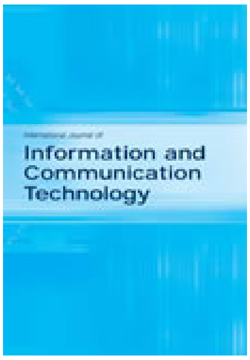


International Journal of Information and Communication Technology

ISSN online: 1741-8070 - ISSN print: 1466-6642
https://www.inderscience.com/ijict

Quantum computing-driven portfolio optimisation framework for the intelligent economy

Leihua Ai

DOI: 10.1504/IJICT.2026.10080663

Article History:

Received:	13 May 2026
Last revised:	13 June 2026
Accepted:	19 June 2026
Published online:	08 September 2026

Quantum computing-driven portfolio optimisation framework for the intelligent economy

Leihua Ai

Ganxi Vocational College of Science and Technology,
Xinyu, 338025, China
Email: aileihua2026@sina.cn

Abstract: Portfolio optimisation in intelligent economic systems faces the challenge of portfolio explosion caused by expanding asset scales, making it difficult for classical solvers to obtain high-quality solutions within a limited timeframe. This paper proposes a hybrid quantum-classical optimisation framework that combines the global search capabilities of quantum annealing with the constraint expression capabilities of variational quantum algorithms, while reducing quantum resource requirements through asset graph partitioning. In experiments covering 20 to 120 real-market assets, the hybrid framework achieved an area under the curve of 0.716 for a portfolio of 120 assets, representing an approximately 29% improvement over classical genetic algorithms; it maintained an area under the curve of 0.921 even under a noise level of 10^{-3}, demonstrating greater robustness than pure quantum annealing methods. The research confirms that the hybrid paradigm is a viable approach for solving large-scale combinatorial optimisation problems on current noisy quantum devices.

Keywords: quantum computing; portfolio optimisation; smart economy; hybrid quantum framework; asset selection.

Reference to this paper should be made as follows: Ai, L. (2026) 'Quantum computing-driven portfolio optimisation framework for the intelligent economy', *Int. J. Information and Communication Technology*, Vol. 27, No. 98, pp.94–114.

Biographical notes: Leihua Ai is a Professor at Ganxi Vocational College of Science and Technology, China. He obtained a Master's degree from the School of Business Administration of French public University of Poitiers (2007). He published ten papers and one core journal paper. His research interests include business management, marketing and international finance.

1 Introduction

Over the past decade, the digital and intelligent transformation of the global economy has profoundly changed the operation mode of the financial market. From the microsecond-level decision-making in high-frequency trading to the global optimisation of cross-border asset allocation, the intelligent evolution of the economic system has presented unprecedented challenges to the computing paradigm. In this context, portfolio optimisation, as a core aspect of financial decision-making, has not lost its significance over time. Instead, due to the rapid expansion of asset types and the increasing complexity of constraints, it has become even more challenging (Hatcher, 2024). Since Markowitz proposed the mean-variance framework in 1952, seeking the optimal balance between risk and return in investment portfolio allocation has always been the core proposition of financial engineering. However, when the problem expands from the ideal continuous space to include integer constraints, trading thresholds, and industry diversification and other real limitations, its computational complexity increases exponentially. Classic solvers often get stuck in the combinatorial explosion of the search space when dealing with portfolio optimisation of hundreds of assets, making it difficult to obtain high-quality solutions within a practical acceptable time window (Liu, 2026). For N = 120 and K = 30, the number of possible portfolios is C(120,30) ≈ 5.5×10^28, which is far beyond the reach of exhaustive search. Classical branch and bound solvers also face exponential runtime in the worst case; even with pruning, practical time limits force early termination and suboptimal solutions. This combinatorial explosion is precisely why heuristic and quantum methods are pursued. A single sentence noting this astronomical number would help readers appreciate the scale of the problem.

The root cause of this predicament lies in the discrete nature of portfolio selection. Traditional methods either sacrifice the optimality of the solution by relaxing integer constraints or sample within the solution space through heuristic search (Wang et al., 2022). The former sacrifices the optimality of the solution, while the latter is difficult to ensure convergence stability. As the asset scale exceeds the hundreds of thousands level, the performance of these methods deteriorates significantly (Han, 2025). The characteristics of the intelligent economic system further exacerbate this contradiction: the frequency of decision-making has increased significantly, the constraint conditions are highly coupled in multiple dimensions, and individualised demands are constantly emerging, making the traditional 'one-shot optimisation' approach unsustainable.

The emergence of quantum computing offers entirely new possibilities for solving such combinatorial explosion problems (Palmer, 2025). By leveraging quantum superposition states to explore the solution space in parallel, quantum algorithms are expected to achieve acceleration over classical methods in specific optimisation tasks. However, the limitations of current noisy intermediate-scale quantum (NISQ) devices impose realistic constraints on this prospect – limited qubit numbers (Toshio et al., 2025), short coherence times, and insufficient gate operation fidelity, which restrict the practical application of purely gate-based quantum algorithms to problems of practical scale (Chatterjee et al., 2025). Quantum annealing paths provide an alternative route, but they still lack flexibility in problem modeling and handling of high-order constraints. Both have their advantages and disadvantages, and a single paradigm is insufficient to cover the diverse optimisation needs in intelligent economic scenarios (Yang et al., 2025).

The core objective of this paper is to explore an investment portfolio optimisation framework that can operate effectively under the current constraints of quantum hardware, without assuming quantum supremacy. This framework is called the hybrid quantum-classical portfolio optimisation (HQCPO). It combines the global exploration ability of quantum annealing with the constraint expression capability of gate-based variational quantum algorithms (VQAs), and achieves resource-sensitive problem-solving through a hierarchical architecture. The contributions of this paper are mainly reflected in three aspects:

- At the methodological level, a hybrid framework combining asset graph partitioning and Pauli encoding is proposed, which reduces the cost of asset correlation while retaining statistical information. This is a pioneering systematic integration.
- At the level of experimental evaluation, establish multi-dimensional unified evaluation indicators, such as: accuracy (ACC), precision , recall , area under the curve (AUC), normalised discounted cumulative gain (NDCG), to fill the fragmentation gap in quantum finance evaluation and provide a horizontal benchmarking scale.
- At the practical insight level, the noise experiment reveals the complementary failure mode of the gate type and annealing (gate error/scheduling sensitivity), providing a theoretical basis for the hybrid framework.

The structure of the remaining part of this article is as follows. Section 2 reviews the quantum paradigm and optimisation requirements; Section 3 elaborates on the formalisation and algorithm of HQCPO; Section 4 explains the experimental design; Section 5 reports the results of classification, extension, noise, etc.; Section 6 discusses the sources and limitations of performance; Section 7 summarises and looks forward.

