# Portfolio Selection Optimisation Using Genetic Algorithm (GA)

Thurka Puvanendran

# Literature Review: Portfolio Optimisation

- Modern Portfolio Theory (MPT)
  - A core part of the framework is Harry Markowitz's mean-variance model (1952),
  - Tells us factors on how to build a good portfolio:
    - The trade-off between risk and return
    - Diversification: risk can be minimized by investing in a range of assets
- Capital Asset Pricing Model (CAPM)
  - Introduced by Sharpe (1964), based on Markowitz's model
  - Used to give a fair price on shares
  - Looks at the relationship between systematic/market risk and return
    - Bearing higher systematic risk rewards the investors with higher return
    - This relationship concludes that the market must be efficient → The best investment is the market

# Literature Review: Portfolio Optimisation

## INDEX TRACKING

- Index tracking is a form of passive fund management.
- For investors who are risk averse
- Proxy for investing the market
- This form of investment is to be as close to a reference financial index (such as CAC, FTSE,...) as possible using a subset of assets, rather than investing in every asset in the reference financial index

# Literature Review: Genetic Algorithm

Based on Darwin's theory of evolution

Population-based algorithm → looks at multiple solutions → works quickly

Fitness function →selection operation → crossover operation → mutation operation → repeat until stopping criteria Is met

Intelligent Population, helps GA look into finding solution quicker

Need to choose each aspects of the GA carefully or a good solution won't be found

# Aims and Objectives

- Develop a portfolio optimisation model that considers practical issues such as cardinality (i.e., the maximum number of assets in the portfolio) and threshold constraints (i.e., minimum or maximum weights for certain assets).
- Evaluate the effectiveness of the model in creating efficient portfolios in different markets.
- Implement a genetic algorithm (GA) in which the portfolio closely follows the market based on historical data.
- Test the GA-based model on four different markets (CAC 40, DAX 40, FTSE 100, S&P 500)
- Fine-tune the GA code by adjusting parameters.

# Index Tracking Model

- On the right is the cardinality-constrained optimization problem used to track the index.
- This model acted as our fitness function telling us how good each portfolios were by tracking the index tracking error.

**SETS:**

| | |
|---|---|
| $T = 1..N$ | denotes time |
| $I = 1..A$ | denotes assets |
| $I' = 1..F$ | denotes assets forced in |
| $I'' = I - I'$ | denotes assets that may or may not be in portfolio |

**DATA:**

| | |
|---|---|
| $r_{i,t}$ | denotes return of asset $i$ at time $t$ |
| $rm_t$ | denotes market return at time $t$ |
| $l$ | denotes minimum investment if an asset is to be held |
| $U$ | denotes maximum investment if an asset is to be held |
| $K$ | denotes maximum stock chosen |
| $T'$ | denotes target return |

**VARIABLES:**

| | |
|---|---|
| $\omega_i \geq 0$ | denotes weight in asset $i$, $i \in I$ |
| $\delta_i = 0/1$ | binary variable, indicating 1 if we have invested in asset $i$, 0 if we haven't, $i \in I''$ |
| $u_t \geq 0$ | amount over the index return $t$, $t \in T$ |
| $o_t \geq 0$ | amount under the index return $t$, $t \in T$ |
| $P_t$ | denotes portfolio return at time $t$, $t \in T$ |

$$\text{minimise:} \quad \sum_{t \in T} (o_t + u_t)$$

$$\text{subject to:} \quad \sum_{i \in I} \omega_i = 1 \qquad (1)$$

$$\sum_{i \in I} r_{i,t} \cdot \omega_i = rm_t + o_t - u_t \qquad t \in T \qquad (2)$$

$$l\delta_i \leq \omega_i \leq U\delta_i \qquad i \in I'' \qquad (3)$$

$$l \leq \omega_i \leq U \qquad i \in I' \qquad (4)$$

$$|I'| + \sum_{i \in I''} \delta_i \leq K \qquad (5)$$

$$P_t = \sum_{i \in I} r_{i,t} \cdot \omega_i \qquad t \in T \qquad (6)$$

$$\sum_{t \in T} \frac{P_t}{|T|} \geq T' \qquad (7)$$

# Genetic algorithm Pseudocode

- Here is the pseudocode representing the AMPL code of the genetic algorithm.
- The GA only forms half of the portfolio called set I''. The rest is formed by the CPLEX solver.

