Daniel Wilde
557 Skyview St
El Cajon, CA 92020
(858) 356-7029

Data-Driven Investing

Optimizing Investment Returns by Analyzing Historical Stock Prices, Valuations Metrics, and other Macroeconomic Data

20th August 2021

BACKGROUND

Achieving maximum investment returns with minimal risk is a goal of many stock market investors. One part of this challenge is picking the right stocks. And the other part involves timing, or when to invest in a particular stock. Many different variables play into the value and price movement of an individual stock, as well as the overall valuation of the stock market.

The purpose of this project is to craft a data solution whereby investors can test the returns on a theoretical portfolio of stocks over a given timeframe using actual historical stock price data. The data solution will consist of financial data typically used for fundamental analysis (valuations and financial reports), historical end-of-day prices, and monthly macroeconomic data. An analyst will then be able to "back-test" a particular investing strategy given certain inputs, thresholds, and assumptions. The solution will serve as a tool for determining **where** and **when** to invest. For the purposes of this project I will focus mainly on dividend-paying stocks (from the S&P 500), fundamental analysis, and I will attempt to tie in at least 1 or 2 macroeconomic data sets to further enhance the analysis. Stock index averages (DOW, NASDAQ, S&P) will serve as the benchmark for comparison.

DATA SETS

- Yahoo Finance - stock price history, stock index history, financial reports
- Seeking Alpha - valuations, dividends, other data for fundamental analysis
- ISM Manufacturing Report on Business - macroeconomic monthly indicator
- JPMorgan Global Manufacturing PMI - macroeconomic data
- Federal Reserve Economic Data - macroeconomic data

SOLUTION DESIGN

Data Gathering

Python scripts and standard ETL tools will be used to scrape data from the web APIs of the Data Sets listed above. Starting with stocks in the S&P 500, and looking back 20 to 40 years, data will be loaded into the tables outlined below. A larger set of stocks and longer timeframe may eventually be used as time and space allow.

Database Tables

- Companies (Stock Tickers) and Major Indexes with metadata
- Prices (End of Day) for each Company / Index
- Quarterly Financial Results (Actuals) for each Company / Index
    - Forward-looking Financials if available
- Dividend Metrics by Quarter for each Company / Index
- Economic Data by Month
    - Economic Indicators - metadata
    - Values by Month for each Economic Indicator
- Portfolio Table
- Investing Strategy (metadata and strategies tested)
- Trades Table (Buy / Sell)

Portfolio Analyzer API

The Portfolio Analyzer API will allow an analyst to enter specific thresholds and rules for opening and exiting positions in stocks by entering between 15 to 20 different variables. These thresholds and rules will be used to simulate trades and “back-test” against actual historical stock prices. Standard data visualization tools will be used to display the Portfolio Results graphically against the major stock market averages and against other scenarios over a specified period of time.