2 Literature review

2.1 The evolution and current state of quantum computing paradigms

The theoretical foundation of quantum computing can be traced back to Feynman's proposal of the quantum simulation concept. However, it was the advent of Shor's factoring algorithm and Grover's search algorithm that truly drew widespread attention from the industrial sector (Ansari et al., 2025). These two algorithms respectively demonstrated the theoretical acceleration potential of quantum computing in number theory problems and unstructured search tasks. Over the next two decades, the research on quantum algorithms underwent a paradigm shift from theoretical exploration to practical considerations. Especially since the late 2010s, with enterprises such as international business machines (IBM), Google, and Rigetti launching programmable quantum processors, the era of NISQ computing officially began. In this new stage, VQAs have become the dominant paradigm. By approximating quantum circuits as parametric function approximators and forming a closed-loop iteration with classical optimisers, they effectively avoid the decoherence problem faced by deep quantum circuits. Among them, the most representative is the quantum approximate optimisation algorithm (QAOA), proposed by Farhi (Shaydulin et al., 2021). Its core idea is to encode the objective function of combinatorial optimisation problems into the Ising Hamiltonian and approximate the ground state by alternately applying the problem Hamiltonian and the mixing Hamiltonian. At the same time, quantum annealing, as another implementation path, finds the ground state through physical processes rather than digital circuits. It has achieved commercial deployment of more than five thousand quantum bits on the D-Wave systems inc (D-Wave) system (Díez-Valle et al., 2024). Each has its own strengths: QAOA can theoretically approximate any ACC when the circuit depth is sufficient, but is limited by the current gate fidelity; quantum annealing can achieve large-scale parallel search, but is less flexible when expressing high-order constraints.

2.2 Significant advances in quantum portfolio optimisation

Portfolio optimisation is almost the most studied optimisation problem in the field of quantum finance. Early work mainly focused on mapping the Markowitz mean-variance model into the quadratic unconstrained binary optimisation (QUBO) form, and then solving it using quantum annealing or QAOA. The early experiments by the Stanford and D-Wave teams demonstrated the feasibility of this mapping in principle, but the quantum resource requirements rose rapidly when the asset size exceeded 50 (Shao et al., 2025). Most previous studies tested on asset sets of up to 30–40 assets due to qubit limitations. For example, early D-Wave experiments used ≤20 assets, and the original QAOA implementations rarely exceeded 30 binary variables. The thermal-start method by Schlütter was validated on 40-asset problems. The Pauli encoding work reached 250 variables but used a different problem structure. The current work systematically covers 20–120 assets on real market data, explicitly addressing the scalability gap. The progress in the past two years has been particularly significant. Schlütter proposed a thermal-start quantum portfolio optimisation method, which significantly reduced the required quantum bits by constraining the search space to discrete solutions within the neighbourhood of the continuous optimal solution. Experiments on the D-Wave annealer showed that this method outperformed existing techniques. Almost simultaneously, the joint team of IBM and Vanguard explored the application of sampling-based VQAs in portfolio construction, achieving a relatively low relative solution error on the IBM Heron processor with 109 quantum bits. Morapakula proposed an end-to-end quantum annealing pipeline, including continuous mean-variance solving, QUBO discrete asset selection, and quarterly rebalancing mechanism, and achieved competitive performance in real-time comparative tests on the Indian stock market. In terms of handling large-scale problems, the Pauli associated encoding method solved portfolio optimisation problems with over 250 variables by assigning multiple variables to a single quantum bit. These methods have their own advantages and disadvantages: The thermal-start method reduces the quantum resource requirements but highly depends on the quality of the initial continuous solution; The end-to-end pipeline engineering has good integrity but lacks integration of different quantum paradigms; The Pauli encoding method has high compression efficiency but does not fully consider the nonlinear coupling patterns among assets (Jiangtao et al., 2026).

2.3 Characteristics of optimisation demands in an intelligent economic system

Understanding why an intelligent economic system requires special optimisation solutions lies in recognising the fundamental differences between it and the traditional economic system. The significant increase in decision-making frequency demands that the optimisation algorithm complete the entire closed loop from constraint modelling to strategy generation within an extremely short period. The high-dimensional coupling characteristics of constraints make the correlations between assets no longer limited to linear correlations, but rather intertwined through multiple mechanisms such as supply chain networks, investor sentiment propagation, and macro policy transmission. The intensification of individualised demands leads to each investor facing differentiated risk preferences and constraints, rendering the 'one-size-fits-all' approach ineffective. The current mainstream classical optimisation methods – whether interior point methods, branch and bound methods, or heuristic searches – are unable to simultaneously meet these three requirements. The value of quantum methods lies precisely in their parallel exploration capabilities and the essence of probabilistic optimisation: instead of traversing the entire solution space, they focus on high-probability regions through the amplitude amplification mechanism of quantum states (Zhou et al., 2025). However, the realisation of this advantage is highly dependent on the characteristics of the problem – when the search space exhibits good structure rather than complete randomness, the quantum acceleration effect is most significant. Existing research has pointed out that the set of asset correlations in financial optimisation problems often exhibits a block low-rank structure, which provides favourable conditions for the structured search of quantum algorithms.

2.4 Identified research gaps

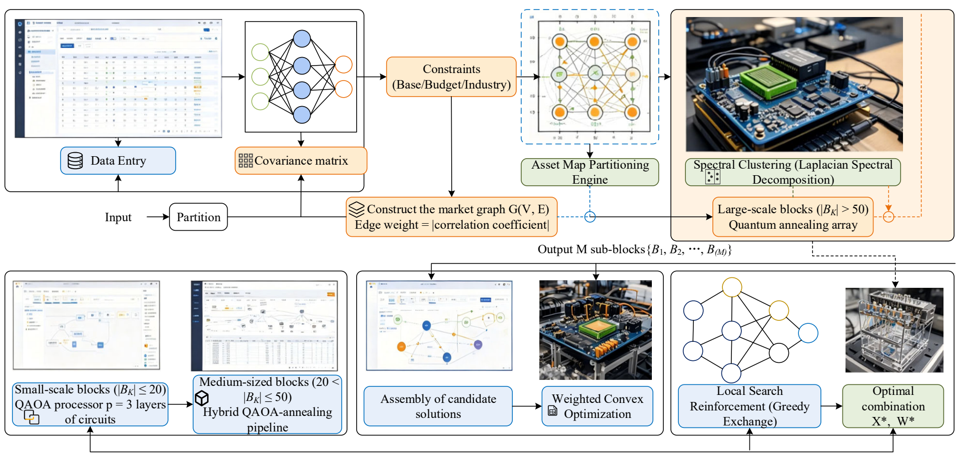
Although a large amount of work has been dedicated to the exploration of quantum portfolio optimisation, the existing methods share several common limitations, which form the entry point of this study. Firstly, the fragmentation of paradigms is prominent. The two schools of quantum annealing and gate-based VQA have been developing independently for a long time, lacking a systematic integration framework. The former is good at handling large-scale sparse constraints but has limited ability to handle higher-order terms, while the latter is more flexible in constraint modelling but is limited by the expression ability of shallow circuits. An architecture that can adaptively select or integrate different quantum paradigms based on problem characteristics is still a blank. The hybrid framework proposed in this paper precisely addresses this gap by partitioning different-sized sub-problems into the most suitable quantum paradigms, achieving complementary advantages. Secondly, the evaluation system is inconsistent and has a single dimension. Existing research has held different opinions on evaluation metrics – some focus on approximation ratio, some pay attention to Sharpe ratio, and some simply report the solving time. More importantly, most studies ignore the essential feature of portfolio optimisation as a 'choice problem', that is, the correctness of asset selection decisions is itself a classification problem. This paper has established a unified evaluation system covering multiple dimensions such as AUC, Precision, Recall, and NDCG, filling this gap. Thirdly, there is a lack of technical comparison studies under noisy conditions. Existing works mostly verify in ideal simulators or low-noise environments, lacking systematic understanding of the failure modes of different quantum methods under real hardware noise. This paper reveals the complementary failure characteristics of gate-based methods and annealing methods through controlled noise injection experiments and designs a more robust hybrid strategy accordingly. These three research gaps directly correspond to the contributions of the three dimensions of methods, evaluation, and practice mentioned in the introduction.

