A blue text on a black background

Description automatically generated

**Imperial College Business School**

**MSc Financial Engineering and Risk Management – Computational Finance with C++**

# **Markowitz Portfolio Optimization and Backtesting with C++**

**CID: 06014448  
Author: Bach Le Hoang Nguyen   
Course:** Computational Finance with C++ **Module Leader:** Professor. Panos Parpas **Submission Date:** 27 May 2025

Table of Contents

[1. Introduction 3](#_Toc199241594)

[2. Software Structure 3](#_Toc199241595)

[2.1. Logical Components and Function Structure 4](#_Toc199241596)

[2.2. Data Input and Storage 4](#_Toc199241597)

[2.3. Estimating Mean and Covariance 4](#_Toc199241598)

[2.4. Portfolio Optimization Using Conjugate Gradient 4](#_Toc199241599)

[2.5. Code: Step by step 5](#_Toc199241600)

[3. Evaluation of Portfolio Performance 7](#_Toc199241601)

[3.1. Overview 7](#_Toc199241602)

[3.2. Performance Evaluation 7](#_Toc199241603)

[3.2.1. Return vs. Risk (Efficient Frontier) 7](#_Toc199241604)

[3.2.2. Summary Statistic 9](#_Toc199241605)

[3.2.3. Portfolio Weight Characteristics 10](#_Toc199241606)

[3.2.4. Solver Performance: Conjugate Gradient Method 10](#_Toc199241607)

[4. Conclusion 11](#_Toc199241608)

[5. Appendix: Source Code 12](#_Toc199241609)

[5.1.read\_data.h – Function declarations 12](#_Toc199241610)

[5.2.read\_data.cpp – Main program logic + CG solver 12](#_Toc199241611)

[5.3.conjgrad.m – MATLAB reference implementation for CG 18](#_Toc199241612)

[5.4.testSolutions.m – Validation script 19](#_Toc199241613)

1. **Introduction**

Portfolio optimization plays a huge role modern investment management. The Markowitz mean-variance framework which was developed in the 1950 is remains one of the foundational models in this domain. By balancing expected return against risk, the model provides a structured way to select portfolios that lie within the efficient frontier, indicate the best possible returns for a given level of risk. The aim of this project is to implement and test the Markowitz portfolio optimization model using C++. This project's goal is to use C++ to put the Markowitz portfolio optimization model into action and evaluate it. The goal isn't simply to solve the optimization problem but it's also to provide a strong back-testing framework that shows how well such a technique would operate in a real-world. The dataset consists of daily returns for 83 FTSE 100 companies over 700 trading days. Using a rolling in-sample window of 100 days followed by a 12-day out-of-sample period, portfolios are rebalanced repeatedly to assess how performance evolves over time. A key feature of this project is the use of the Conjugate Gradient method to solve the linear system arising from the Karush-Kuhn-Tucker (KKT) conditions. The objective is to assess the model's performance across various return targets, to comprehend the risk-reward trade-off, and to illustrate performance indicators including the Sharpe ratio, realized volatility, and out-of-sample returns. This project connects theoretical finance with real algorithmic implementation, offering quantitative insights and hands-on coding expertise with C++.

## **Software Structure**

This project was developed using modular C++ code to implement the Markowitz portfolio optimization model in a rolling-window backtesting framework.

* There are four main components in the program: (1) a module for reading the data file, (2) a set of functions to calculate statistics like average returns and covariances, (3) a solver that finds the optimal portfolio weights, and (4) a main program that ties everything together, runs the back-testing, and exports the results.

### **Logical Components and Function Structure**

| **Component** | **Role** | **Related Functions** |
| --- | --- | --- |
| **CSV Reader** | Reads the dataset from file | readData(), string\_to\_double() |
| **Data Processor** | Computes mean and covariance of asset returns | computeMean(), computeCovariance() |
| **Solver** | Solves the KKT system to get portfolio weights | conjugateGradient() |
| **Controller** | Runs the full workflow step by step | main() |

These functions are declared in **read\_data.h** and implemented in read\_data.cpp. The CSV reader class is also defined in **csv.h** and csv.cpp, and is used to parse **asset\_returns.csv**

### **Data Input and Storage**

The dataset contains the daily returns for 83 FTSE 100 companies over a period of 700 days. The function **readData()** loads this information into a 2D matrix.

