STATS 290 Project Report

Jeffrey Wong

March 15, 2011

Original Project Proposal

Below is the original project proposal submitted in January 2011.

Final Project Proposal: Trading Package

Algorithmic trading has become very popular in the past 10 years with the advent of high powered computation. In order to optimize a portfolio, traders can compute statistics such as the minimum variance portfolio, delta neutral hedging, etc. However, such statistics operate on very large datasets, and are not trivial to compute. For example, the Black-Scholes formula for option pricing is computationally streneous for large input. Also, with the markets being such a large entity, and with the advance of a global economy, traders need to keep track of large amounts of data all around the world. With this package, I would like to provide a suite of tools to extract, manage, and anallyze such data.

Currently, there are packages that can calculate option prices using Black-Scholes, but they do not represent the state-of-the-art pricing techniques. Also, these packages, such as fOption, are slow in pricing and would not be able to keep up in real time. They also do not provide tools for volatility forecasting, an essential for traders. Other packages exist to calculate minimum variance portfolios and other statistics, but no one package provides all these tools in a single suite.

Package Features

- Download both stock quotes and options chains in real time
- Be able to spawn a thread that runs in the background to continuously collect data on a list of assets
- Use real data to compute portfolio statistics
- Interface with CUDA to compute option prices in parallel
- Extrapolate implied volatility of a particular option at any given time. Provide methods to study volatility surfaces

- Use historic data to backtest trading strategies

Use Case

- User provides an input file of 100 symbols
- Package retrieves real time data for each symbol and can save the data on the users local hard drive
- User can call a function to determine the minimum variance portfolio of these 100 stocks
- User can provide a utility function of risk and return to determine the maximum utility portfolio
- To price options for these symbols, the user can call a function to extract the implied volatility over time. Support functions will be provided to model the volatility using non-parametric regression. Predicted values will be fed into the Black-Scholes formula
- User can run a simulation using historic data to test a particular strategy. The simulation will retrieve data on the 100 symbols on a given day, and will pass the data to a user-written function. The simulation will report statistics on return, risk, etc. at the end of the simulation.

RFinance

RFinance is the final package that was produced for STATS 290.

Summary of Features

- Download stock data and options data in real time through the Yahoo Finance API
- Calculates the minimum variance portfolio
- CUDA implementation of Black Scholes, but having trouble linking to R. Stand alone version can be found in the /src folder. Must have CUDA driver >= CUDA 2.3 to run.
- Determine Implied Volatility of an option through optimization functions
- Back-testing engine allows traders to measure the performance of a particular trading strategy
- (New) User can maintain a virtual portfolio, calculate its value in real time and display all risks associated with it

- (New) Time series clustering. Identifies which assets behave similarly to each other, which can be leveraged with n-pairs trading. (Only half complete)
- (New) Sweave file which can dynamically generate detailed financial reports based on a watchlist. User can schedule a task to compile this sweave file which will pull all relevant information for each asset listed in the watchlist, and compile it in 1 single pdf. See Daily Financial Report under /inst/doc

Difficulties

- Cannot spawn separate threads in R. If a user wants to collect data through the 'monitor' function and run other R commands, the user must open another instance of R.
- The optimalActions function of the back-testing engine can only support trading of 1 asset. Ideally, the engine identifies local minima and maxima, purchasing at minimums and selling at maximums. However, the solution becomes significantly harder when there are finite funds available and multiple assets to trade. This is a constrained optimization problem, but it is very difficult to set up. [Update March 15: FIXED! optimalActions can now be used to compute optimal trading strategy for any n assets. Super cool application of LP and constrOptim!!!]
- Availability of options data is very limited, so could not test the performance of non parametric models in predicting volatility curves
- Wanted to do the calculation of the Michaud Resampled Portfolio on CUDA because of its heavy matrix calculations, but ran out of time.

Wishlist

Features that would be nice to have:

- Fit time series to AR(p) processes and determine a pooled covariance matrix, which can then be used in calculating the Mahalanobis distance for better clustering performance (thanks to Noah Simon)
- Somehow obtain options data and show examples of how non parametric regression can be used to fit volatility surfaces
- Move the calculation of the Michaud Resampled Portfolio to CUDA using CUBLAS
- Implement a spider that can crawl Google Finance and report on relevant news articles. (Already have a working JAVA version from CS229 Fall 2010.)