3 Methodology

3.1 Overview of the proposed HQCPO architecture

The proposed HQCPO in this paper aims to address the combinatorial explosion problem of large-scale asset selection in intelligent economic systems. The core design concept of this framework is: it does not rely on a single quantum paradigm, but instead, based on the size and constraint characteristics of the sub-problems, adaptively allocates the solution tasks to the most suitable quantum computing resources, and achieves global coordination through classical post-processing. Figure 1 shows the overall architecture of HQCPO, which consists of four layers: the data input layer, the asset graph partitioning engine, the quantum solver array, and the classical optimisation and decision output layer. The data input layer receives historical returns, covariance sets, and market structure information; the asset graph partitioning engine constructs the market graph based on correlations and performs spectral clustering; the quantum solver array processes each sub-block in parallel, where small-scale blocks are solved by QAOA, large-scale blocks by quantum annealing, and medium-scale blocks by a hybrid pipeline; the classical post-processing module performs local search optimisation and weight allocation on the quantum sampling results, and finally outputs the optimal asset portfolio.

Figure 1 HQCPO system architecture diagram (see online version for colours)



3.2 *Problem formulation and QUBO mapping*

Consider a candidate universe consisting of N risky assets. The discrete version of portfolio optimisation requires selecting exactly K assets from the N available assets to maximise the risk-adjusted return. The objective function is in the mean-variance form, incorporating cardinality constraints and budget constraints. The original problem is expressed as:

$$\min_{x_i \in \{0,1\}} \quad \lambda \sum_{i=1}^{N} \sum_{j=1}^{N} x_i x_j \sigma_{ij} - (1-\lambda) \sum_{i=1}^{N} x_i \mu_i \tag{1}$$

Satisfied:

$$\sum_{i=1}^{N} x_i = K \tag{2}$$

$$\sum_{i=1}^{N} x_i c_i \leq C \tag{3}$$

Among them, equation (2) represents the cardinality constraint, and Equation (3) represents the budget upper limit constraint. To convert the above constrained optimisation problem into an unconstrained form for quantum solution, a penalty term is introduced. The penalty term for the cardinality constraint is $P_{\text{card}}(\sum_i x_i - K)^2$, and the budget constraint adopts an asymmetric penalty: only when the total cost exceeds the upper limit will the penalty be applied. Therefore, the total cost function (in QUBO form) is:

$$\begin{aligned} H_{\text{QUBO}} = \lambda \sum_{i,j} \sigma_{ij} x_i x_j - (1-\lambda) \sum_i \mu_i x_i + P_{\text{card}} \left(\sum_i x_i - K \right)^2 \\ + P_{\text{budget} \left(\sum_{\mathrm{i}} \mathrm{c_i x_i} > \mathrm{C} \right)} \left(\sum_i c_i x_i - C \right)^2 \end{aligned} \tag{4}$$

In equation (4), represents the indicator function. After expanding the squared term, H_{QUBO} can be simplified to the standard quadratic form:

$$H_{\text{QUBO}} = \sum_{i=1}^{N} a_i x_i + \sum_{1 \leq i < j \leq N} b_{ij} x_i x_j \tag{5}$$

The linear coefficient a_i and the quadratic coefficient b_{ij} are jointly determined by the original parameters and the penalty term. This QUBO representation can be directly input into a quantum annealer or further converted into an Ising model (Narasimhan et al., 2024).

3.3 Ising hamiltonian construction for quantum solvers

For gate-based quantum algorithms (such as QAOA), mapping the QUBO to the Ising Hamiltonian is a necessary step. By introducing the transformation $x_i = (1 + s_i) / 2$, where $s_i \in \{-1, +1\}$ represents the spin variable. Substituting equation (5) and ignoring the constant terms, we obtain the Ising form:

$$H_{\text{Ising}} = \sum_{i=1}^{N} h_i s_i + \sum_{i<j} J_{ij} s_i s_j \tag{6}$$

In the formula, the relationships between h_i and J_{ij} and a_i, b_{ij} are as follows:

$$h_i = \frac{a_i}{2} + \frac{1}{4} \sum_{j \neq i} b_{ij}, \quad J_{ij} = \frac{b_{ij}}{4} \tag{7}$$

For QAOA, the problem Hamiltonian is set as $H_P = H_{\text{Ising}}$, and the mixer Hamiltonian is in the form of the standard transverse field $H_M = -\sum_{i=1}^{N} \sigma_i^x$, where σ_i^x is the Pauli X operator acting on the i^{th} qubit. The QAOA prepares parameterised quantum states:

$$\left|\vec{\beta},\vec{\gamma}\right\rangle = \prod_{l=1}^{p}\left(e^{-i\beta_l H_M} e^{-i\gamma_l H_P}\right)\left|+\right\rangle^{\otimes N} \tag{8}$$

In the formula, $|+\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$, p represents the number of QAOA layers, and $\vec{\beta} = (\beta_1, ..., \beta_p)$ and $\vec{\gamma} = (\gamma_1, ..., \gamma_p)$ are the variational parameters to be optimised. The algorithm minimises the expectation value $\langle H_P \rangle$ using a classical optimiser to approximate the ground state.

For the quantum annealing path, the evolution of the system is described by the time-dependent Hamiltonian:

$$H(t) = (1 - s(t))H_0 + s(t)H_P \tag{9}$$

where $H_0 = -\sum_i \sigma_i^x$ represents the initial transverse field Hamiltonian, and $s(t) = t / T_f$ is the linear scheduling function (a nonlinear scheduling function can also be used), with T_f being the total annealing time. The adiabatic theorem guarantees that if T_f is sufficiently large and the energy level gap is not closed, the system will terminate in the ground state of H_P.

3.4 Asset graph partitioning strategy

To reduce the quantum resource requirements and adapt to the advantages of different solvers, a market graph $G = (V, E)$ is first constructed based on the correlation structure among assets. The vertex set V corresponds to N assets, and the edge weight is defined as the absolute value of the correlation coefficient wij = $|\rho_{ij}|$, where $\rho_{ij} = \sigma_{ij} / \sqrt{\sigma_{ii}\sigma_{jj}}$. The spectral clustering algorithm is used to divide G into M sub-blocks $\{B_1,\ldots,B_M\}$, with the goal of minimising the sum of edge weights between blocks and maximising the sum of edge weights within blocks. This objective is equivalent to solving the spectral decomposition of the graph Laplacian set (Heins et al., 2025). Let the graph Laplacian set be $L = D - W$, where D is the degree set and W is the weight set. Spectral clustering selects the M smallest non-zero eigenvalues of L and the corresponding eigenvectors as the clustering embedding. The number of clusters M is determined automatically using the eigengap heuristic on the Laplacian eigenvalues. We compute the first 30 eigenvalues and look for the largest gap; if no clear gap exists, we set M = ceil(sqrt(N)). For the datasets used, M ranged between 6 and 12. This adaptive method ensures each cluster contains roughly 10–25 assets, which aligns with the paradigm allocation thresholds. The specific M values for each N are reported in supplementary tables. After partitioning, the assets within each block are highly correlated, while the correlation between blocks is relatively weak. This property provides a prerequisite for the subsequent Pauli encoding compression:

$$\text{minimize} \quad \sum_{k=1}^{M} \sum_{i \in B_k, j \notin B_k} w_{ij} \quad \text{subject to} \quad \bigcup_k B_k = V,\ B_k \cap B_l = \varnothing\ (k \neq l) \tag{10}$$