### **Estimating Mean and Covariance**

Then, the program computes two important statistics for each rolling window of 100 days, these are calculated using the given formulas:

* **computeMean()** — loops through the return matrix and averages the values
* **computeCovariance()** — measures how the returns of assets vary together

Both functions use **pass-by-reference**, meaning they directly update the output variables (vectors or matrices) without returning them.

### **Portfolio Optimization Using Conjugate Gradient**

After computing the mean and covariance, the program sets up a system of equations that represent the Markowitz optimization problem with return and budget constraints.

The solver function **conjugateGradient()** takes the matrix and right-hand side, and solves for the vector of weights w. It uses an iterative algorithm that avoids matrix inversion and instead updates the solution step by step until the result is accurate enough.

### **Code: Step by step**

#### The entire workflow of the program is run by the **main()** function which is located in **read\_data.cpp**.

#### **Step 1: Setup**

#### The program starts with setting up the environment. It defines by the inputs of the number of assets (83), the total number of return days (700), the in-sample window size (100 days), and the step size for rebalancing (12 days). Memory is then allocated for a matrix to store all return values, and the **readData()** function is called to load data from the input CSV file.

#### **Step 2: Output File**

#### Next, the program prepares an output file named results.csv to log the results of each backtest step. It writes the column headers, which include the rolling window index, the target return, the out-of-sample return, portfolio risk, and the solver residual.

#### **Step 3: Start Rolling Backtest**

* The core of the program is a rolling-window backtest. It progresses through the dataset in 12-day steps, simulating a realistic rebalancing strategy. For each 100-day in-sample window, it first computes the mean return vector using **computeMean()** and then calculates the covariance matrix using **computeCovariance().**
* For every rolling window, the program then evaluates 21 different return targets, ranging from 0.0 to 0.10. For each target return:

1. It builds the optimization system matrix and right-hand side vector.
2. It solves the system using the **conjugateGradient()** function to obtain the optimal portfolio weights.
3. When the target return is 0.03, the system (matrix Q, vector b, and resulting weights) is exported to CSV for external verification through Octave (since MATLAB server is down)
4. It also computes the in-sample expected return and logs the residual error from the optimization.

* Following this, the program evaluates the out-of-sample performance by using the next 12 days of return data. Its re-computes the mean and covariance over that short window and applies the previously computed portfolio weights to determine the realized return and risk. These results are then saved in **results.csv**. This process repeats until the dataset is fully processed. After all windows have been evaluated, the program releases the dynamically allocated memory and exits.

|  |
| --- |
| A diagram of a company  Description automatically generated with medium confidence |

**Figure 1**: Diagram of the full program workflow, showing data input, parameter estimation

## **Evaluation of Portfolio Performanc****e**

### **Overview**

The project aims to evaluate how well the Markowitz portfolio optimization performs. So, to be able to do that, I used rolling backtest with daily returns from 83 FTSE 100 companies over 700 trading days. For each test, I built portfolios targeting 21 different return levels, from 0.0% up to 10.0%. These portfolios were updated every 12 days, based on the most recent 100 days of return data.

Three main indicators are being used to assess the strategy's efficiency and stability: out-of-sample return, standard deviation, and Sharpe ratio. These helped determine how consistent and useful the model is in a real-world scenario.

### **Performance Evaluation**

### **Return vs. Risk (Efficient Frontier)**

|  |
| --- |
| A graph with a line going up  Description automatically generated |
| A graph with a green line  Description automatically generated |

The two graphs together provide a clear view of how the Markowitz portfolio optimization performs out-of-sample. The first chart, the **efficient frontier**, shows the relationship between portfolio risk and realized return across a range of target return levels from 0% to 10%. As can be seen, the portfolios targeting higher returns tend to experience higher risk, creating the classic rising upward concave curve which is usually seen by modern portfolio theory.

