Snakes on a Plane

Shrinking the Sample Covariance Matrix: A Pythonic Approach

|  |  |
| --- | --- |
| Jason Strimpel | For Dr. B. Boonstra |
| Illinois Institute of Technology | Independent Study, MSF 597 |

Introduction

When estimating the sample covariance matrix of stock returns as the input to a mean-variance optimization problem and the number of stocks under consideration is large relative to the number of historical return observations, the most extreme values in the sample covariance matrix have a lot of error. The mean-variance optimization routines will place large bets on the extreme value which is unreliable because these extreme values are largely error.

Ledoit and Wolf show that following the methods described in their paper, “Honey, I Shrunk the Sample Covariance Matrix” (http://papers.ssrn.com/sol3/papers.cfm?abstract\_id=433840), tracking error is reduced relative to a benchmark index and the information ratio is increased. This has obvious benefits in the market for active equity portfolio managers. While the authors implement their method in MATLAB, this experiment replicates the implementation in Python which has become a compelling choice for work in quantitative finance, trading, and statistical applications.

Environment Used

Below describes the environment (including the Python modules) used for the experiment.

* Mac OSX 10.6 (http://store.apple.com/us/product/MC573Z/A?fnode=MTY1NDAzOA)
* Python 2.7 (http://www.python.org)
* Numpy 1.6.0 (http://numpy.scipy.org/)
* SciPy 0.9.0 (http://www.scipy.org/)
* Pandas 0.4.0dev (http://code.google.com/p/pandas/)
* PyTables 2.2.1 (http://www.pytables.org/moin)
* matplotlib 1.0.1 (http://matplotlib.sourceforge.net/)
* CVXOPT 1.1.3 (http://abel.ee.ucla.edu/cvxopt/)

Calls for Help

Posts to forums seeking help.

* http://stackoverflow.com/questions/7409108/portfolio-variance-of-a-portfolio-of-n-assets-in-python

Custom Modules, Classes, and Functions Used

The experiment requires XX classes, helper (private, or non-importable) methods, and regular methods.

* class Portfolio(portfolio, benchmark) contains all methods to calculate portfolio statistics. The constructor takes two arguments:
  + portfolio: user-defined dictionary containing the expected returns of each portfolio holding, the holding period of each holding, the share quantity held of each portfolio holding, and the maximum weight any position can have in the portfolio.
  + benchmark: user defined benchmark weights of the stocks held in the portfolio.

Portfolio class helper methods:

* Portfolio.\_get\_historic\_data(ticker, start, end) returns a pandas.Series with the weekly, open, high, low, close, adjusted close, and volume between start and end for ticker.
* Portfolio\_get\_historic\_returns(ticker, start, end, offset=1) returns a pandas.Series with the offset-period return between **start** and **end** for **ticker**.
* Portfolio.\_build\_portfolio(shares) returns a pandas.DataFrame with the share quantity held and the last traded price for each portfolio holding.

Portfolio class regular methods:

* Portfolio.get\_portfolio\_historic\_returns() returns a pandas.DataFrame with the period returns for each position in the portfolio.
* Portfolio.get\_portfolio\_historic\_position\_values(shares=None) returns a Pandas.DataFrame with the periodic position value (shares times price) for each position in the portfolio.
* Portfolio.get\_portfolio\_historic\_values(shares=None) returns a Pandas.Series with the periodic portfolio value.
* Portfolio.get\_benchmark\_weights() returns a pandas.DataFrame with the benchmark weights.
* Portfolio.get\_portfolio\_weights() returns a pandas.DataFrame with the weight of each portfolio holding as a percent of the total portfolio value.
* Portfolio.get\_active\_weights() returns a pandas.DataFrame with the active weight for each portfolio holding. Active weight is defined as:

where:

= portfolio weights

= benchmark weights

* Portfolio.get\_holding\_period\_returns() returns a pandas.DataFrame with the return of the portfolio holding between start and end.
* Portfolio.get\_expected\_stock\_returns() returns a pandas.DataFrame with the expected return of each portfolio holding.
* Portfolio.get\_active\_returns() returns a pandas.DataFrame with the active return for each portfolio holding. Active return is defined as:
* Portfolio.get\_expected\_excess\_stock\_returns() returns a pandas.DataFrame with the expected excess stock returns for each stock in the benchmark. Expected excess stock return is defined as:

where:

= expected stock returns

* Portfolio.get\_covariance\_matrix() ad
* Portfolio.get\_expected\_benchmark\_return() adf
* Portfolio.get\_benchmark\_variance() asd
* Portfolio.get\_expected\_portfolio\_return() ad
* Portfolio.get\_portfolio\_variance() ads
* Portfolio.get\_expected\_excess\_portfolio\_return() ad
* Portfolio.get\_tracking\_error\_variance() ads

Yahoo class:

* class Yahoo(ticker\_list) provides access to a myriad of stock information from Yahoo! Finance (http://finance.yahoo.com). There are many methods, only one of which is used. The constructor takes one argument:
  + ticker\_list: a list of ticker symbols.
* Yahoo.get\_LastTradePriceOnly(ticker) returns the last price of ticker returned to Yahoo! Finance from the applicable exchange.

Test Suite

Ledoit and Wolf published MATLAB code (http://www.ledoit.net/covCor.m) along with the paper, the results of which are used as the test suite to which my results are compared.

Range of Inputs

Analysis

Comparison of Results

Out of sample v. sample

Answers to Questions Posed

Further Investigation

The main purpose of this project was to implement a relevant experiment in Python using many of the modules available for quantitative financial researchers. The secondary purpose was to build a framework with which one has the ability to build from and enhance as time goes on. Given the main point was to show that using a “shrunk” covariance matrix reduces tracking error relative to a benchmark index and increases the information ratio, some details were left out.

* Expected return model. There are many models available to forecast expected returns for use in a mean-variance optimized portfolio. This experiment assumes these expected returns are given. However, the program was built with the ability to incorporate a forecasting model in the future.
* Optimizer. While I was able to implement an optimizer to return active weights given constraints, I did not fully understand how to use the code therefore improvements could be made to the optimization problem.

Conclusion