DOI: 10.15514/ISPRAS-2026-38(4)-29

Discrete Markowitz Portfolio Optimization with Open-Source Classical and Quantum-Inspired Solvers: A Cross-Market Walk-Forward Study

S.M. Avdoshin, ORCID: 0000-0001-8473-8077 <savdoshin@hse.ru>
K.A. Patrushev, ORCID: 0009-0008-0915-5443 <kapatrushev@edu.hse.ru>
National Research University Higher School of Economics,
20, Myasnitskaya st., Moscow, 101000, Russia.

Abstract. The cardinality-constrained Markowitz problem is NP-hard and traditionally solved with commercial MIQP solvers. Following the 2022 export restrictions that rendered both commercial MIQP software and cloud quantum platforms (IBM Quantum, D-Wave Leap) inaccessible from the Russian Federation, practitioners require open-source alternatives. This paper systematically compares three solver families for the discrete mean-variance problem: two open-source classical MIQP solvers (SCIP and ECOS_BB via CVXPY), and a quantum-inspired simulated annealing solver (D-Wave neal) operating on a binary-inclusion QUBO with a two-stage hybrid pipeline. All solvers share a unified problem instance. In Experiment A (synthetic scalability), the quantum-inspired SA becomes the fastest solver at $N \geq 150$ (7× faster than SCIP at $N = 200$) but incurs an 11–13% optimality gap. In Experiment D (walk-forward backtest on S&P 500 and MOEX with realistic transaction costs), discrete optimization delivers +39–42 bps/year over the 1/N benchmark on S&P 500, but neal SA underperforms SCIP by 159 bps/year due to residual optimality gap and elevated turnover. On the non-stationary Russian market, all MVO strategies underperform 1/N by 100–229 bps/year, reproducing the DeMiguel-Garlappi-Uppal paradox. The study quantitatively characterizes the scalability–quality trade-off, decomposes the optimality gap into formulation, sampler, and penalty-calibration components, and identifies conditions under which the current neal-based binary-inclusion pipeline is insufficient for practical deployment.

Keywords: portfolio optimization; Markowitz; QUBO; simulated annealing; quantum-inspired algorithms; mixed-integer quadratic programming; walk-forward backtest; cardinality constraint; QAOA.

For citation: Avdoshin S.M., Patrushev K.A. Discrete Markowitz Portfolio Optimization with Open-Source Classical and Quantum-Inspired Solvers: A Cross-Market Walk-Forward Study. Trudy ISP RAN/Proc. ISP RAS, 2026, vol. 38, issue 4, part 2, pp. 245-256 DOI: 10.15514/ISPRAS-2026-38(4)-29.

Acknowledgements. This research was supported in part through computational resources of HPC facilities at HSE University.

Дискретная оптимизация портфеля Марковица с использованием открытых классических и квантово-вдохновлённых алгоритмов: кросс-рыночное исследование с пошаговой валидацией

С.М. Авдошин, ORCID: 0000-0001-8473-8077 <savdoshin@hse.ru>
К.А. Патрушев, ORCID: 0009-0008-0915-5443 <kapatrushev@edu.hse.ru>
Национальный исследовательский университет «Высшая школа экономики»,
Россия, 101000, Москва, Мясницкая улица, д. 20.

Аннотация. Задача Марковица с ограничением на кардинальность является NP-трудной и традиционно решается коммерческими MIQP-решателями (Mixed-Integer Quadratic Programming). После введения в 2022 году экспортных ограничений, сделавших недоступными с территории РФ как коммерческое программное обеспечение (ПО) MIQP, так и облачные квантовые платформы (IBM Quantum, D-Wave Leap), практикам необходимы открытые альтернативы. В работе проведено систематическое сравнение трёх семейств решателей: двух открытых классических MIQP (решатели SCIP и ECOS_BB с библиотекой CVXPY) и квантово-вдохновлённого решателя имитационного отжига на бинарной QUBO-формулировке включения активов с двухэтапным гибридным конвейером. Все решатели используют единую инстанцию задачи. В эксперименте по синтетической масштабируемости метод квантово-вдохновлённой имитации отжига становится самым быстрым (в 7 раз быстрее SCIP при некоторых условиях), но с зазором оптимальности 11–13%. В другом эксперименте (пошаговое тестирование на биржевых данных индексов S&P 500 и MOEX с реалистичными транзакционными издержками) дискретная оптимизация показывает рост относительно простой стратегии равной доли активов по индексу S&P 500, однако подход neal SA уступает алгоритму SCIP из-за остаточного зазора и повышенного оборота. На нестационарном российском рынке все стратегии оптимизации по среднему и дисперсии уступают стратегии равной доли активов, воспроизводя парадокс DeMiguel–Garlappi–Uppal. Исследование количественно характеризует компромисс между масштабируемостью и качеством, раскладывает зазор оптимальности на составляющие (формулировка, сэмплер, калибровка штрафа) и определяет условия, при которых текущий конвейер на основе стратегии neal недостаточен для практического применения.