The second chart which is the **average Sharpe ratio** for each target. As the model increases the target return, the Sharpe ratio begins to decline steadily, meaning risk is rising faster than return. This trend shows that there is inefficiency as we set for higher returns. Despite achieving a higher return, the **risk required to reach those returns outweighs the benefit**. Hence, leading to poor risk-adjusted performance. Together, these graphs suggest that the most reliable and stable portfolios lie in the mid-range of the target return, where both returns, and risk-adjusted performance are strong.

### **Summary Statistic**

|  |
| --- |
| A table of statistics with numbers  Description automatically generated |

The summary statistics support what the graphs show visually. At lower target returns, portfolios are more stable and efficient, delivering better risk-adjusted performance. As target returns increase, returns do rise, but risk grows faster, causing a steady drop in efficiency.

The notable inefficiencies emerge in the higher target bands, particularly from 7% to 10%. Although the average return increases at these levels, the portfolio volatility rises even faster. As a result, the Sharpe ratio continues to decline, falling to its lowest point at the 10% target. This confirms the theory that was mentioned above, the minimum realized return worsens substantially, highlighting the increased downside exposure in higher-return target portfolios.

### **Portfolio Weight Characteristics**

|  |
| --- |
| A close-up of a graph  Description automatically generated |

As can be observed, there are positive performance is mostly concentrated in the mid to high target return. However, there are also negative returns become more frequent and intense particularly around windows 16 to 24, where deep red zones reflect sharp underperformance. The right panel displays out-of-sample risk. A clear pattern emerges: the periods with the worst returns also exhibit the highest risk. Specifically, the same window ranges that show high negative return before. The portfolios targeting returns above 6% show volatility exceeding 1.0, indicating extreme instability. This strong alignment between high risk and poor return highlights a key breakdown in portfolio efficiency.

This instability is due to the model doesn’t have constraint for short selling. Without this constraint, the optimizer often assigns large negative weights to in attempt to reach high return targets, resulting in highly leveraged and unrealistic in real world. While this can make in-sample results look attractive, but it creates significant instability when tested out-of-sample, especially during volatile periods. Overall, the heatmaps make clear that higher return targets come with even higher risk. This reinforces again, the conclusion that moderate return targets offer a better balance of performance and stability, while aggressive strategies lead to unreliable and risk-heavy outcomes.

### **Solver Performance: Conjugate Gradient Method**

|  |  |
| --- | --- |
| ans (final residual norm from Octave CG) | 1.1083e-07 |
| ||Q \* x\_cpp - b|| | 6.45414e-06 |
| ||x\_cpp - x\_direct|| | 0.164557 |
| ||x\_cpp - x\_octave|| | 0.0277488 |

To verify the accuracy of the Conjugate Gradient (CG) solver implemented in C++, a validation test was conducted using Octave. The solution vector generated (sol) contains 83 asset weights and includes both positive and negative values, which is consistent with the model's unconstrained short-selling framework. The final residual norm reported by Octave was   
ndicating that the CG method successfully reached a stable and precise solution. On the other side, the C++ implementation produced a similarly low residual of confirming that the solution accurately satisfies the linear system and satisfy the requirement by the coursework.

Beyond just checking residuals, a direct comparison was made between the solution from the C++ CG solver and two other benchmarks: Octave’s exact solution obtained via matrix inversion, and Octave’s own CG solver. The difference between the C++ CG solution and the exact direct solution was approximately 0.1646, while the difference between the C++ and Octave CG solutions was much smaller at 0.0277. This means the C++ implementation behaves consistently with standard numerical methods, with only minor numerical deviations which could be due to iterative precision or floating-point operations. Overall, the test confirms that the C++ solver is both accurate and stable.