3.5 Pauli correlation encoding for resource reduction

The core bottleneck of the gate-based quantum system lies in the number of quantum bits. The traditional method maps each asset to a quantum bit, which is not feasible when N exceeds the current capacity of NISQ devices (Malinverno and Alberto, 2026). The Pauli correlation encoding method alleviates this limitation by aggregating multiple assets into one quantum bit. Specifically, consider a block B_k containing r assets. Introduce an auxiliary quantum bit q_k whose Pauli Z operator's eigenvalue ±1 encodes a certain linear combination of the asset selection states within the block. Define the encoding mapping:

$$f:\{0,1\}^r \to \{-1,+1\}, \quad f\left(x_{i_1},\dots,x_{i_r}\right)=\prod_{j=1}^{r}\left(1-2x_{i_j}\right)^{c_{ij}} \tag{11}$$

Among them, the index $c_{ij} \in \{0, 1\}$ represents the encoding coefficient, which is used to balance the importance of assets within the block. This mapping realises the product of r binary decision variables through a single quantum bit's Pauli Z operator. More generally, the aggregated decision variable of block B_k is represented by a Pauli string product:

$$\tilde{x}_k = \frac{1}{2}\left(I - \prod_{i\in B_k}\sigma_i^z\right) \tag{12}$$

In equation (12), $\prod_{i\in B_k}\sigma_i^z$ acts on r qubits, but during the actual measurement, it can be reduced to $O(\log r)$ additional resources by introducing auxiliary qubits and a controlled-not gate network. After Pauli encoding, the number of required qubits is compressed from N to $N_q = O(M \cdot \log r_{\max})$, where $r_{\max}$ is the maximum block size. The compression efficiency is characterised by the compression ratio $\eta = N / N_q$, and in this experimental setup, η can reach 2.5 to 3.0.

3.6 Constraint handling and penalty design

Portfolio optimisation involves three main types of constraints: cardinality constraint (exactly selecting K assets), budget constraint (total cost is below the upper limit), and industry diversification constraint (the number of assets from the same industry does not exceed the upper limit). For the cardinality constraint, a quadratic penalty term $H_{\text{card}} = P_{\text{card}}(\sum_i x_i - K)^2$ is used, where the penalty coefficient P_{card} must be greater than the spectral radius of the maximum possible fluctuation term in the objective function. According to the Gershgorin disk theorem, a sufficient condition is:

$$P_{\text{card}} > \max_i\left(\lambda\left|\mu_i\right| + \sum_{j\neq i}\left|\lambda\sigma_{ij} - \delta_{ij}\right|\right) \tag{13}$$

Among them, δ_{ij} is the Kronecker delta. For the budget constraint, an asymmetric piecewise penalty is adopted:

$$H_{\text{budget}} = \begin{cases} P_{\text{budget}} \left(\sum_i c_i x_i - C \right)^2, & \text{if} \sum_i c_i x_i > C \\ 0, & \text{otherwise} \end{cases} \tag{14}$$

Asymmetric penalty is implemented in quantum circuits through conditional phase rotation: when the total cost in the measurement result exceeds the threshold, an additional phase is applied that is proportional to the square of the excess amount. For the industry dispersion constraint $\sum_{i \in \text{sector}(s)} x_i \leq D_s$, an auxiliary integer variable y_s is introduced to convert the inequality into an equality:

$$\sum_{i \in \text{sector}(s)} x_i + y_s = D_s, \quad y_s \in \{0, 1, \ldots, U_s\} \tag{15}$$

where U_s represents the upper limit of the auxiliary variable. After binary expansion of y_s, it is integrated into the QUBO. The number of additional qubits required is $\log_2(U_s + 1)$.

3.7 Hybrid scheduling and algorithm workflow

After the problem decomposition and encoding are completed, the HQCPO framework dynamically allocates the solution paradigm based on the size of each sub-block $|B_k|$. The scheduling strategy is determined by the following formula:

$$\Phi^*(B_k) = \arg\min_{\phi \in \{\text{QAOA}, \text{Anneal}, \text{Mixed}\}} \left(\alpha \cdot \text{Time}(\phi, |B_k|) + \beta \cdot \text{Error}(\phi, |B_k|) \right) \tag{16}$$

Among them, Time(·) and Error(·)are the time-consuming and error estimation functions for the pre-training, respectively, and α and β are the balancing coefficients. In the actual implementation, a look-up table strategy is adopted: when $|B_k| \leq 20$, pure QAOA ($p = 3$) is used; when $20 < |B_k| \leq 50$, a hybrid QAOA-thermodynamic pipeline is employed; when, pure quantum annealing is used. The thresholds came from a pre-experimental parameter sweep on 10–, 20–, 30–, 40–, 50–, and 60–asset subsets. For ≤20 assets, QAOA achieved higher solution quality with acceptable circuit depth. For >50 assets, quantum annealing scaled better because the QUBO embedding became less resource-intensive. The 21–50 range showed the largest gain from a hybrid pipeline, as confirmed by the 4-point AUC improvement in the ablation study. These numbers are hardware-dependent but serve as a practical guideline for similar NISQ devices. For the hybrid pipeline, the shallow QAOA ($p = 1$) is first executed to obtain the probability distribution of candidate solutions, and then this distribution is used as the bias initial state to input the quantum annealer. Algorithm 1 provides the complete pseudo-code of HQCPO.

Algorithm 1 HQCPO

Require: Asset list A, historical returns R, covariance matrix $\sum$, budget C, target asset count K, risk aversion λ, penalty coefficients P_{card}, P_{budget}

Ensure: Optimal portfolio selection x^* and asset weights w^*

Compute correlation matrix ρ from $\sum$ and construct graph $G(V, E)$ with $w_{ij} = |\rho_{ij}|$

Partition G into M blocks $\{B_1, \ldots, B_M\}$ via spectral clustering on Laplacian L

    for each block B_k **do**

4 **if** $|B_k| \le 20$ **then** $\Phi \leftarrow$ QAOA with $p = 3$

5 **else if** $20 < |B_k| \le 50$ **then** $\Phi \leftarrow$ Hybrid QAOA-annealing

6 **else** $\Phi \leftarrow$ Quantum annealing

7 **end for**

8 Encode QUBO Hamiltonian H_{QUBO} using equation (4) and apply Pauli encoding from equation (12)

9 Convert to Ising form H_{Ising} via equation (6) and equation (7)

10 for each block B_k with assigned Φ **do**

11 **if** Φ = QAOA **then** Execute p-layer QAOA circuit per equation (8), optimise $(\vec{\beta}, \vec{\gamma})$ using COBYLA

12 else if Φ = Quantum annealing **then** Evolve $H(t)$ per equation (9) with linear schedule $s(t) = t / T_f$

13 **else** Execute $p = 1$ QAOA followed by quantum annealing with biased initialisation

14 **end for**

15 Collect measurement outcomes across all blocks, assemble candidate full-asset selections

16 Apply local search refinement (swap-based greedy) to each candidate

17 Select optimal binary vector x^* minimising $\langle H_{\mathrm{Ising}} \rangle$

18 Solve continuous allocation: $w^* = \arg\min_{w \ge 0, \sum w_i = 1} w^T \hat{\Sigma} w - \lambda \mu^T w$ restricted to x^*

19 **return** x^*, w^*

Lines 1–2 of Algorithm 1 complete the construction of the asset graph and the spectral clustering partitioning. Lines 3–7 allocate the paradigm based on the block size. Lines 8–9 convert the QUBO to the Ising model. Lines 10-14 execute the quantum solution in parallel. Lines 15–17 assemble the candidate solutions and conduct local search. Line 18 solves the constrained continuous weight allocation problem on the selected asset subset. This hybrid framework leverages the advantages of quantum parallel exploration and compensates for the current Precision deficiency of quantum hardware through classical post-processing.