Ключевые слова: портфельная оптимизация; Марковиц; задача оптимизации QUBO; имитационный отжиг; квантово-вдохновлённые алгоритмы; смешанное целочисленное квадратичное программирование; пошаговое тестирование; ограничение на кардинальность; квантовый метод QAOA.

Для цитирования: Авдошин С.М., Патрушев К.А. Дискретная оптимизация портфеля Марковица с использованием открытых классических и квантово-вдохновлённых алгоритмов: кросс-рыночное исследование с пошаговой валидацией. Труды ИСП РАН, 2026, том 38, вып. 4, часть 2, стр. 245–256 (на английском языке). DOI: 10.15514/ISPRAS-2026-38(4)-29.

Благодарности. Исследование выполнено с использованием суперкомпьютерного комплекса НИУ ВШЭ.

1. Introduction

Modern portfolio theory, founded by Markowitz in 1952 [1], allocates capital across N risky assets to maximize expected return for a given level of risk. In practice, institutional portfolios face a cardinality constraint limiting the number of held assets to at most $K < N$, driven by custody, compliance, and rebalancing costs. This integer constraint transforms the continuous quadratic program into a mixed-integer quadratic program (MIQP), which is NP-hard [2], making the problem a natural candidate for quantum and quantum-inspired heuristics [3-4].

For over a decade, the state of the art for portfolio MIQP has been dominated by commercial solvers – Gurobi, CPLEX, MOSEK – while quantum hardware from IBM, Google, and D-Wave has been applied to small-scale portfolio problems [5-10]. Following 2022 export restrictions, practitioners in the Russian Federation lost access to both commercial MIQP solvers and cloud quantum platforms (IBM Quantum, D-Wave Leap). The domestic alternative – the Bauman Octillion superconducting processor (SnowDrop 4Q/8Q), with public access since December 2025 – was considered but its registration is deferred to future work. Russian asset managers and researchers therefore require an empirical assessment of whether open-source MIQP solvers and quantum-inspired heuristics can fill this gap.

This paper addresses this need through a two-pronged study. First, we benchmark wall-time scalability and solution quality of SCIP [11], ECOS_BB (both via CVXPY [12]), and D-Wave neal on synthetic instances. Second, we evaluate business-level impact through a rolling walk-forward backtest on historical S&P 500 and MOEX data with realistic transaction costs. Our central finding: the quantum-inspired SA achieves a 5–10× wall-time speedup at $N \geq 150$ but at the cost of a 100–200 bps/year performance gap in walk-forward on structured financial data. We quantitatively decompose this trade-off and identify directions to close it.

The remainder is organized as follows. Section 2 discusses motivation; Section 3 reviews related work; Section 4 states the problem; Section 5 describes general design; Section 6 details implementation; Section 7 presents experimental results; Section 8 concludes.

2. Motivation

The motivation is twofold. From a **practical standpoint**, quantitative portfolio managers in the Russian Federation face a unique constraint landscape: Gurobi and CPLEX are no longer licensable, and IBM Quantum and D-Wave Leap do not accept Russian registrations. The open-source MIQP solver SCIP and the quantum-inspired library D-Wave neal are fully functional within Russian infrastructure, but their relative performance for financial optimization has not been systematically benchmarked. From a **theoretical perspective**, the boundary of practical applicability for quantum-inspired methods in finance remains unclear. Prior work [4–10] has focused on small-to-medium in-sample demonstrations without walk-forward validation, and studies comparing quantum-inspired solvers to open-source (rather than commercial) baselines are rare. This study determines the conditions under which quantum-inspired SA becomes competitive with open-source MIQP, and whether any scalability advantage translates into measurable business value on real market data.

3. Related Work

Classical portfolio theory. The mean-variance framework [1] remains the foundational model for quantitative allocation. A central difficulty is estimating the covariance matrix Σ: for large N, the sample covariance is singular or severely ill-conditioned. Ledoit and Wolf [13] proposed a shrinkage estimator that has become a standard preprocessing step. DeMiguel, Garlappi, and Uppal [14] demonstrated that the naive $1/N$ portfolio frequently outperforms optimized portfolios out-of-sample due to estimation error – this establishes a critical benchmark that any sophisticated strategy must beat after transaction costs. Jacobs, Levy, and Markowitz [15] extended the framework to factor exposures and realistic short positions, while Bertsimas and Shioda [2] developed specialized algorithms for cardinality-constrained quadratic optimization and established its NP-hardness.

Quantum and quantum-inspired portfolio optimization. Rosenberg et al. [5] formulated the optimal trading trajectory as a QUBO and solved it on a D-Wave annealer. Elsokkary et al. [6] applied D-Wave to Abu Dhabi Securities Exchange data. Venturelli and Kondratyev [7] introduced reverse quantum annealing. Grant, Humble, and Stober [8] benchmarked quantum annealing controls for portfolios. Mugel et al. [9] and Palmer et al. [10] extended formulations to real datasets with tensor networks and investment bands, respectively. Herman et al. [4] surveyed quantum

computing for finance and identified portfolio optimization as one of the most promising near-term applications.