## **Conclusion**

The model performs logically in most cases. However, there are a few problems during testing and analysis. The first issue relates to how the optimizer handles aggressive return targets. When we aim for high returns like above 5%, the model would start allocating extreme weights, including large negative positions. This behavior can be seen in the C++ terminal output, where it would show weights like –1.5, –2.5, and in some cases, even more extreme. These shows the existence of short positions under the unconstrained Markowitz framework, but they are unrealistic in practice. In addition, other issue is that the model relies on just 100 days of past data to decide which assets to invest in. It’s because of the uneven amount of assets and data, the results can be misleading. This becomes a bigger problem when aiming for higher returns, since the model puts more weight on assets that happened to perform well recently. While the portfolio often meets the return target in the sample period, it usually performs worse afterward, as some of the OOS\_r outputs in the terminal are very negative. Hence, the actual return even drops below zero, showing that the model is focused on short-term patterns that make it become inefficient. This leads to less reliable performance and greater risk, especially in out-of-sample testing.

Overall, the model is stable and accurate. The residuals calculated by the Conjugate Gradient are consistently within the tolerance of 1e-6, which confirms that the numerical solution is valid, and that the optimization is solving the system correctly.

There are a few simple ways to make the model more realistic. First, implementing the limit of how much short selling it can do or completely constrain short selling, so the portfolio would avoid extreme position. Second, it could use better methods for measuring risk like shrinkage estimator, Minimum-Variance Matrix Adjustment to smooth out noisy data. Finally, adding some penalty for frequent trading and including transaction costs would help the model reflect real-world conditions and avoid rebalancing between aggressive positions.

## **Appendix: Source Code**

## **read\_data.h – Function declarations**

|  |
| --- |
| #ifndef READ\_DATA\_H  #define READ\_DATA\_H  #include <string>  #include <vector>  double string\_to\_double(const std::string& s);  void readData(double\*\* data, const std::string& fileName);  void computeMean(double\*\* data,int numberAssets,int t\_start,int window,std::vector<double>& mean);  void computeCovariance(double\*\* data,int numberAssets,int t\_start,int window,const std::vector<double>& mean,std::vector<std::vector<double> >& cov);  void conjugateGradient(const std::vector<std::vector<double> >& Q,const std::vector<double>& b,std::vector<double>& x,double epsilon = 1e-6);  #endif |