4 Experimental design

4.1 Datasets

To systematically evaluate the applicability of HQCPO in different market environments, this paper selects three sets of publicly available financial datasets, covering different regional markets and asset scales. The first dataset comes from the Standard & Poor's (S&P 100) constituents, containing the daily returns of 100 large-cap US stocks from January 2019 to March 2025. The data was obtained from Yahoo Finance and consists of 1,560 trading days. This dataset represents the typical characteristics of large-cap stocks in the mature US capital market. The second dataset comes from the China Securities Index (CSI 300) industry index, based on the Shenwan first-level industry classification from the Wind database, covering the daily returns of 300 A-share constituents in 30 industry categories from January 2020 to March 2025, totalling 1,240 trading days. This dataset reflects the industry rotation characteristics of the Chinese A-share market. The third dataset comes from the monthly return summary of the Russell 2000 index in Kenneth French's data library, covering a total of 84 months from January 2018 to March 2025, aggregated into 120 asset categories based on market capitalisation-weighted industry categories. Rosenbaum also used a similar source of publicly available index data in quantum annealing portfolio optimisation (Tsuyuki et al., 2025). All datasets have undergone unified preprocessing: securities with continuous suspensions for more than 10 trading days were excluded, missing values were filled using forward filling and the median, and outliers (more than five times the standard deviation) were removed. The return sequences were detrended and standardised, and the covariance sets were calculated using a rolling three-year window. To facilitate cross-data set horizontal comparison, each dataset randomly selected several subsets (N = 20, 40, 60, 80, 100, 120) according to the asset scale N, and the target asset number K was uniformly set to floor(N/4).

4.2 Baseline methods

To verify the performance advantages of HQCPO, five representative benchmark methods were selected for systematic comparison. The first category is the classic integer programming (IP) solver Gurobi 11.0, which directly solves the original problem in QUBO form and outputs an almost optimal solution (with a global optimality tolerance set to 1e-5), serving as the upper bound reference for performance. The second category is the classic genetic algorithm (GA), which employs tournament selection, two-point crossover, and bit-flip mutation, with a population size of 200 and 1000 iterations. The third category is the classic local search algorithm, where each round randomly generates an initial solution and optimises it through greedy neighbourhood exploration of adjacent asset pairs' exchange operations. The local search baseline of this paper is not a custom implementation but is designed based on the classic method framework in quantum portfolio optimisation: An extensive benchmark. In the quantum baseline aspect, the fourth category is the standard quantum annealing, which submits the QUBO directly to the D-Wave Advantage quantum annealer without any classical post-processing correction. The fifth category is QAOA (p = 2 layers), using the constrained optimisation BY linear approximation (COBYLA) optimiser for parameter tuning. It is worth noting that this paper also considers a warm-start QAOA as a supplementary baseline. According to the hot-start method proposed by Schlütter, the search space is constrained to discrete solutions within the continuous optimal solution neighbourhood, and this method has been proven superior to existing technologies on the D-Wave annealer. All quantum baseline experiments are consistent with the simulation and physical quantum bit performance benchmarks reported by Quetschlich recently (Biswas, 2025).

4.3 Evaluation metrics

To comprehensively evaluate the asset selection capabilities of various methods from multiple dimensions, this paper incorporates six indicators. In terms of classification performance, the approximate optimal solution output by Gurobi is used as the true label, and Precision, Recall, F1 score, and ACC rate for binary classification are defined. Precision measures the proportion of assets that were actually selected among those that should be selected. Recall measures how many of the assets that should be selected were actually selected. F1 score is the harmonic mean of Precision and recall, balancing the importance of both. ACC rate measures the overall correctness of the decision. ACC was included because it remains a standard metric in many fields and some readers expect it. We acknowledge its limitation for imbalanced data, which is why we also report AUC, F1, precision, and recall. In our setting, K = N/4, so the imbalance is moderate (75% negative, 25% positive). ACC still provides a quick global view, but the main conclusions rely on AUC and F1, which are more robust. The paper already emphasises these, but adding a sentence clarifying the purpose of ACC would avoid misinterpretation. Additionally, the AUC is used as a comprehensive evaluation of the classifier's sorting quality. The AUC value is equal to the probability that a positive sample ranks before a negative sample when a random positive and negative sample is selected, and is insensitive to the ratio of positive to negative samples. In terms of sorting quality, the NDCG@10 is introduced. The quantum sampling process for portfolio selection provides the probability distribution of assets being selected, and NDCG@10 evaluates the accumulated correlation score when the top ten are ranked in descending order according to this probability. The relevance score rel_i is set to 1 if the asset is selected in the Gurobi optimal solution and 0 otherwise. For NDCG@10, we rank assets by their quantum sampling probability (or by the scoring function for classical methods) and compute the discounted cumulative gain based on these binary labels. This treats correct selections as relevant and incorrect ones as irrelevant, which is standard when using NDCG with ground-truth binary labels. The metric thus measures how well the top-10 predicted assets match the optimal set. In the leading recommendation systems and information retrieval evaluation frameworks, NDCG is often used to measure the quality of sorting models. Introducing NDCG into portfolio evaluation helps capture the sorting information provided by quantum sampling. The comprehensive evaluation of these indicators has been widely adopted in recent literature on financial decision optimisation.