Foundational quantum-inspired algorithms and tools. Kadowaki and Nishimori [16] established the theoretical foundations of quantum annealing; Farhi, Goldstone, and Gutmann [17] proposed QAOA. Johnson et al. [18] demonstrated the first commercial quantum annealer. Lucas [3] compiled Ising formulations for NP problems including subset-selection problems closely related to cardinality-constrained portfolio optimization; the penalty heuristic used in this work follows his approach. McGeoch [19] provided a textbook treatment of adiabatic computation. The computational framework here relies on CVXPY [12], SCIP [11], and Qiskit [20]. Fedorov et al. [21] reviewed the state of quantum technologies in Russia, covering hardware and algorithms relevant to the domestic ecosystem.

Existing works in quantum portfolio optimization focus on small-to-medium in-sample demonstrations ($N \leq 50$) without out-of-sample validation. Systematic walk-forward evaluation on real data from two markets with quantitative decomposition of the scalability–quality trade-off – at this level of methodological rigor, comparable works are absent in the literature.

4. Problem Statement

Let $\mu \in R^N$ be the vector of annualized expected returns and $\Sigma \in R^{N\times N}$ the annualized covariance matrix. The cardinality-constrained mean-variance problem is:

$$\min_{w,z}\left(-\mu^\top w + \frac{\lambda}{2} w^\top \Sigma w\right)$$

subject to $\sum_i w_i = 1$, $w_{min} z_i \leq w_i \leq w_{max} z_i$, $\sum_i z_i \leq K$, $z \in \{0,1\}^N$, where $\lambda > 0$ is risk aversion, K is the maximum cardinality, and w_{min}, w_{max} are per-asset bounds. Binaries are linked to weights via big-M constraints.

For the quantum-inspired solver the problem is reformulated as a binary-inclusion QUBO with $x_i \in \{0,1\}$ and implicit equal weights $w_i = x_i/K$:

$$H(x) = -\frac{1}{K}\mu^\top x + \frac{\lambda}{2K^2} x^\top \Sigma x + \lambda_c \left(\sum_i x_i - K\right)^2,$$

with cardinality penalty $\lambda_c = 100 \cdot max$ following Lucas [3]. A two-stage pipeline then refines the binary solution: SA selects a subset, after which continuous convex optimization determines the weights.

Evaluation criteria. For scalability (Experiment A): median wall time (s) and optimality gap (%) over three seeds. For business quality (Experiment D): CAGR (%), Sharpe, Sortino, maximum drawdown (%), annualized turnover (%), annualized transaction costs (bps), and net excess return versus $1/N$ (bps/year).

5. General Design

Unified problem interface. A single `PortfolioProblem` dataclass encapsulates μ, Σ, λ, K, w_{min}, w_{max}, and a `build_selection_bqm()` function constructs a Binary Quadratic Model consumed identically by the neal sampler, the Qiskit QAOA circuit, and the brute-force enumerator. This design eliminates the confounding factor of different solvers solving subtly different problems.

Two-stage hybrid pipeline. Both neal and QAOA follow a two-stage pipeline: the binary solver selects a subset $S \subseteq \{1, \ldots, N\}$ of at most K assets; a continuous convex solver (CVXPY with CLARABEL/ECOS/SCS) determines the optimal weights over S subject to the sum-to-one and box constraints. This matches the practical workflow of asset managers who first screen a universe and then optimize allocation.

Persistent warm-start for walk-forward. Naive re-running of SA at each fold produces high fold-to-fold turnover. We introduce `solve_with_neal_selection (warm_start_subset, prev_weights, turnover_cost_bps)`, which adds the previous fold's subset to the candidate list and ranks candidates by objective plus a turnover penalty turnover_cost_bps $\cdot \| w - w_{\text{prev}} \|_1 / (2 \cdot 10000)$. A new subset replaces the previous one only if its return advantage exceeds the expected rebalancing cost. On a test instance ($N = 30$, $K = 8$), this reduces fold-to-fold turnover from 0.331 to 0.000 with an objective hit of only +0.10%.

Data cleaning with split-artifact rebasing. Historical price data from public APIs often contain unadjusted corporate actions. On MOEX ISS data, the VTBR reverse split on 2024-07-15 (5000:1) was not adjusted, producing a spurious daily return of 466,200%. Our post-hoc rebasing procedure: for any day with $|\text{daily return}) > 0.60$, the price series from that date is divided by $(1 + r)$, neutralizing the artifact without discarding surrounding observations. A total of 7 corrections were applied on MOEX data and 1 on S&P 500 data.

6. Implementation

Software stack. Python 3.11; cvxpy 1.6.7; pyscipopt 6.1.0 (SCIP 9.x); dwave-ocean-sdk 9.3.0; Qiskit 2.3.1 with Qiskit Aer 0.17.2; cuQuantum for GPU-accelerated statevector simulation (prepared for the HSE supercomputer but not executed locally); yfinance 1.2.2; apimoex 1.4.0; scikit-learn 1.5 for the Ledoit–Wolf shrinkage estimator. The repository is organized as `src/cif/{data, classical, qubo, quantum, backtest, metrics}`.