## **read\_data.cpp – Main program logic + CG solver**

|  |
| --- |
| #include "read\_data.h"  #include "csv.h"  #include <iostream>  #include <fstream>  #include <sstream>  #include <vector>  #include <cmath>  #include <numeric>  using namespace std;  //g++ -c read\_data.cpp  // g++ -c csv.cpp  // g++ -o portfolioSolver csv.o read\_data.o  // ./portfolioSolver  double string\_to\_double(const std::string& s) {  istringstream i(s);  double x;  if (!(i >> x))  return 0.0;  return x;  }  void readData(double \*\*data, const std::string& fileName)  {  char tmp[20];  ifstream file (strcpy(tmp, fileName.c\_str()));  Csv csv(file);  string line;  if (file.is\_open())  {  int i=0;  while (csv.getline(line) != 0) {  for (int j = 0; j < csv.getnfield(); j++)  {  double temp=string\_to\_double(csv.getfield(j));  //cout << "Asset " << j << ", Return "<<i<<"="<< temp<<"\n";  data[j][i]=temp;  }  i++;  }    file.close();  }  else {cout <<fileName <<" missing\n";exit(0);}  }  void computeMean(double\*\* data,int numberAssets,int t\_start,int window,vector<double>& mean) {  for (int i = 0; i < numberAssets; ++i) {  double sum = 0.0;  for (int j = 0; j < window; ++j)  sum += data[i][t\_start + j];  mean[i] = sum / window;  }  }  void computeCovariance(double\*\* data,int numberAssets,int t\_start,int window,const vector<double>& mean,vector<vector<double> >& cov) {  for (int i = 0; i < numberAssets; ++i) {  for (int j = 0; j < numberAssets; ++j) {  double sum = 0.0;  for (int k = 0; k < window; ++k) {  double ri = data[i][t\_start + k] - mean[i];  double rj = data[j][t\_start + k] - mean[j];  sum += ri \* rj;  }  cov[i][j] = sum / (window - 1);  }  }  }  void conjugateGradient(const vector< vector<double> >& Q,const vector<double>& b,vector<double>& x,double epsilon) {  int n = Q.size();  vector<double> r(n);  vector<double> p(n);  vector<double> Qp(n);  // r0 = b - Q \* x0  for (int i = 0; i < n; ++i) {  double Qxi = 0.0;  for (int j = 0; j < n; ++j){  Qxi += Q[i][j] \* x[j];  }  r[i] = b[i] - Qxi;  p[i] = r[i];  }  double rs\_old = 0.0;  for (int i = 0; i < n; ++i){  rs\_old += r[i] \* r[i];    }  for (int k = 0; k < n; ++k) {    // Qp\_k = Q \* p\_k  for (int i = 0; i < n; ++i) {  Qp[i] = 0.0;  for (int j = 0; j < n; ++j)  Qp[i] += Q[i][j] \* p[j];  }  double alpha\_num = rs\_old;  double alpha\_den = 0;  for (int i = 0; i < n; ++i){  alpha\_den += p[i] \* Qp[i];  }  double alpha = alpha\_num / alpha\_den;  // x = x + alpha \* p  for (int i = 0; i < n; ++i)  x[i] = x[i] + alpha \* p[i];  // r = r - alpha \* Qp  for (int i = 0; i < n; ++i){  r[i] -= alpha \* Qp[i];  }  double rs\_new = 0.0;  for (int i = 0; i < n; ++i){  rs\_new += r[i] \* r[i];  }  if (sqrt(rs\_new) < epsilon){  break;  }  // beta\_k = (r\_(k+1)^T \* r\_(k+1)) / (r\_k^T \* r\_k)  double beta = rs\_new / rs\_old;  // p\_(k+1) = r\_(k+1) + beta\_k \* p\_k  for (int i = 0; i < n; ++i)  p[i] = r[i] + beta \* p[i];  rs\_old = rs\_new;  }  }  int main() {  const int numberAssets = 83;  const int numberReturns = 700;  const int window = 100;  const int step = 12;  const int numberTargets = 20;  double\*\* returnMatrix = new double\*[numberAssets]; // a matrix to store the return data  //allocate memory for return data  for (int i = 0; i < numberAssets; ++i)  returnMatrix[i] = new double[numberReturns];    //read the data from the file and store it into the return matrix  string fileName="asset\_returns.csv";  readData(returnMatrix, fileName);  // returnMatrix[i][j] stores the asset i, return j value  vector<double> meanInSample(numberAssets);  vector< vector<double> > inCov(numberAssets, vector<double>(numberAssets));  ofstream results("results.csv");  results << "window,target,Return,Risk, Residual\n";  // Rolling backtest  int rollingWindow = 0;  for (int t\_start = 0; t\_start + window + step <= numberReturns; t\_start += step) {  computeMean(returnMatrix, numberAssets, t\_start, window, meanInSample);  computeCovariance(returnMatrix, numberAssets, t\_start, window, meanInSample, inCov);  for (int t = 0; t <= numberTargets; ++t) {  double target = t \* 0.10 / numberTargets;  int dimension = numberAssets + 2;      vector< vector<double> > Q(dimension, vector<double>(dimension, 0.0));  vector<double> b(dimension, 0.0);    for (int i = 0; i < numberAssets; ++i)  for (int j = 0; j < numberAssets; ++j)  Q[i][j] = inCov[i][j];  for (int i = 0; i < numberAssets; ++i) {  double ri = meanInSample[i];  Q[i][numberAssets] = -ri;  Q[i][numberAssets + 1] = -1.0;  Q[numberAssets][i] = -ri;  Q[numberAssets + 1][i] = -1.0;  }  b[numberAssets] = -target;  b[numberAssets + 1] = -1.0;  vector<double> weights(dimension, 1.0);  conjugateGradient(Q, b, weights, 1e-6);    // CSV Export for validation  if (fabs(target - 0.03) < 1e-6) {  ofstream outW("weights.csv");  for (int i = 0; i < dimension; ++i)  outW << weights[i] << "\n";  outW.close();  ofstream outB("b.csv");  for (int i = 0; i < dimension; ++i)  outB << b[i] << "\n";  outB.close();  ofstream outQ("Q.csv");  for (int i = 0; i < dimension; ++i) {  for (int j = 0; j < dimension; ++j) {  outQ << Q[i][j];  if (j < dimension - 1) outQ << ",";  }  outQ << "\n";  }  outQ.close();  }  cout << "[CG] sample weights: ";  for (int i = 0; i < 5; ++i) cout << weights[i] << " ";  cout << "\n";  // Check    double sumw = 0.0;  for (int i = 0; i < numberAssets; ++i)  sumw += weights[i];  cout << "Sum of weight = " << sumw << "\n";  double actualReturn = 0.0;  for (int i = 0; i < numberAssets; ++i)  actualReturn += meanInSample[i] \* weights[i];  cout << "Actual return = " << actualReturn << ", Target = " << target << "\n";  vector<double> Qx(dimension, 0.0);  for (int i = 0; i < dimension; ++i)  for (int j = 0; j < dimension; ++j)  Qx[i] += Q[i][j] \* weights[j];  double residual = 0.0;  for (int i = 0; i < dimension; ++i)  residual += pow(Qx[i] - b[i], 2);  residual = sqrt(residual);  cout << "Residual 'Qx - b' = " << residual << "\n";  // Out-of-sample  vector<double> meanOutSample(numberAssets);  vector< vector<double> > outCov(numberAssets, vector<double>(numberAssets));  computeMean(returnMatrix, numberAssets, t\_start + window, step, meanOutSample);  computeCovariance(returnMatrix, numberAssets, t\_start + window, step, meanOutSample, outCov);    double realizedReturn = 0.0;  for (int i = 0; i < numberAssets; ++i)  realizedReturn += meanOutSample[i] \* weights[i];  double realizedVar = 0.0;  for (int i = 0; i < numberAssets; ++i)  for (int j = 0; j < numberAssets; ++j)  realizedVar += weights[i] \* outCov[i][j] \* weights[j];  double realizedRisk = sqrt(realizedVar);  results << rollingWindow << "," << target << "," << realizedReturn << "," << realizedRisk << "," << residual << "\n";  cout << "Window["<< t\_start << "] target="<< target  << ", OOS\_r="<< realizedReturn  << ", sigma="<< realizedRisk << "\n";  }  rollingWindow++;  }  results.close();  // Cleanup  for (int i = 0; i < numberAssets; ++i)  delete[] returnMatrix[i];  delete[] returnMatrix;  return 0;  } |