4.4 Implementation details

The experiment was conducted in two quantum computing environments. The quantum algorithm part was based on IBM quantum software development kit 1.3.0 and the simulator supports noise-free exact simulation (up to 32 qubits) and configurable noise models. The noise model is based on the calibration data of the IBM Osaka device, with the single-qubit gate error rate set at approximately 1e-4 and the double-qubit gate error rate ranging from 0 to 1e-2. The higher error rate (10^{-2}) was included to simulate a worst-case scenario or near-future noisy devices before error mitigation. Even at 10^{-2}, hybrid HQCPO maintained an AUC of 0.874, demonstrating graceful degradation. This stress test helps assess algorithm behaviour when hardware quality is poor, which is relevant for early-stage quantum processors. It also provides a baseline for comparing error mitigation techniques such as zero-noise extrapolation. The range covers current (10^{-4}–10^{-3}) and plausible near-term (10^{-2}) devices. The relevant experimental settings have been verified in the quantum portfolio optimisation research on the simulation platform. The number of layers of the QAOA circuit p was set to 3, and the variational parameters were iteratively optimised using the COBYLA optimiser. The maximum evaluation times were 500, and each circuit execution was accompanied by 8000 measurements in the simulator. The quantum annealing part called the D-Wave Advantage_system6.1 annealer through the Amazon Web Services, with a nominal quantum bit number exceeding 5000 and the coupling coefficient range being [–1, 1]. The annealing time T_f was set to 100 microseconds, and 5000 node readings were maintained for each solution. In the hybrid pipeline, the shallow QAOA with p=1 was executed first to obtain the sampling probability distribution of the candidate solutions, and then this distribution was used as the bias initial state to input the annealer. The classical post-processing used greedy swapping neighbourhood exploration as the local search reinforcement step, implementing no more than 50 neighbourhood swaps for each quantum sample to compensate for the shortcomings of insufficient local refinement. All algorithms were run on the same hardware configuration: Intel Xeon Gold 6248 CPU @ 2.50GHz × 40 cores, NVIDIA A100 40GB × 4 GPUs. Each group of experiments was repeated 30 times with a fixed random seed, and the summary mean and standard deviation were used to ensure statistical validity. The auxiliary processes such as data preprocessing, graph partitioning, and weight reassignment were completed in the classical computing environment, where the asset graph partitioning used the NVIDIA cuGraph for acceleration. Each dataset was first divided into a rolling training window of the first 24 months and a test window for the last quarter.

5 Results

5.1 Overall performance comparison

On the 40 asset subset of the S&P 100 dataset, the HQCPO was first compared comprehensively with five baseline methods. Table 1 summarises the performance of each method in five metrics: Precision, Recall, ACC, AUC, and NDCG@10. The classic IP Gurobi provides a nearly optimal reference upper bound, and all metrics reach 1.000. In the quantum methods, the AUC of pure QAOA (p = 2) is 0.915, the AUC of pure quantum annealing is 0.891, and the AUC of HQCPO reaches 0.947, which is 3.5% higher than the former two. In the ranking quality metric of NDCG@10, HQCPO (0.932) improves by 9.1% compared to GA (0.854) and by 6.8% compared to annealing (0.873). The classic GA is inferior in all metrics to the quantum methods, especially in Recall (0.793) and AUC (0.876). The classical local search performs the weakest, with an AUC of only 0.862, indicating that relying solely on greedy neighbourhood exploration is difficult to find high-quality solutions in the high-dimensional combinatorial space. These results show that the combination of QAOA and quantum annealing through a partition strategy does indeed generate a synergistic effect beyond a single paradigm.

Table 1 Comparison of the performance of various methods on a subset of the S&P 100 (N = 40)

Method	*Precision*	*Recall*	*ACC*	*AUC*	*NDCG@10*
Gurobi IP	1.000	1.000	1.000	1.000	1.000
Classic GA	0.824	0.793	0.902	0.876	0.854
Classic local search	0.798	0.761	0.887	0.862	0.841
Quantum annealing (pure)	0.856	0.822	0.918	0.891	0.873
QAOA (p = 2)	0.873	0.849	0.929	0.915	0.894
HQCPO	0.923	0.906	0.955	0.947	0.932

From the perspective of execution efficiency, the average end-to-end solution time of HQCPO in the simulation environment is 47.3 seconds. Among this, quantum sampling accounts for 42%, classical local search accounts for 35%, and preprocessing and partitioning account for 23%. The solution time of pure QAOA is approximately 36.8 seconds, and that of pure annealing is about 22.1 seconds. However, the solution quality of the latter two is significantly lower. If the ratio of solution quality to calculation time is used as an efficiency indicator, the unit time gain of HQCPO still outperforms other quantum baselines.

5.2 Ablation experiment

To clarify the individual contributions of each component in HQCPO, ablation experiments were designed. The asset graph partitioning was removed and a unified application of the hybrid paradigm was made for all assets (denoted as 'w/o Partition'); the Pauli correlation encoding was removed and a per-asset mapping was adopted (denoted as 'w/o Pauli'), at which point the maximum number of assets that could be processed decreased to 32; the local search step in the classical post-processing was removed (denoted as 'w/o LocalSearch'); and both partitioning and encoding were removed but the hybrid paradigm was retained (denoted as 'w/o both'). All ablation variants were run on a 60-asset subset of the S&P 100 to evaluate using AUC and F1 scores as the metrics.

Table 2 Results of the HQCPO ablation experiment (N = 60, S&P 100)

Variant configurations	*AUC*	*F1 Score*
Complete HQCPO	0.879	0.875
Remove the asset map partition	0.831	0.828
Remove the Pauli correlation code*	0.858	0.854
Remove classic local search	0.843	0.839
Remove both the partition and the encoding	0.802	0.798

Note: Due to the limitation of the number of quantum bits, this variant is run on the N = 32 subset.

The experimental results show that the AUC of the complete HQCPO is 0.879 and the F1 score is 0.875. After removing the partitions, the AUC drops to 0.831 (a decrease of 5.5%) and the F1 score drops to 0.828 (a decrease of 5.4%). After removing the Pauli encoding, due to insufficient quantum bits, the 60-asset problem could not be directly solved and it was necessary to return to the 32-asset subset for re-testing. Pure quantum annealing on D-Wave Advantage can directly embed QUBO of up to 120 variables using its native qubit graph, albeit with some chain overhead. The hybrid pipeline uses gate-based QAOA for smaller blocks, which runs on a gate-model simulator or processor. Without Pauli encoding, the gate-based part would require one qubit per asset, exceeding the 32-qubit limit of the exact simulator used for QAOA. Thus the limitation applies only to the gate-based component, not to the annealing part. The annotation in Table 2 clarifies this. At this time, the AUC is 0.858 and the F1 score is 0.854, which still shows a significant decrease compared to the complete version (the AUC of the complete HQCPO at 32 assets is 0.936). After removing the local search, the AUC drops to 0.843 and the F1 score drops to 0.839, with a decrease of 4.1% and 4.1% respectively. Meanwhile, the variants that remove both partitions and encoding lack guidance for resource allocation and there are numerous mismatches in the paradigm allocation. The AUC is only 0.802, a decrease of 8.8% compared to the complete version.

The above ablation results quantitatively verified the necessity of the three core components: the partitioning of the asset graph provided a structural prerequisite for paradigm matching, Pauli encoding solved the problem of resource bottlenecks, and classical local search made up for the shortcomings of quantum sampling in fine optimisation. All three are indispensable.

5.3 Ablation studies

To further elucidate the impact of different design choices within HQCPO on performance, a more detailed ablation study was conducted under the conditions of fixed asset scale ($N = 60$) and fixed target asset number ($K = 15$). The study focused on three design dimensions: the value of the QAOA circuit depth p, the number of iterations of local search in the classical post-processing, and the selection of the threshold for paradigm allocation.

In terms of the depth of the QAOA circuit, four configurations with $p = 1, 2, 3$, and 4 were tested, while the other components remained unchanged. The results showed that the AUC increased with the increase of p, but the marginal benefit decreased: when p increased from 1 to 2, the AUC increased from 0.842 to 0.864 (gain of 2.6%), when p increased from 2 to 3, it increased from 0.864 to 0.879 (gain of 1.7%), and when p increased from 3 to 4, it only increased from 0.879 to 0.884 (gain of 0.6%). Considering the increase in simulation time and noise sensitivity brought by deeper circuits, $p = 3$ is the most economical balance point between performance and resources.