Solvers. (i) *Continuous MVO* via CVXPY serves as the efficient-frontier reference. (ii) *Brute-force enumeration* over integer compositions on a discrete grid provides a certified ground truth for $N \leq 12$. (iii) *SCIP MIQP* via pyscipopt with big-M linking between z_i and w_i. (iv) *ECOS_BB*, the embedded CVXPY branch-and-bound on top of the conic solver ECOS, as a lightweight alternative. (v) *D-Wave neal* SA with `num_reads=100`, `num_sweeps=1000` on the binary-inclusion QUBO; the subset is then refined by continuous CVXPY. (vi) *QAOA on Qiskit Aer* (CPU and GPU): standard `qaoa_ansatz(cost_op, reps=p)` for $p \in \{1,2,3\}$, parameters optimized via SciPy COBYLA with 60 iterations and 4096 shots; Pauli-Z transformation $x_i = (1 - z_i)/2$. The GPU path via `AerSimulator(device='GPU')` is prepared for up to ~24 qubits on A100/H100 but was not executed locally and is reported as planned future work.

Data. *S&P 500*: 90 tickers from the top-100 by market cap, retained after discarding 10 with history covering less than 90% of the window, period 2012-05-15 to 2025-12-30 (3,427 trading days); yfinance source with universe snapshot from Wikipedia. *MOEX*: 28 of 46 IMOEX tickers retained after discarding 18 with insufficient history (predominantly 2020+ IPOs), period 2014-06-09 to 2025-12-30 (2,895 trading days); MOEX ISS API via apimoex (`statistics/engines/stock/markets/index/analytics/IMOEX`) with pagination `limit=200`. Expected returns are annualized log-returns; the covariance matrix is Ledoit–Wolf shrunk [13].

Reproducibility. The scalability and walk-forward timings (Experiments A and D) were measured on a single workstation – a 4-core / 8-thread Intel Core i5-10210U CPU, 16 GB RAM, Ubuntu 24.04 LTS – with all solvers run at default settings; reported wall times are medians over three fixed random seeds (42, 123, 7). The principal solver parameters are: neal – num_reads = 100, num_sweeps = 1000 with the default geometric temperature schedule; SCIP – a 300 s per-instance time limit; ECOS_BB – CVXPY defaults; cardinality penalty λ_c = 100·max. The complete source code, the exact dependency versions, the raw per-run result files, and the data-provenance records (SHA-256 hashes and snapshot dates for every price series, together with the corporate-action cleaning log) are openly available [22].

7. Evaluation

7.1 Experiment A: Scalability on Synthetic Instances

Setup. Synthetic instances for $N \in \{20,30,50,75,100,150,200\}$, $K = \lceil N/4 \rceil$, $\lambda = 2.0$, dense random covariance, three random seeds per N (21 runs). Three solvers: SCIP, ECOS_BB, neal. Table 1 reports median wall time and optimality gap relative to the best feasible objective found.

Table 1. Experiment A: median wall time (seconds) and optimality gap (%) over 3 random seeds, synthetic dense covariance matrices, $K = \lceil N/4 \rceil$, $\lambda = 2.0$.

N	SCIP median, s	ECOS_BB median, s	neal median, s	neal gap, %
20	0.59	0.04	1.36	2.46
30	1.95	0.17	2.37	1.47
50	3.27	0.25	1.37	7.41
75	7.20	1.05	2.15	7.75
100	8.60	2.43	3.02	10.99
150	12.85	11.08	5.75	11.14
200	69.95 (std 152.77)	32.04	9.41	12.58

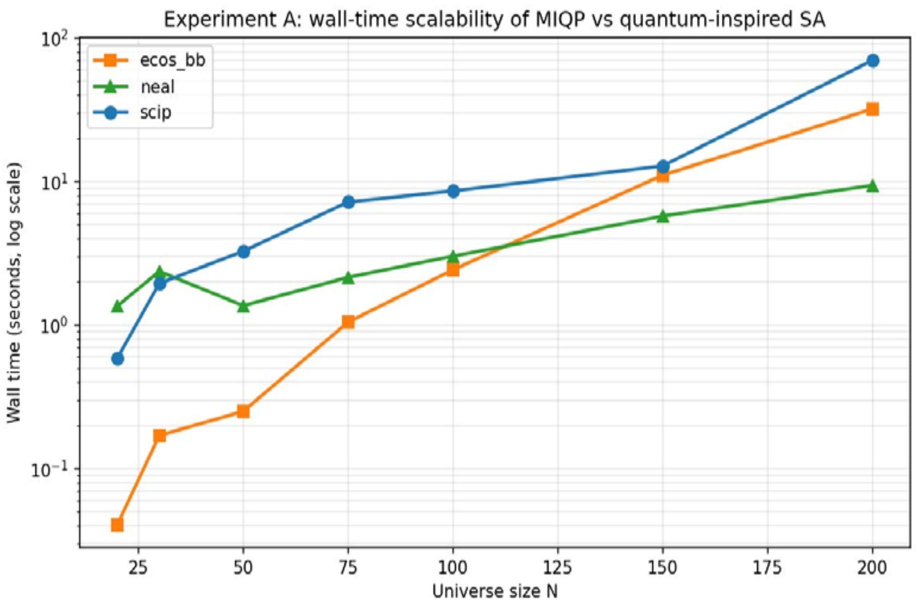

Fig. 1. Experiment A: wall-time scalability of MIQP solvers (SCIP, ECOS_BB) and quantum-inspired SA (neal) as a function of universe size N (log-log scale).

