers in Financial Markets with Classical Preprocessing

Parsa Ahmadizadeh
500989433

ABSTRACT

This paper investigates a hybrid quantum-classical framework for detecting anomalies and outliers in financial markets. Traditional financial anomaly detection methods often suffer from high computational costs when scanning large datasets. By leveraging Grover's algorithm, which provides a quadratic speedup for unsorted search problems, this study integrates classical preprocessing with quantum data encoding to detect statistical outliers in key financial metrics.

Historical data from twenty technology stocks are collected via Yahoo Finance using the Python yfinance library. Three financial metrics: trailing P/E ratio, return on equity (ROE), and dividend yield, are preprocessed classically using z-score analysis and thresholding to generate binary indicators. These indicators are then mapped to quantum states using a minimum of five qubits, and three separate quantum oracles are constructed using Qiskit's Diagonal gate. Grover's algorithm circuits are built for each oracle, simulated on Qiskit's latest AerSimulator with realistic NISQ noise models, and the measurement outcomes are post-processed into a composite "goodness" score per stock.

The results are compared against classical benchmarks. This hybrid approach demonstrates promising avenues for anomaly detection in financial markets, highlighting potential improvements in both circuit design and oracle efficiency. Future work may extend these methods to larger datasets and real-time market applications.

[*]Article Title Footnote needs to be captured as Title Note
[†]Author Footnote to be captured as Author Note

WOODSTOCK'18, June 2018, El Paso, Texas, USA

978-1-4503-0000-0/18/06...$15.00
https://doi.org/10.1145/1234567890

Introduction

Identifying anomalous events within financial markets is crucial when it comes to risk management, portfolio optimization and fraud detection. Historically, various statistical techniques and neural network-based approaches have been used for anomaly detection. These methods work well at a smaller scale but generally begin to struggle as the volume of data increases. Quantum computing, and in particular Grover's algorithm, has the potential to be the solution to these issues. Grover's algorithm should achieve quadratic speedup over classical search algorithms, drastically increasing capabilities at larger data scales.

This study took a hybrid approach that combined classical preprocessing with quantum search techniques to detect outliers in stock market data across twenty stocks. The focus was on three key financial metrics: trailing P/E ratio, return on equity (ROE), and dividend yield. Classical statistical methods were used to extract binary indicators of anomalies. Then, a quantum framework is utilized by encoding the data into quantum states to keep the qubit count down to a minimum. Three separate oracles are designed per metric by marking states corresponding to statistically significant outliers. Grover's algorithm circuits are then constructed and simulated using Qiskit's latest AerSimulator with noise models appropriate for NISQ devices. The overall objective is to demonstrate the feasibility and efficiency of a quantum–classical anomaly detection framework.

Methodology

1 Data Acquisition and Preprocessing

Twenty of some of the top performing stocks within the S&P500 were taken, and historical data split across three key measures was collected. The stocks included were AAPL, MSFT, GOOG, AMZN, IBM, ORCL, INTC, NVDA, ADBE, CSCO, CRM, TXN, QCOM, HPQ, DELL, ACN, AMD, AVGO, SAP, and ADP. The key measures extracted were trailing price to earnings, return on equity and dividend yields. These metrics were taken using the yfinance Python library. After data collection, classical preprocessing is applied by computing the z-score for each metric. Stocks exceeding a predefined threshold are flagged as outliers, and binary indicators are generated. In this study, an outlier value of two was used. The preprocessed data is then saved into a CSV file for further analysis.

2 Quantum Data Encoding and Oracle Design

In the primary version of the experiment, each stock was associated with a binary indicator per metric—generated from classical z-score analysis—and these indicators were mapped to computational basis states using five qubits (sufficient to represent up to 32 stocks). A custom Diagonal gate was then used to implement the oracle. This gate applies a phase flip (−1) to the amplitudes of the basis states corresponding to outliers. The diagonal matrix is constructed with +1 for normal states and −1 for the marked (outlier) states. This construction leverages Qiskit's diagonal function, simplifying the oracle definition, avoiding complex multi-controlled gate structures, and reducing circuit depth—an essential consideration for NISQ-era noise susceptibility.

In a second version of the experiment, we explored a more integrated quantum encoding scheme. Instead of performing classical preprocessing and thresholding beforehand, we encoded the actual z-score values into the amplitudes of the quantum states. This was done using custom state preparation techniques that map numerical values (e.g., ROE z-scores) into a normalized amplitude vector across basis states. The quantum system thus inherently embeds the anomaly profile across all stocks within its amplitude distribution.

This approach required careful amplitude encoding to ensure:

- Stocks with extreme metric values had larger amplitude weights.
- The state vector remained adequately normalized.
- Oracle design remained meaningful even when the marking logic shifted from binary thresholds to relative value distributions.