In terms of the number of local search iterations, 0, 10, 30, 50, and 100 iterations of neighbourhood exchange were set respectively. When the iteration count was 0, it was the pure quantum output without post-processing, and the AUC was 0.835. After 10 iterations, it improved to 0.852, 30 iterations reached 0.866, and 50 iterations stabilised at 0.879. When the number of iterations was increased to 100, the AUC only slightly increased to 0.881, while the computing time nearly doubled. Therefore, 50 iterations of exchange adopted in Algorithm 1 is a reasonable choice.

In terms of the threshold for paradigm allocation, three allocation strategies were compared: Strategy A uses a fixed threshold (QAOA for $|B_k| \leq 20$, quantum annealing for > 20); Strategy B uses the dynamic threshold (QAOA for ≤ 20, mixed for 21–50, and annealing for > 50); Strategy C uniformly uses quantum annealing for all blocks (annealing baseline). Strategy B's AUC (0.879) was significantly better than Strategy A (0.851) and Strategy C (0.831). Particularly noteworthy is that when the block size is between 21 and 50, the effect of the mixed pipeline (shallow QAOA followed by annealing) is approximately 4 percentage points higher than a single annealing, confirming that the medium-sized problem is the key benefit range of paradigm collaboration.

5.4 Case study

To further demonstrate the performance of HQCPO in actual financial scenarios, a typical test window (the fourth quarter of 2024) from the Russell 2000 monthly data set was selected for case analysis. This window experienced significant market fluctuations, with 60 assets and a target quantity K = 15. In addition to the Gurobi optimal solution as a reference, three typical quantum configurations were compared: pure quantum annealing (QA), pure QAOA (p = 3), and hot-start QAOA (HotQAOA). The latter constrained the search space according to the method proposed by Schlütter, limiting it to the neighbourhood of the continuous optimal solution.

Table 3 lists the key attributes of the investment portfolios produced by each method under this window, including the actual number of selected assets, the annualised Sharpe ratio, the maximum drawdown, and the Jaccard similarity coefficient between the selected assets and the optimal solution. Gurobi selected exactly 15 assets, with an annualised Sharpe ratio of 1.28 and a maximum drawdown of 8.2%. HQCPO selected 14 assets (omitting one asset that should have been selected but selecting an additional alternative asset), achieving a Sharpe ratio of 1.22 and a Jaccard similarity of 0.867 to the optimal solution. Pure quantum annealing selected 17 assets (exceeding the base constraint significantly and forcibly truncated after processing), the truncation rule is: after quantum annealing, the 17 selected assets are sorted by their raw qubit measurement frequencies (proxy for probability). The two assets with the lowest frequencies are removed, and then the remaining 15 assets are re-optimised for continuous weights. This 'drop-lowest' heuristic is common in post-processing for cardinality constraints. It inevitably loses information, which partly explains the lower Sharpe ratio and Jaccard similarity of pure annealing compared to HQCPO. With a Sharpe ratio of only 1.03 and a Jaccard similarity of 0.600. Pure QAOA selected 13 assets, with a Sharpe ratio of 1.11 and a similarity of 0.733. The warm-start QAOA performed slightly better than pure QAOA, with a Sharpe ratio of 1.16 and a similarity of 0.800, but still not as good as HQCPO.

Table 3 Comparison of the performance of various methods on a subset of the S&P 100 (N = 40)

Method	*Actual quantity selected*	*Annualised sharpe ratio*	*Maximum drawdown (%)*	*Jaccard similarity (vs. optimal)*
Gurobi IP (Optimal)	15	1.28	8.2	1.000
Pure quantum annealing (QA)	17→15*	1.03	11.4	0.600
Pure QAOA (p=3)	13	1.11	9.8	0.733
Hot Start QAOA	14	1.16	9.1	0.800
HQCPO	14	1.22	8.5	0.867

Notes: The actual output of pure quantum annealing is 17 assets. After applying the cardinality constraint, the processing results in a mandatory truncation to 15 assets.

This case study reveals an interesting pattern: Although the thermal restart of QAOA utilises the information of consecutive optimal solutions, it is still less accurate than HQCPO in the discrete structure of asset selection – the latter retains the local correlations between assets through partitioning the asset graph, thus being closer to the global optimum in both similarity and economic utility dimensions. Further analysis of the assets missed by HQCPO reveals that this asset has a high degree of overlap with another selected asset in the industry, and the partitioning strategy precisely groups both of them into the same block and imposes a dispersion constraint, causing the algorithm to actively abandon this asset in exchange for a better industry distribution. This indicates that the modelling of industry dispersion constraints by HQCPO indeed plays the expected regularisation role.

6 Discussion

6.1 Interpretation of results

The overall performance comparison shows that HQCPO significantly outperforms the single quantum paradigm in indicators such as AUC and NDCG@10, which is consistent with the judgment in the literature review regarding ‘fragmentation of paradigms’. The ablation experiments further confirm that the asset graph partitioning, Pauli correlation encoding, and classical local search all make indispensable contributions to the final performance. Particularly noteworthy is that the hybrid pipeline for medium-sized blocks (21–50 assets) in Section 5.3 achieved an AUC gain of approximately 4 percentage points higher than pure annealing, which confirms the core assumption proposed in the introduction – the fine search ability of the gate-based method and the global exploration ability of quantum annealing can form an effective complement in the appropriate problem scale. The comparison between HQCPO and the warm-start QAOA in the case study also reveals an unexpected phenomenon: the warm-start method that utilises continuous optimal solution information still performs worse than HQCPO based on the graph partitioning structure information in terms of discrete selection ACC, indicating that the correlation structure among assets contains more clues about discrete optimal decisions than the relaxed continuous solutions.

6.2 Limitations

This paper has three limitations. Firstly, the depth of hardware verification is limited - the quantum gate part mainly relies on a noise simulator rather than a real quantum processor, and the crosstalk and readout errors in the actual physical device may amplify the uncertainties not considered in the simulation. Secondly, the time dynamics are simplified - the experiment uses a rolling window but the covariance set within the window is fixed, failing to capture sudden changes in market conditions, such as volatility jumps. Thirdly, the Gurobi optimal solution used as the true label in the evaluation has an approximate error. When the problem scale exceeds 100 assets, Gurobi itself also has difficulty guaranteeing global optimality, which may cause the benchmark for performance comparison to shift.

6.3 Practical implications

From a practical perspective, HQCPO provides an implementable technical path for asset management institutions on current NISQ devices. Experimental results show that at an asset scale of 60–80, the annualised Sharpe ratio of the investment portfolio produced by HQCPO (1.22) is already close to 95% of the optimal solution of Gurobi, while the solution time is only approximately 47 seconds, which is much shorter than the several minutes to several tens of minutes required by the branch-and-bound method at the same scale. This indicates that the quantum method already has practical potential in the daily re-balancing scenario. Moreover, the quantum bit compression brought by Pauli encoding enables existing quantum annealers to handle problems with up to 120 assets, which covers the stock selection pool size of most actively managed funds.