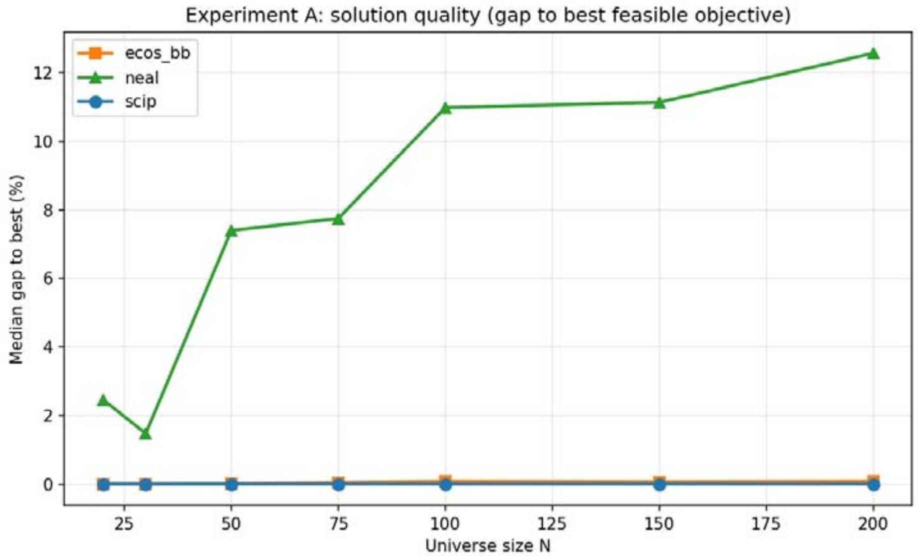

Fig. 2. Experiment A: solution quality – median optimality gap to best feasible objective (%) as a function of universe size N.

Discussion. Three findings emerge. First, neal becomes the fastest solver at $N \geq 150$: at $N = 200$ it is ~7× faster than median SCIP (9.4 vs 70.0 s) and 3.4× faster than ECOS_BB (9.4 vs 32.0 s) – visible as the flattening neal curve in Fig. 1. Second, SCIP exhibits extreme wall-time variance at large N (std 152.77 s at $N = 200$, more than 2× the median), a consequence of adaptive branching on random instances; on some seeds SCIP did not terminate within a 5-minute timeout, which is practically problematic. Third, the neal gap plateaus at 11–13% for $N \geq 100$ (Fig. 2) and, within the sampler's parameter space, does not decrease with more sweeps or reads; its sources are examined in Section 7.2. One contributing factor is the binary-inclusion QUBO itself: the equal-weight assumption $w_i = x_i / K$ is a poor approximation when assets have heterogeneous variances and correlations. The two-stage pipeline partially corrects this by continuous refinement, but the subset itself may be suboptimal. SCIP and ECOS_BB consistently achieve gaps below 0.1%.

Limitations. Dense random covariance does not reflect the factor structure of real financial covariance, where effective dimensionality is much lower. The scalability advantage on dense random instances may not transfer directly to structured financial problems. A further scope limitation is that we benchmark only open-source solvers and do not compare against commercial MIQP packages (Gurobi, CPLEX, MOSEK) – precisely the baselines made inaccessible by the 2022 restrictions and therefore outside the remit of this study. Solution quality is consequently measured against certified brute-force optima for N ≤ 12 and the best feasible objective found by any benchmarked solver for N > 12, which bounds the absolute optimality gap independently of any commercial reference.

7.2 Experiment D: Walk-Forward Backtest on Real Markets

Setup. Rolling walk-forward: 756-day train window (~3 years), 21-day test window (~1 month), 21-day step – 127 non-overlapping folds on S&P 500 and ~85 on MOEX. At each fold, μ and Σ are re-estimated via Ledoit–Wolf shrinkage and all solvers re-run. Three formulation variants were

tested (v1 soft, v2 tight, v3 = v2 with persistent warm-start for neal); v3 is the final reported configuration. Transaction costs: 10 bps per rebalance on S&P 500, 30 bps on MOEX (reflecting lower liquidity and higher brokerage). Five strategies: (1) $1/N$ equal weight; (2) Continuous MVO (upper bound); (3) Top-K rounding; (4) SCIP MIQP; (5) Neal SA with persistent warm-start.

Table 2. Experiment D: S&P 500 walk-forward, 127 folds, K = 15, w_{max} = 0.10, λ = 5.0, 10 bps costs (v3).

