Paul F. Mende

Project B: Algorithmic Trading Strategy Simulation

Objective: Use U.S. equity daily return data to implement, backtest, and analyze an equity long/short trading strategy.

Consider a market-neutral algorithmic trading strategy designed to exploit potential mean reversion in stock returns. The objective is to simulate the strategy numerically using historical market data and analyze how the strategy would have performed, had it been implemented during the appropriate time period in the past.

Universe: The securities universe U for this problem consists of all publicly traded U.S. equities during the period from 1980–2001, subject to selected filters. The primary data source for this project is derived from the University of Chicago's Center for Research in Security Prices (CRSP) database, and it consists of time-series data of daily prices, returns, volumes, etc., for all U.S. listed securities on the NYSE, AMEX, and NASDAQ exchanges.

Data and Computing Environment: This project will involve the use of SQL, Microsoft Access, and R for data extraction and analysis. The data is available in the database **crispy** on the SQL Server **hedge.mit.edu**. The selected equities have attribute `project_flag = 1` in the **equity** table. Trade dates for the simulation are all market days of the 1990's, from 1/1/1990 to 12/31/1999, inclusive*, found in the **tradedate** table. Additional technical details and assistance with set up can be discussed during lecture and recitation.

Additional Details: All returns are log returns (*i.e.*, continuously compounded), and they are adjusted for corporate actions such as splits and dividends. For annualizing returns, use 252 as the number of market days per year (and factors of $\sqrt{252}$ for volatilities, where appropriate). Be sure that all numbers are reported in appropriate units with an appropriate number of significant figures. If the mean daily return is 0.0004354275, report the annualized return as 11.0% or possibly 10.97%, not 0.1097277. Reporting more than three or four sigfigs is rarely justified. In contrast, keep full precision values for internal use during calculations and for the final bulk data upload. All graphs and tables should have clearly labeled title, caption, headings, and axes. Code and any lengthy technical details should be relegated to an appendix.

*Take care with the endpoints both in retrieving source data and in analyzing results, taking account of the time-series lag structure in both parts of this assignment.

Strategy: Consider a contrarian strategy of "selling the winners and buying the losers." Let the return on stock i in period t be $R_i(t)$ and define a "market index" to be given by the equal-weighted average return of all N stocks in the universe,

$$\overline{R}(t) \equiv \frac{1}{N} \sum_{i \in U} R_i(t) \ . \tag{1}$$

At (or near) the end of period t, the excess return of security i above the mean is $R_i(t) - \overline{R}(t)$. Use this excess return to define a trading signal,

$$s_i(t) = -\left[R_i(t) - \overline{R}(t)\right], \tag{2}$$

that can be used to rank order to candidate stock trades. A simple strategy consists of going long the top quintile of stocks according to this signal and shorting the bottom quintile. Neglect transaction costs, and rebalance the portfolio at the end of each day so that it holds the desired positions for the following tradedate. The individual weights $w_i(t)$ may be positive or negative, corresponding to long or short stock positions. Overall, the longs and shorts are balanced if $\sum_i w_i(t) = 0$.

In addition to specifying the sign of the weights, let's further require that the stocks within each quintile be equally weights and that the long/short portfolio be "fully-invested" by choosing the $w_i(t)$ on each day so that the sum of the long weights and the sum of the short weights separately add up to 100%.[†] That is,

$$\sum_{i \in L} w_i = 1 = \sum_{j \in S} |w_j|. \tag{3}$$

The time series of daily *portfolio returns*, $\pi(t)$ is given by

$$\pi(t) = \frac{\Delta MV}{MV} = \sum_{i \in U} w_i(t-1) R_i(t) \ . \tag{4}$$

1. **Strategy Simulation:** For these questions, use historical market data to simulate the strategy P/L and analyze the returns.
   (a) Plot and label a graph with the time series of daily portfolio returns $\pi(t)$ and another graph with the time series of daily market returns $\overline{R}(t)$.
   (b) What are the annualized mean return, volatility, and Sharpe ratio of the strategy and of the market average?
   (c) Are they consistent over time? Is the strategy's distribution of daily returns stationary?

[†]Lo and MacKinlay used a different weighting scheme and also a different normalization. In their case the total amount invested, or total gross market exposure, changes day to day.

(d) Are there any events (*i.e.*, market days) that are "unusual events"? Are there any stocks that are "outliers"? If so, do they have a significant effect on the overall strategy returns?

(e) What is the correlation between the strategy returns and the market return? Is the strategy in fact market-neutral or just dollar neutral?

(f) What is the correlation between the long and short sub-portfolios?[‡]

(g) How realistic is this simulation? Assume that there were no transaction costs and no market impact. How closely do you think actual returns on this strategy would track the simulated values $\pi(t)$? List data issues that need to be considered and what their impact might be.

2. **A Family of Strategies:** Now generalize the original strategy and use a lag of k days. That is, consider over- or under-reaction relative to excess returns k days in the past. How do the answers depends on the lag k? For a given value of k, use signals defined by

$$s_i^{(k)}(t) = -\left[R_i(t-k+1) - \overline{R}(t-k+1)\right]. \qquad (5)$$

(a) Make a table showing annualized mean return, volatility, and Sharpe ratio, comparing values for $k = 1, 2, \ldots, 5$. How do the results change with increasing lag? Discuss any patterns that you think are significant. Is there a value of k that is optimal?

(b) Create a flat file with the portfolio weights for these cases. The format is

- Column 1: Portfolio Manager ID `pid` (Your MIT ID number)
- Column 2: Date `d`.
- Column 3: Stock ID `id`
- Column 4: Time lag `k`
- Column 5: Portfolio weight `w` $= w_i^{(k)}(t)$
- Column 6: Version ID `vid` (Set equal to 0 for all rows)

There should be no missing or invalid values (e.g., NULL or NaN) and no duplicates or extras.

(c) Upload the file to the server in table `ProjectB` of database `upload`. (Test uploads can be done in the table `upload.dbo.ProjectB_TEST`.)

(d) Check and validate your upload carefully.

(e) If you find a mistake and need to fix your submission, upload a new file in which the `vid` column is incremented (e.g., set equal to 1 for all rows). Do not delete or update any rows in the database.

The performance table (first item) is deliverable with your project report.

[‡]The long and short sub-portfolios, respectively, consist on each day of those positions whose weights are positive or negative.

3. **Strategy R&D** (*Optional, extra credit*): Code and simulate your own strategy using the same data set. Answer the same questions as above, and compare your strategy to the initial one.

   Your strategy can be a simple variation on the strategy above, or it can be anything financially reasonable that can be evaluated based on the available data set. For example, you could choose to go long the top $x\%$ and short the bottom $y\%$, where x and/or y differ from 10. As another example, you could weight the individual postions as Lo and MacKinlay did, $w_i(t) \propto -(R_i(t) - \overline{R}(t))$.

References

Lo, A. and C. MacKinlay, 1990, "When Are Contrarian Profits Due to Stock Market Overreaction?", *Review of Financial Studies* 3, 175–206.