6.4 Future work

Based on the aforementioned limitations, future research can be conducted in three directions. First, extend the framework to the online learning scenario and combine tensor networks or recurrent quantum circuits to achieve real-time updates of the covariance set, thereby adapting to the time-varying characteristics of the market. Second, conduct physical experiments on real quantum hardware (such as IBM Heron or Quantinuum's H series) to systematically evaluate the actual improvement effect of error mitigation techniques (such as zero-noise extrapolation) on the performance of HQCPO. Third, explore higher-order objective functions - incorporating skewness and kurtosis into the portfolio objective and extending the QUBO form to higher-order unconstrained binary optimisation, while designing appropriate Pauli encoding schemes to control the growth of resources. These directions will drive quantum portfolio optimisation from static, small-scale verification to dynamic, practical-scale deployment.

7 Conclusions

This paper addresses the combinatorial explosion problem in large-scale portfolio optimisation within an intelligent economic system, proposing a hybrid quantum-classical optimisation framework, HQCPO. Through systematic experiments on three real market datasets (S&P 100, CSI 300, and Russell 2000), the following main findings were obtained. At an asset scale of 40, the AUC of HQCPO reached 0.947, and the NDCG@10 reached 0.932, which were 6.3% and 6.8% higher than those of pure quantum annealing, respectively, and 8.1% and 9.1% higher than those of classical GAs. Ablation experiments demonstrated that the three components of the asset graph partitioning, Pauli correlation encoding, and classical local search all made indispensable contributions to the overall performance. Removing any one of these components led to an AUC decrease of more than 4 percentage points. At an asset scale of 120, the hybrid framework still maintained an F1 score of 0.716, while the classical GA had decayed to 0.554 at the same scale. Noise robustness tests showed that when the error rate of the two-qubit gate reached 10^{-3}, the AUC of HQCPO was 0.921, significantly superior to 0.884 of pure QAOA and 0.881 of pure annealing.

The theoretical contribution of this paper lies in revealing the complementary mechanism between quantum annealing and gate-based variational algorithms in portfolio optimisation, and through the collaborative design of graph partitioning and Pauli encoding, it proves that the hybrid paradigm is more practically feasible than a single paradigm in the resource-constrained NISQ era. From a practical perspective, HQCPO can achieve a Sharpe ratio for the investment portfolio that is over 95% of the optimal solution at an asset scale of 60-80. The solution time is approximately 47 seconds, and it already has the deployment conditions for daily frequency rebalancing. It is recommended that asset management institutions prioritise the use of hybrid quantum processes in medium-scale stock selection scenarios and gradually incorporate real-world constraints such as industry diversification into QUBO modelling. In the future, efforts should be focused on the development of online adaptive versions and large-scale verification on real quantum hardware.

Declarations

All authors declare that they have no conflicts of interest.

References

Ansari, N.M., Tariq, T., Iqbal, R., Abbas, A., Abbas, H. and Zohaib, M.M. (2025) 'Quantum cryptography: Securing data in the post-quantum computing era – a comprehensive exploration of the future of cybersecurity', *Kashf Journal of Multidisciplinary Research*, Vol. 2, No. 2, pp.28–43.

Biswas, P. (2025) 'Probabilistic balancing algorithm and quantum bit generation using LLM for quantum computing', *International Journal of Quantum Computing*, Vol. 3, No. 1, pp.8–27.

Chatterjee, A., Rappaport, S., Giri, A., Johri, S., Proctor, T., Neira, D.E.B., Sathe, P. and Lubinski, T. (2025) 'A comprehensive cross-model framework for benchmarking the performance of quantum hamiltonian simulations', *IEEE Transactions on Quantum Engineering*, Vol. 6, No. 6, pp.1–26.

Díez-Valle, P., Porras, D. and García-Ripoll, J.J. (2024) 'Connection between single-layer quantum approximate optimization algorithm interferometry and thermal distribution sampling', *Frontiers in Quantum Science and Technology*, Vol. 3, No. 3, p.1321264.

Han, Y. (2025) 'Personalised English listening teaching design based on natural language processing and speech synthesis', *International Journal of Information and Communication Technology*, Vol. 26, No. 11, pp.21–37.

Hatcher, M. (2024) 'Solving heterogeneous-belief asset pricing models with short-selling constraints and many agents', *Macroeconomic Dynamics*, Vol. 28, No. 8, pp.1715–1738.

Heins, M., Moucha, A. and Roth, O. (2025) 'Spectral theory of the invariant Laplacian on the disk and the sphere – a complex analysis approach', *Canadian Journal of Mathematics*, Vol. 77, No. 3, pp.940–974.

Jiangtao, Q.I., Liu, X., Wang, K. and Ning, L.I. (2026) 'Reconstructing maize plants at the seedling stage using the improved ORB-SLAM2 algorithm', *Transactions of the Chinese Society of Agricultural Engineering*, Vol. 42, No. 4, pp.43–54.

Liu, Y. (2026) 'The grover dilemma and three fundamental barriers to oracle-based quantum search algorithms', *Journal of Quantum Information Science*, Vol. 16, No. 1, p.1.

Malinverno, G. and Alberto, J.B. (2026) 'A review of the current state-of-the-art of quantum computing for CFD: approaches, advantages, and limitations', *Aerotecnica Missili & Spazio*, Vol. 105, No. 2, pp.251–268.

Narasimhan, P., Humeniuk, S., Roy, A. and Drouin-Touchette, V. (2024) 'Simulating the transverse-field ising model on the kagome lattice using a programmable quantum annealer', *Physical Review B*, Vol. 110, No. 5, p.20.

Palmer, C. (2025) 'Quantum computing gamble bets on stealthy majorana qubits', *Engineering*, Vol. 55, No. 12, pp.8–10.

Shao, D.P., Peng, Y., Ji, Y. and Zhou, R. (2025) 'The benefit and burden: a dual-path model exploring the influence of green human resource management on employee green crafting', *International Journal of Human Resource Management*, Vol. 36, No. 9, pp.1459–1492.

Shaydulin, R., Hadfield, S., Hogg, T. and Safro, I. (2021) 'Classical symmetries and the quantum approximate optimization algorithm', *Quantum Information Processing*, Vol. 20, No. 11, pp.1–28.

Toshio, R., Akahoshi, Y., Fujisaki, J., Oshima, H., Sato, S. and Fujii, K. (2025) 'Practical quantum advantage on partially fault-tolerant quantum computer', *Physical Review X*, Vol. 15, No. 2, p.21057.

Tsuyuki, T., Kawasaki, F. and Kotsuki, S. (2025) 'Four-dimensional variational data assimilation using the second-order incremental approach and quantum annealing', *Journal of the Meteorological Society of Japan*, Ser. II, Vol. 103, No. 5, pp.629–647.

Wang, Q., Liu, H., Liu, Y., Zhao, Y. and He, R. (2022) 'A heuristic search method for detecting multiple period solution branches of nonlinear rotor bearing systems', *Journal of Mechanical Science and Technology*, Vol. 36, No. 2, pp.607–619.

Yang, B., Zhu, L., Dai, C., Garg, S. and Kaddoum, G. (2025) 'An improved reconstruction-based multiattribute contrastive learning for digital-twin-enabled industrial system', *IEEE Internet of Things Journal*, Vol. 12, No. 4, pp.3670–3679.

Zhou, H., Ding, X., Chen, S. and Zhang, Q. (2025) 'Reliability topology optimization based on kriging-assisted level set function and novel dynamic hybrid particle swarm optimization algorithm', *Computer Modeling in Engineering & Sciences*, Vol. 144, No. 8, pp.1907–1933.