Data Driven Methods in Finance:

Stock Screening & Ranking

Fall 2023: IEOR 4576

Naftali Cohen

Understanding Stock Screening & Ranking

Stock screening and ranking are essential tools used by investors and portfolio managers to make informed investment decisions.

Definition:

- Stock Screening: Filtering stocks based on specific criteria to identify potential investment opportunities.
- Stock Ranking: Assigning ranks to stocks based on their performance or other factors.

Sequential Stock Screening

What is Sequential Stock Screening?

- It involves selecting stocks based on various factors related to the securities.
- Stocks are ranked by a specific factor, and the top-ranked ones are chosen for investment.
- This process is repeated for each factor, narrowing down the pool until a desired number of stocks are selected.

Example: If an analyst believes that stocks with a low Price-to-Book (P/B) ratio yield higher returns:

- Rank all stocks based on their P/B ratios.
- Select those with the lowest ratios. For a 60-stock portfolio, this would mean picking the lowest 60.

Sequential Stock Screening

Managerial Screens Not Tied to Stock "Attractiveness":

Exchange Listing:

- Some portfolio managers might prefer stocks only from specific exchanges, such as NYSE, AMEX, or NASDAQ.
- NASDAQ can be further segmented into the Nasdaq National Market and Nasdaq Small-Cap.

Benchmark Association:

- Managers might opt for stocks only from specific benchmarks like the S&P 500 or Russell 1000/2000.
- Screens can also be set to exclude ADRs or only include stocks from indices such as the DJIA, DJU, or DJT.

* American Depositary Receipts (ADRs) offer US investors a means to gain investment exposure to non-US stocks without the complexities of dealing in foreign stock markets.

The Stock Ranking Process

- Ranking involves evaluating stocks based on specific attributes or metrics.
- If higher values of an attribute are deemed positive, stocks with these high values are selected for a portfolio.

An Effective Stock Screen Should:

Be Easy to Automate & Replicate:

- Facilitates consistent results over time.
- Enables scalability for large datasets.

Reflect Essential Beliefs:

- Align with the investment philosophy or strategy.
- Ensure genuine representation of desired attributes.
- Reflect the correct order of importance.

The Stock Ranking Process: practical example

Scenario: A Portfolio Manager (PM) believes high profit margins and high Book-to-Price (B/P) ratios are critical for a portfolio.

- Universe: S&P 500
- Date: December 2020

Screening Process:

- Step 1: Rank all stocks by profit margin and select the top 30%.
- Step 2: Of the remaining stocks, rank by B/P and select the top 30%.

Ticker	Book-to-Price Ratio (B/P)	Net Profit Margin (NPM)
PBCT	1.38	21.64
BK	1.07	24.10
FITB	1.06	18.21
AFL	1.04	20.87
KEY	0.99	16.61
CMA	0.96	16.17
TFC	0.96	17.01
BAC	0.94	19.50
MTB	0.91	21.49
STT	0.90	19.63

The Stock Ranking Process: practical example

The Dilemma:

Should stocks be sorted using the Price-to-Book (P/B) or the Book-to-Price (B/P) ratio?

Factor Importance:

Choosing the correct version of a factor is *pivotal* when analyzing stocks.

P/B & P/E Ratios: Commonly used but may have limitations.

- Cannot effectively account for companies with negative earnings.

E/P & B/P Ratios: More versatile and recommended.

- E/P can be compared across a broader range of companies.
- B/P offers an insightful inverse perspective of P/B, making it preferable.

Lakonishok's Value Screening Principles

- Originator: Josef Lakonishok - Professor at the University of Illinois; CEO & founder of LSV Asset Management.
- Primary Concept: Identifying companies undervalued due to pessimistic market forecasts.
- Key Attributes of Screen:
  - Equity Selection: Focus on top 30% market capitalization of NYSE, AMEX, and NASDAQ.
  - Metrics for Value: P/E and P/B ratios set below industry medians.
  - Earnings Forecast: Next year's consensus estimate should exceed the current year's forecast.
  - Performance Check: Stock returns should outpace S&P 500 in recent months, signaling potential rebound.