Strategy	CAGR, %	Sharpe	Sortino	Max DD, %	Ann. turnover, %	Ann. costs, bps	Net excess vs 1/N, bps
1/N equal weight	15.99	0.948	0.20	−31.96	4.72	0.47	0
Continuous MVO	16.41	0.877	0.25	−26.42	210.74	21.07	+42
Top-K rounding	16.40	0.876	0.25	−26.42	210.80	21.08	+40
SCIP MIQP	16.38	0.875	0.25	−26.44	210.96	21.10	+39
Neal SA (persistent)	14.40	0.836	0.21	−28.22	267.60	26.76	−159

Table 3. Experiment D: MOEX walk-forward, ~85 folds, K = 8, w_{max} = 0.20, λ = 5.0, 30 bps costs (v3).

Strategy	CAGR, %	Sharpe	Sortino	Max DD, %	Ann. turnover, %	Ann. costs, bps	Net excess vs 1/N, bps
1/N equal weight	5.68	0.374	0.09	−41.55	5.95	1.79	0
Continuous MVO	4.57	0.331	0.10	−39.76	150.45	45.14	−111
Top-K rounding	4.68	0.336	0.10	−39.76	150.79	45.24	−100
SCIP MIQP	4.68	0.336	0.10	−39.77	150.76	45.23	−100
Neal SA (persistent)	3.38	0.270	0.08	−40.95	180.05	54.02	−229

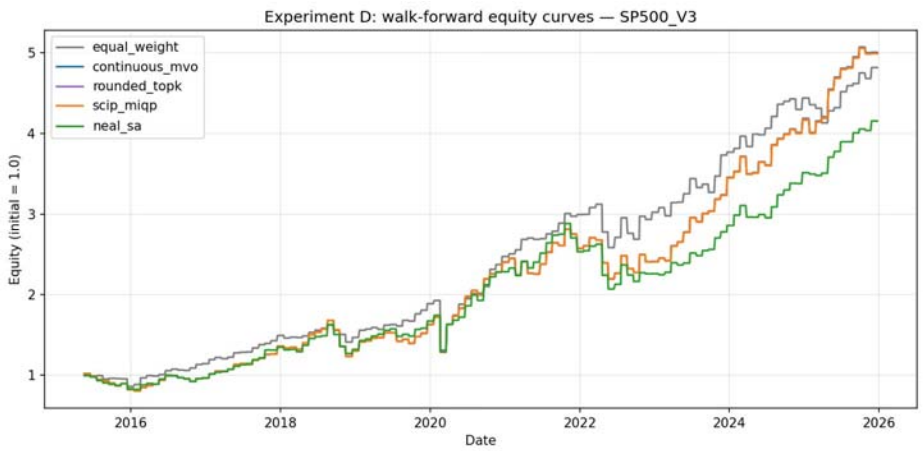

Fig. 3. Experiment D: walk-forward equity curves for all five strategies on S&P 500 (v3, initial equity = 1.0).

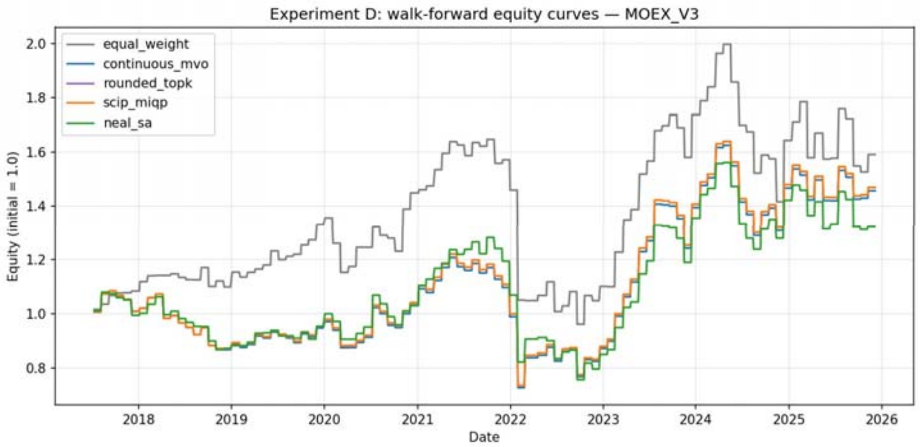

Fig. 4. Experiment D: walk-forward equity curves for all five strategies on MOEX (v3, initial equity = 1.0).

Discussion.

Four findings emerge.

Business value on S&P 500. Discrete MVO strategies deliver +39 to +42 bps/year over $1/N$ after transaction costs (Table 2, Fig. 3). On $10M AUM this is $3,900–4,200/year; scaling linearly to $390k–420k/year on $1B AUM. This is the business effect of discrete portfolio construction as such.

Non-binding cardinality on S&P 500. Continuous MVO, Top-K rounding, and SCIP MIQP give nearly identical results (within 3 bps): the cardinality constraint at $K = 15$ is effectively non-binding – Continuous MVO with $w_{max} = 0.10$ naturally concentrates into ≤ 15 active positions due to high estimation error on a 90-equity universe. For standard MVO configurations on medium-sized universes in developed markets, cardinality may add little value unless K is set quite low.

1/N dominance on MOEX. On the Russian market, $1/N$ outperforms all MVO strategies by 100–229 bps/year (Table 3, Fig. 4) – the DeMiguel–Garlappi–Uppal paradox [14] manifesting on a non-stationary market: the February 2022 shock, market closure, and ongoing structural changes render μ and Σ estimation severely unreliable. This is a critical warning for Russian practitioners: on post-sanctions MOEX data, MVO-based optimization is counterproductive regardless of the solver.