$P_p$ := sets of assets for portfolio $p$;

$C_a$ := capital market value for asset $a$ ;

$W_a$ := weight of asset $a$ ;

$E_p$ := tracking error for portfolio $p$;

$B_g$ := best solution for Generation $g$;

**for** $p = 1..P$ **do**

    **if** $p = 1$ **then**

        SORT ($C_a$) by capital market value;

        Let $I' =$ the $\frac{K}{2}$ assets of highest capital market value;

    **else if** $p = 2$ **then**

        SORT ($W_a$) by weight from the LP relaxed solution;

        Let $I' =$ the $\frac{K}{2}$ assets of highest weight;

    **else**

        Generate randomly $\frac{K}{2}$ assets Let $I' = \frac{K}{2}$ randomly generated assets;

    **end**

    $I'' = I - I'$;

    SOLVE;

    SAVE $E_p$;

    Let $P_p = I' \cup \{i \in I'' | \delta_i = 1\}$;

**end**

Let $g = 1$;

Let $B_1 = \min_{p \in P} E_p$;

**repeat**

    **for** $p = 2..P$ **do**

        Let $T = P_m \cup P_f$;

        $W'_i = max(w_{m,i}\ , w_{f,i}) \quad i \in T$;

        SORT ($W'_i$) by weight;

        Let $I' = \frac{K}{2}$ assets of highest weight in $T$;

        SOLVE;

        SAVE $E_p$;

        Let $P_p = I' \cup \{i \in I'' | \delta_i = 1\}$;

        Let $g = g + 1$;

    **end**

    Let $B_g = \min_{p \in P} E_p$

**until** $B_g = B_{g-1}$ *where* $B_g$ *is the best portfolio generation*;

# Results

| Market | GA solution limit 10 | CPU TIME | GA solution limit 20 | CPU TIME2 |
|---|---|---|---|---|
| CAC 40 | 0.000889017 | 2090.25 | 0.000889017 | 2094.83 |
| DAX 40 | 0.00174613 | 298.109 | 0.00174613 | 654.766 |
| FTSE 100 | 0.000547758 | 17536.1 | 0.000547943 | 29656.8 |
| S&P 500 | 0.000292391 | 45158.2 | 0.000278454 | 92738.7 |

GA solution limit 10 v. GA solution limit 20

| Market | GA child-loop 1 | CPU TIME | GA child-loop 2 | CPU TIME2 |
|---|---|---|---|---|
| CAC 40 | 0.000889017 | 2094.83 | 0.000889017 | 2216.7 |
| DAX 40 | 0.00174613 | 654.766 | 0.00174613 | 478.453 |
| FTSE 100 | 0.000547943 | 29656.8 | 0.000547758 | 27115.3 |
| S&P 500 | 0.000278454 | 92738.7 | 0.000281021 | 91472.4 |

GA child-loop 1 v. GA child-loop 2

| Market | GA (Year 1) | CPU TIME | GA (Year 2) | CPU TIME2 |
|---|---|---|---|---|
| CAC 40 | 0.000889017 | 2094.83 | 0.00179318 | 9.95312 |
| DAX 40 | 0.00174613 | 654.766 | 0.00429514 | 2.39062 |
| FTSE 100 | 0.000547758 | 17536.1 | 0.000526216 | 41699 |
| S&P 500 | 0.000278454 | 92738.7 | 0.000741659 | 66381.4 |

GA (Year 1) v. GA (Year 2)

| Market | Benchmark | CPU TIME | GA | CPU TIME2 |
|---|---|---|---|---|
| CAC 40 | 0.000889017 | 746.578 | 0.000889017 | 2094.83 |
| DAX 40 | 0.001746129 | 84 | 0.00174613 | 654.766 |
| FTSE 100 | 0.000559224 | 715.312 | 0.000547758 | 17536.1 |
| S&P 500 | 0.000372042 | 4004.06 | 0.000278454 | 92738.7 |

Benchmark v. GA

# Conclusion

- The genetic algorithm worked extremely well with us in achieving our expected results
- Year 1 and Year 2 data were not similar but that is most likely when the data was taken
- Child-loop 2 is quicker but doesn't necessarily produce a better solution
- To improve the code, running the code at a higher solution limit and time limit would lead us to smaller index tracking errors.
- Additionally, add mutation operation to maintain diversity
- Compare different types of GA to know how successful the code was