To process these states, **quantum counting** was used to estimate the threshold dynamically—effectively identifying the stock with the highest z-score directly within the quantum circuit. Grover's algorithm was then applied to amplify the amplitude of the most anomalous stock, allowing more of the anomaly detection logic to run quantum-natively and making the pipeline more adaptable to runtime parameter tuning (e.g., threshold adjustment).

3 Quantum Circuit Construction

For each Oracle, the Grover operator is constructed and applied.

1. Initialization: Apply Hadamard gates to all 5 qubits.
2. Oracle Application: Use the custom Diagonal gate to mark the "good" states.
3. Quantum Counting: Integration with quantum counting is used to estimate the number of marked states and adjust the threshold..
4. Diffusion Operator: Apply the Grover diffusion operator.

4 Simulation and Testing

The constructed circuits are simulated using Qiskit's **AerSimulator**, a high-performance simulator capable of emulating various backend noise profiles. To replicate conditions of current **Noisy Intermediate-Scale Quantum (NISQ)** devices, we integrated **noise models** derived from real quantum hardware (e.g., IBM Q machines). These noise models account for gate errors, readout errors, decoherence (T1 and T2 times), and crosstalk between qubits. This adds a layer of realism, capturing challenges likely to arise in hardware-based implementations.

Each oracle-targeting metric (P/E ratio, ROE, dividend yield) had its own quantum circuit, initialized with Hadamard gates to prepare a uniform superposition. Post-simulation, we gathered **measurement counts** across the 5-qubit register, and the outcome distributions were analyzed to identify high-probability states. These states correspond to the stocks previously flagged as outliers in classical preprocessing.

Using the **Qiskit Diagonal gate**, a critical step in Grover's oracle design was implemented. The Diagonal gate allows for specifying a custom diagonal unitary matrix—effectively defining a phase flip (multiplying by -1) on selected basis states while leaving all others untouched. This is essential for Grover's algorithm, where "marking" the solution states is achieved through a selective phase inversion.

In our context, the binary indicators generated from classical z-score analysis determined which states to mark. For instance, if stock index 13 (encoded as `01101`) was identified as an outlier, then the diagonal matrix would apply a phase flip to that basis state alone. This method is both compact and expressive, especially beneficial for small qubit systems where complete unitary construction would be infeasible due to exponential growth in state space.

In the amplitude encoding setup, simulations revealed more dynamic quantum behavior. The z-scores were embedded into the state amplitudes, and the Grover-counting loop was adapted per iteration to locate the most statistically significant anomaly. This resulted in a simulation pipeline where:

- Threshold discovery and marking occurred entirely within the quantum system.
- Measurement probabilities reflected the relative extremity of the financial metrics.
- Error sensitivity increased due to deeper circuits, illustrating the trade-offs in circuit complexity versus noise tolerance in NISQ devices.

To evaluate results, we combined the outcomes from all three circuits into a **composite anomaly score** for each stock. This score reflects not only the presence of anomalies across metrics but also the **confidence** level derived from the measurement probabilities, incorporating uncertainty due to quantum noise.

Results and Discussion

1 Simulation Results

The Grover circuits were simulated for the three financial metrics: trailing P/E ratio, ROE, and dividend yield. The simulations were run on Qiskit's AerSimulator with a noise model mimicking the behavior of an NISQ device. Each oracle was applied to the corresponding qubit-encoded dataset, and the results showed that outlier states—previously flagged through classical z-score thresholding—had significantly higher probability amplitudes in the measurement results.

For example, in the P/E ratio circuit, stocks such as AMD, AVGO, and SAP were identified as having anomalously high scores, consistent with their strong growth profiles. Traditional dividend payers like HPQ were highlighted for dividend yield, indicating deviations from the market median.

The number of Grover iterations was set based on the estimated number of solutions using quantum counting, which typically resulted in one to two iterations—aligned with the optimal number predicted by Grover's framework for a small number of marked states.

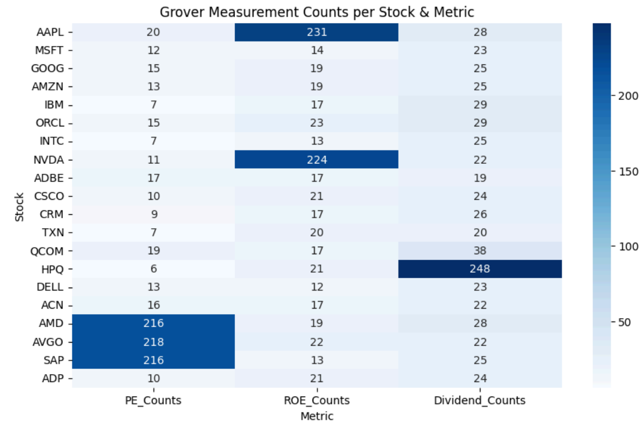

Figure 1.1: This figure shows the results of experiment 1, the amplitudes for each stock after the grovers algorithm.