Simultaneous Screening

Sequential vs. Simultaneous:

- Sequential screening filters stocks step by step.
- Simultaneous screening evaluates stocks using all factors at once.

Advantages of Simultaneous Screening:

- No prioritization of factors required.
- Avoids prematurely eliminating potential stocks.
- Retains a larger pool for diversified portfolio selection.
-

Simultaneous Screening

Challenge: Different factors use varied units – direct aggregation isn't feasible.

Solution:

- Utilize the Z-score method to standardize or normalize factors.
- Enables a balanced multi-factor assessment without bias.

Example:

Using B/P ratio & stock size, direct summation is not insightful. Standardization makes these factors comparable.

z-score

Standardize values to make clear statements about an observation's distance from the population mean.

Example: If the average B/P ratio for S&P 500 stocks was 0.3577 with a standard deviation of 0.3794, a company with B/P ratio of 1.1165 is 2 standard deviations away from the mean.

Comparative Analysis: Allows comparing different factors in standardized units. For example, a company with a B/P ratio Z-score of 2 and profit growth Z-score of 2 is equally desirable from both factors.

Probability & Normalization: If the original distribution is normal, Z-scores can indicate probability. A Z-score of 2 indicates <2.27% chance of finding a stock with a higher value.

Assigning Z-Scores:

1. Ensure Z-scores represent the desired qualities (e.g., high B/P ratio is good, use B/P not P/B).
2. Compute the mean of the factor for all stocks.
3. Calculate the standard deviation for all stocks.
4. Z-score formula for a stock:

$$z_{i,k} = \frac{\beta_{i,k} - \bar{\beta}_k}{S(\beta_k)}$$

Aggregate z-score: Combining Multiple Factors

When screening for multiple factors simultaneously, these factors are combined into a single screening value per stock. This is the stock's aggregate Z-score.

Calculation:
After determining individual Z-scores for each factor for each stock, simply add them to find the aggregate Z-score.

$$\bar{z}_i = \left(\frac{1}{K}\right)\left(z_{i,1} + z_{i,2} + z_{i,3} + \cdots + z_{i,K}\right)$$

Handling Outliers:
1. Consider capping extreme Z-scores: All above 3 to 3 and all below -3 to -3.
2. Alternatively, exclude stocks with extreme Z-scores, though this might lead to missed opportunities.

Weighting Factors:
1. Equal Weighting: Most common method, seen as stable and straightforward.
2. Other methods exist, but equal weighting often preferred for its stability.

Ad Hoc Aggregate z-score

Custom Weights by Portfolio Manager:

- Factors can be weighted based on the manager's beliefs or investment style.
- Example: 80% weight to the value factor if deemed most important, with 20% spread among other factors.

Concerns:

- Risk of "data snooping": undue influence from past research.
- Not recommended due to lack of quantitative backing.

Information Ratio Weighting:

- Factors are weighted based on their historical information ratios.
- Higher information ratio indicates better predictive power for stock returns.
- More merit than other ad hoc methods but neglects correlation between factors.

Optimal Aggregate z-score

Equal vs Ad Hoc Weighting:

- Equal weighting is simple but might miss out on valuable data insights.
- Ad hoc weighting is subjective and may not capture correlations between factors.