## **conjgrad.m – MATLAB reference implementation for CG**

|  |
| --- |
| function [x] = conjgrad(A, b, x)  r = b - A \* x;  p = r;  rsold = r' \* r;  for i = 1:length(b)  Ap = A \* p;  alpha = rsold / (p' \* Ap);  x = x + alpha \* p;  r = r - alpha \* Ap;  rsnew = r' \* r;  if sqrt(rsnew) < 1e-10  break;  end  p = r + (rsnew / rsold) \* p;  rsold = rsnew;  end  end |

## **testSolutions.m – Validation script**

|  |
| --- |
| clear;  data=load('asset\_returns.csv');  assets=83;  tR=0.03;  means=zeros(assets,1);  for ii=1:assets  means(ii)=mean(data(:,ii));  %result=sprintf('%0.9f\n',means(ii));  %fprintf(result)  end  covMat=cov(data);  e=ones(assets,1);  A=[covMat,-means,-e; -means',0,0;-e',0,0];  rh=[zeros(assets,1);-tR;-1];  sol=A\rh  x=ones(assets+2,1);  x=conjgrad(A, rh, x);  norm(sol-x)  %eigs(covMat,20,'sm')  %length(covMat)  Q = csvread('Q.csv');  b = csvread('b.csv');  x\_cpp = csvread('weights.csv');  x\_direct = Q \ b;  x\_oct = conjgrad(Q, b, ones(size(b)));  fprintf("||Q\*x\_cpp - b|| = %g\n", norm(Q\*x\_cpp - b));  fprintf("||x\_cpp - x\_direct|| = %g\n", norm(x\_cpp - x\_direct));  fprintf("||x\_cpp - x\_octave|| = %g\n", norm(x\_cpp - x\_oct)); |