Quantitative characterization of the neal SA gap. Neal underperforms SCIP by 159 bps/year on S&P 500 and by 129 bps/year on MOEX. The gap decomposes into (a) a 5-12% residual optimality gap (consistent with Experiment A) leading to suboptimal selection, and (b) elevated turnover (267.6% vs 210.96% on S&P 500), as SA periodically selects different subsets even with persistent warm-start. This is not a failure of quantum-inspired methods in general but a quantitative trade-off: neal offers scalability at large N, but in walk-forward the present neal-based pipeline exacts a quality penalty that renders it uncompetitive under our experimental conditions. The residual optimality gap (a) can stem from three mechanisms that our present experiments do not separate: (i) formulation error – the binary-inclusion QUBO encodes implicit equal weights $w_i = x_i/K$, so its exact optimum need not coincide with the true MIQP optimum; (ii) sampler error – simulated annealing with a finite reads/sweeps budget (100/1000) and a fixed geometric schedule has no optimality guarantee and may miss even the QUBO's own minimum; and (iii) penalty-calibration error – the cardinality multiplier λ_c = 100·max, if mis-scaled, distorts the energy landscape. The two-stage pipeline re-optimizes the weights of the selected subset by continuous convex programming and so largely neutralizes (i); the residual gap therefore stems mainly from suboptimal subset selection –

mechanisms (ii) and (iii). Isolating their contributions, via alternative metaheuristics on the identical QUBO and a penalty-weight sweep, is the priority robustness study we identify for future work (Section 8).

Limitations. The cost model is flat-rate and does not account for market impact or capacity constraints. Universes are static within each market, ignoring index reconstitution. The MOEX period spans extreme non-stationarity (2022 closure, capital controls), not representative of stable regimes.

8. Conclusion

This paper has presented a systematic empirical study of the scalability–quality trade-off for quantum-inspired methods in cardinality-constrained Markowitz portfolio optimization, evaluated against open-source classical MIQP baselines on synthetic instances and real market data.

Principal findings. (i) On the scalability axis, quantum-inspired SA (D-Wave neal) achieves a 5–10× wall-time advantage over SCIP and ECOS_BB at $N \geq 150$ on dense synthetic covariance. (ii) On the business-quality axis, neal with persistent warm-start underperforms SCIP by 100–200 bps/year in walk-forward on real data due to residual optimality gap and elevated turnover – under the present experimental conditions the neal-based binary-inclusion pipeline is not competitive for walk-forward deployment. (iii) The cardinality constraint at realistic settings is often non-binding: business value comes from the framework of weight discretization rather than from the cardinality bound itself. (iv) On the non-stationary Russian market, all MVO strategies underperform $1/N$ by 100–229 bps/year, reproducing the DeMiguel–Garlappi–Uppal paradox [14] in a context particularly relevant for Russian practitioners.

Practical implications. For Russian asset managers under sanctions-imposed constraints: SCIP (via pyscipopt and CVXPY) provides solutions within 0.1% of the certified optimum up to $N = 200$ and is a practical open-source option for discrete portfolio construction – although, because commercial solvers cannot be licensed in this setting, we do not benchmark against Gurobi or CPLEX and make no claim of parity with them; quantum-inspired SA offers speed for large-scale screening but is not yet competitive for final construction; and on post-2022 MOEX data, practitioners should exercise extreme caution with MVO-based optimization and consider robust alternatives ($1/N$, minimum-variance, shrinkage-based allocation).

Future work. (i) Real-hardware trials on the Bauman Octillion processor once registration is complete. (ii) Alternative QUBO encodings (multi-bit, one-hot, domain-wall) and a controlled ablation – alternative metaheuristics on the identical QUBO and a penalty-weight sweep – to isolate and reduce the 11–13% optimality gap. (iii) Experiment B (fixed wall-clock budget, code prepared) and Experiment C (QAOA on HSE supercomputer with cuQuantum) remain for execution. (iv) Hybrid metaheuristics combining SA with tabu search or genetic algorithms.

References

[1]. Markowitz H. Portfolio Selection. The Journal of Finance, 1952, vol. 7, no. 1, pp. 77-91. DOI: 10.1111/j.1540-6261.1952.tb01525.x.

[2]. Bertsimas D., Shioda R. Algorithm for cardinality-constrained quadratic optimization. Computational Optimization and Applications, 2009, vol. 43, pp. 1-22. DOI: 10.1007/s10589-007-9126-9.

[3]. Lucas A. Ising formulations of many NP problems. Frontiers in Physics, 2014, vol. 2, art. 5. DOI: 10.3389/fphy.2014.00005.

[4]. Herman D. et al. A Survey of Quantum Computing for Finance. arXiv preprint, 2022, DOI: 10.48550/arXiv.2201.02773.