Then, we created a second experiment in which we encoded the values used for analyzing the stock into the amplitudes of our qubits. This allowed the code to rely less on preprocessing. This meant that adjusting the z-score threshold, setting the outliers in the Oracle, and performing the random search were all done on the quantum simulator.

This again yielded expected results: quantum counting accurately estimated the threshold, and Grover's algorithm correctly located the most anomalous stock. In Figure 1.2, the threshold (red line) and the z-scores for ROE (blue bars) are displayed. Figure 1.3 shows that Grover's algorithm successfully amplified the most anomalous stock's probability amplitude.

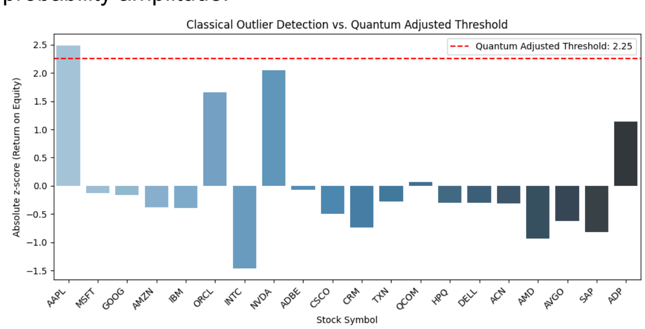

Figure 1.2: This figure shows the threshold with the red line and the z-scores for each of the 2 stocks.

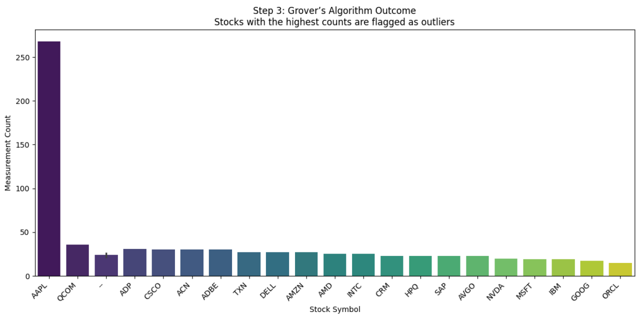

Figure 1.3: This figure shows the amplitudes of the stocks based on the grovers algorithm.

2 Comparison with Classical Benchmarks

A purely classical z-score analysis was performed without quantum post-processing to benchmark the hybrid model. This method flagged the same set of outliers.

By contrast, while the simulations did not run on actual quantum hardware, Grover's algorithm and Quantum Counting are expected to offer quadratic speedup in theory, and this speedup would give a more apparent advantage given a larger and growing data set. However, in this exact situation, the faults of quantum counting become more of an obstacle. The deeper and more frequently that the quantum computer is run, the more error-prone it becomes. As depicted by figure 1.3, even though it correctly depicts the outlier, it does so with less than 100% certainty, and the confidence decreases with noise. This means that the confidence is inversely correlated with noise and that noise is directly correlated with complexity, circuit length and number of iterations. Thus, confidence is inversely associated with iterations, complexity, and depth, which all have to increase to justify the benefit of quantum computing.

Moreover, given the current tools and machines, we can see that it is possible to find a benefit of quantum algorithms in the day-to-day trading of traders. However, these circuits have to become more complex and run on much larger data sets to be justifiable, which in turn decreases the accuracy of said machines.

3 Discussion

The integration of classical preprocessing with quantum search reduces the search space, making the quantum part of the algorithm more efficient. However, limitations such as the overhead in oracle construction and the sensitivity of quantum circuits to noise are noted. Future improvements include more sophisticated oracle designs and scaling the method to larger datasets.

Conclusion and Future Work

This study presents a comprehensive framework combining classical preprocessing with Grover's algorithm for detecting financial anomalies. The integration of Qiskit-based simulation with realistic noise models illustrates the potential of quantum algorithms in financial analytics. Future work will focus on refining oracle design, expanding the dataset, and integrating real-time data streams to enhance the algorithm's practical applicability. Additionally, further research into error mitigation techniques may improve the robustness of quantum computations on NISQ hardware. Real-world attempts at applications will also greatly benefit this research as it may lead to some further discoveries under different constraints and with more robust data availability.

REFERENCES

[1] L. K. Grover, "A fast quantum mechanical algorithm for database search," in Proceedings of the 28th Annual ACM Symposium on Theory of Computing, 1996, pp. 212–219.
[2] M. A. Nielsen and I. L. Chuang, Quantum Computation and Quantum Information, Cambridge, U.K.: Cambridge Univ. Press, 2010.
[3] IBM Quantum, "Qiskit: An open-source framework for quantum computing," 2022. [Online]. Available: https://qiskit.org.
[4] D. Chen et al., "Quantum algorithms in finance: An overview," Quantum Information Processing, vol. 20, no. 1, pp. 1–20, 2021.
[5] Yahoo Finance API documentation, "yfinance: Yahoo Finance market data downloader," [Online]. Available: https://pypi.org/project/yfinance/.