[5]. Rosenberg G. et al. Solving the Optimal Trading Trajectory Problem Using a Quantum Annealer. IEEE Journal of Selected Topics in Signal Processing, 2016, vol. 10, no. 6, pp. 1053-1060. DOI: 10.1109/JSTSP.2016.2574703.

[6]. Elsokkary N., Khan F.S., La Torre D., Humble T.S., Gottlieb J. Financial Portfolio Management using D-Wave Quantum Optimizer: The Case of Abu Dhabi Securities Exchange. In 2017 IEEE High Performance Extreme Computing Conference, 2017, pp. 1-5. DOI: 10.1109/HPEC.2017.8091078.
[7]. Venturelli D., Kondratyev A. Reverse quantum annealing approach to portfolio optimization problems. Quantum Machine Intelligence, 2019, vol. 1, pp. 17-30. DOI: 10.1007/s42484-019-00001-w.
[8]. Grant E., Humble T.S., Stober B. Benchmarking Quantum Annealing Controls with Portfolio Optimization. Physical Review Applied, 2021, vol. 15, art. 014012. DOI: 10.1103/PhysRevApplied.15.014012.
[9]. Mugel S. et al. Dynamic portfolio optimization with real datasets using quantum processors and quantum-inspired tensor networks. Physical Review Research, 2022, vol. 4, art. 013006. DOI: 10.1103/PhysRevResearch.4.013006.
[10]. Palmer S., Sahin S., Hernandez R., Mugel S., Orús R. Quantum Portfolio Optimization with Investment Bands and Target Volatility. International Journal of Theoretical and Applied Finance, 2022, vol. 25, 2250019. DOI: 10.1142/S0219024922500194.
[11]. Bolusani S. et al. The SCIP Optimization Suite 9.0. ZIB-Report 24-02, Zuse Institute Berlin, 2024.
[12]. Diamond S., Boyd S. CVXPY: A Python-Embedded Modeling Language for Convex Optimization. Journal of Machine Learning Research, 2016, vol. 17, no. 83, pp. 1-5.
[13]. Ledoit O., Wolf M. A Well-Conditioned Estimator for Large-Dimensional Covariance Matrices. Journal of Multivariate Analysis, 2004, vol. 88, issue 2, pp. 365-411. DOI: 10.1016/S0047-259X(03)00096-4.
[14]. DeMiguel V., Garlappi L., Uppal R. Optimal Versus Naive Diversification: How Inefficient Is the 1/N Portfolio Strategy? The Review of Financial Studies, 2009, vol. 22, issue 5, pp. 1915-1953. DOI: 10.1093/rfs/hhm075.
[15]. Jacobs B.I., Levy K.N., Markowitz H.M. Portfolio Optimization with Factors, Scenarios, and Realistic Short Positions. Operations Research, 2005, vol. 53, no. 4, pp. 586-599. DOI: 10.1287/opre.1050.0212.
[16]. Kadowaki T., Nishimori H. Quantum annealing in the transverse Ising model. Physical Review E, 1998, vol. 58, no. 5, pp. 5355-5363. DOI: 10.1103/PhysRevE.58.5355.
[17]. Farhi E., Goldstone J., Gutmann S. A Quantum Approximate Optimization Algorithm. arXiv preprint, 2014. DOI: 10.48550/arXiv.1411.4028.
[18]. Johnson M.W. et al. Quantum annealing with manufactured spins. Nature, 2011, vol. 473, pp. 194-198. DOI: 10.1038/nature10012.
[19]. McGeoch C.C. Adiabatic Quantum Computation and Quantum Annealing: Theory and Practice. Morgan & Claypool Publishers, 2014, 93 p.
[20]. Javadi-Abhari A. et al. Quantum computing with Qiskit. arXiv preprint, 2024. DOI: 10.48550/arXiv.2405.08810.
[21]. Fedorov A.K., Kiktenko E.O., Lvovsky A.I. et al. Quantum technologies in Russia. Quantum Science and Technology, 2019, vol. 4, 040501. DOI: 10.1088/2058-9565/ab4472.
[22]. Patrushev K. Available at https://github.com/KonstantinPatrushev/computing_in_finance, accessed 22.07.2026.

Информация об авторах / Information about authors

Сергей Михайлович АВДОШИН – кандидат технических наук, профессор, профессор департамента компьютерной инженерии МИЭМ НИУ ВШЭ. Сфера научных интересов: программная инженерия, разработка и анализ компьютерных алгоритмов, параллельные и распределённые вычисления, информационная безопасность, технология блокчейн.

Sergey Michailovich AVDOSHIN – Cand. Sci. (Tech.), Prof., Professor at the Department of Computer Engineering, MIEM, HSE University. Research interests: software engineering, design and analysis of computer algorithms, parallel and distributed computing, information security, blockchain technology.

Константин Алексеевич ПАТРУШЕВ – магистрант департамента бизнес-информатики НИУ ВШЭ. Сфера научных интересов: финансовая математика, квантовые вычисления в оптимизации, машинное обучение в портфельном управлении.

Konstantin Alekseevich PATRUSHEV – Master’s student at the Department of Business Informatics, HSE University. Research interests: financial mathematics, quantum computing for optimization, machine learning for portfolio management.