Optimal Weighting:

- analyze monthly return data in tandem with factor Z-scores, conducting cross-sectional regressions to determine ideal weights.

$$r_{i,t} = \gamma_i + \delta_1 z_{i,1,t-1} + \delta_2 z_{i,2,t-1} + \cdots + \delta_K z_{i,K,t-1} + \epsilon_{i,t}$$

Example:

Selected Factor Exposures and Z-Scores of Selected Stocks

Ticker	Earnings-to-Price Ratio (E/P)	Earnings-to-Book Ratio (B/P)	Debt-to-Equity Ratio (D/E)	Log of Market Capitalization (SIZE)	Momemtum (M12M)	Z-Score of E/P	Z-Score of B/P	Z-Score of D/E	Z-Score of SIZE	Z-Score of M12M	Aggregate Z-Score
AAPL	0.0254	0.0290	3.96	14.6291	82.31	0.117	−0.866	−0.056	3.962	1.431	−0.645
MSFT	0.0282	0.0734	1.44	14.3353	42.53	0.136	−0.749	−0.224	3.688	0.582	−0.699
AMZN	0.0106	0.0507	2.41	14.3066	76.26	0.016	−0.809	−0.159	3.662	1.302	−0.599
GOOGL	0.0323	0.1927	0.41	13.9152	31.03	0.164	−0.435	−0.292	3.297	0.337	−0.588
TSLA	0.0008	0.0240	1.67	13.4134	743.44	−0.051	−0.880	−0.208	2.830	15.535	2.397
FB	0.0385	0.1793	0.24	13.3949	33.09	0.206	−0.470	−0.303	2.813	0.381	−0.479
BRK.B	0.0659	0.7637	0.98	13.2060	2.37	0.393	1.070	−0.254	2.637	−0.274	−0.239
JNJ	0.0410	0.1556	1.65	12.9344	10.85	0.223	−0.533	−0.210	2.384	−0.094	−0.516
WMT	0.0439	0.1847	1.92	12.9186	23.34	0.242	−0.456	−0.191	2.369	0.173	−0.444
JPM	0.0659	0.6223	10.97	12.8670	−5.52	0.392	0.697	0.410	2.321	−0.443	−0.417
S&P 500											
Mean	0.0083	0.3577	4.80	10.3737	15.24	0	0	0	0	0	0
SD	0.1468	0.3794	15.04	1.0741	46.88	1	1	1	1	1	

Note: The Market capitalization is in millions of dollars. Momentum is defined as the monthly cumulative return (in percentage terms) over the past 12 months. The stocks were selected from the S&P 500. Ticker symbols are for the following companies: Apple Inc. (AAPL), Microsoft Corporation (MSFT), Amazon (AMZN) Alphabet Inc. or Google (GOOGL), Tesla Inc. (TSLA), Facebook Inc. (FB), Berkshire Hathaway B shares (BRK.B), Johnson & Johnson (JNJ), Walmart Inc. (WMT), and J.P. Morgan Chase & Co. (JPM). All values were computed as of December 2020. The data were obtained from Compustat.

Factor Groups:

A Possible Categorization of Factors into Composite Groups

Group Number	Factor Group	Factors
1	Valuation composite	Earnings-to-price ratio (E/P) Book-to-price ratio (B/P) Sales-to-price ratio (S/P)
2	Profitability composite	Gross profit margin (GPM)
3	Financial-soundness composite	Inverse interest coverage ratio (IICR) Debt-to-equity ratio (D/E)
4	Technical composite	12-month momentum (M12M)

z-score by Factor Groups:

Ticker	Valuation Composite			Profitability Composite	Financial Soundness Composite		Technical Composite	Valuation Composite Z-Score	Profitability Composite Z-Score	Financial Composite Z-Score	Technical Composite Z-Score	Aggregate Z-Score
	E/P	B/P	S/P	GPM	IICR	D/E	M12M					
AAPL	0.0254	0.0290	0.1217	42.26	0.0428	3.96	82.31	−0.426	−0.038	0.148	1.431	0.279
MSFT	0.0282	0.0734	0.0875	76.75	0.0450	1.44	42.53	−0.393	1.273	0.231	0.582	0.423
AMZN	0.0106	0.0507	0.2129	47.06	0.0825	2.41	76.26	−0.408	0.144	0.188	1.302	0.307
GOOGL	0.0323	0.1927	0.1554	61.31	0.0079	0.41	31.03	−0.254	0.686	0.276	0.337	0.261
TSLA	0.0008	0.0240	0.0421	29.89	0.7534	1.67	743.44	−0.515	−0.508	0.016	15.535	3.632
FB	0.0385	0.1793	0.1203	89.23	N/A	0.24	33.09	−0.265	1.747	0.152	0.381	0.504
BRK.B	0.0659	0.7637	0.5136	25.73	0.0853	0.98	2.37	0.452	−0.666	0.234	−0.274	−0.064
JNJ	0.0410	0.1556	0.1952	74.65	0.0104	1.65	10.85	−0.253	1.193	0.234	−0.094	0.270
WMT	0.0439	0.1847	1.3290	26.71	0.1052	1.92	23.34	0.185	−0.629	0.197	0.173	−0.018
JPM	0.0659	0.6223	0.3437	73.54	N/A	10.97	−5.52	0.267	1.151	−0.205	−0.443	0.192
S&P 500												
Mean	0.0083	0.3577	0.6131	43.26	0.4523	4.80	15.24					
SD	0.1468	0.3794	0.9311	26.31	1.7088	15.04	46.88					

Note: For factor and ticker symbols, see Table 5.2. All values are as of December 2020. The data were obtained from Compustat. Some companies, such as J.P. Morgan and Facebook, had a missing value for interest expense for December 2020. We thus gave them a "neutral" Z-score of zero for this particular factor.

Outliers:

Outliers are extreme observations that diverge significantly from the rest of the data. They may result from data errors, measurement inaccuracies, or genuinely represent unique occurrences.

Challenges:

- Determining if an outlier is a data error or a novel phenomenon.
- Deciding whether to modify, keep, or remove them.

Formal Approach:

1. Null Hypothesis: All observations are from distribution A, meaning no outliers.
2. Alternative Hypothesis: Most data from distribution A, but some from distribution B (outliers).

Outliers:

Detection:

- Common techniques are based on normal distribution assumptions.
- Examples: Grubb's test, Dean-Dixon's test, and Rosner's test.
- Non-normal approaches also exist.

Handling Outliers:

1. Labeling: Identifying potential outliers.
2. Action:
   - Remove observations (trimming).
   - Replace with less extreme values (winsorization).
   - Employ creative alternatives.

Modified z-score

Standard Z-Score Labeling:

- Outliers labeled as stocks with Z-score > 3.
- Drawback: Not reliable if extreme values influence mean and standard deviation calculations.

Modified Z-Score:

- Uses median instead of mean. MAD (median-[X-median(X)]) instead of standard deviation
- Outliers: Observations with absolute modified Z-score ≥ 3.

Approaches to Handle Outliers:

- Winsorization: Replaces outliers with a predetermined value (e.g., >3 with 3 and <−3 with −3).
- Trimming:
  - Removes values above and below 6
  - Removes top N/2 and bottom N/2 extreme Z-score values and recomputes the Z-scores.

Handling Outliers with Interquartile Range

IQR ensures that the data is less skewed and more reflective of the central tendency.

Procedure:

1. Compute Interquartile Range (IQR):
   - IQR = Q3 (third quartile) - Q1 (first quartile)
   - Represents middle 50% of factor values across stock factors.
2. Determine Outliers:
   - Using the IQR coefficient, decide what deviation constitutes an outlier.
   - Example: If IQR coefficient = 3:
     - Upper Bound (UB) = Q3 + 3*IQR
     - Lower Bound (LB) = Q1 – 3*IQR
3. Identify Outliers:
   - Stocks with factor values > UB or < LB.

Treatment for Outliers:

1. Set outlier values to missing and compute Z-scores for remaining stocks.
3. Fix outlier Z-scores at max and min of the non-outlier stocks' Z-scores.

The Ranking method

An alternative to handle outliers in stock factor analysis.

- Procedure:
  1. Rank Stocks: Every evaluation period, rank stocks based on their factor value from highest to lowest.
  2. Compute Ranking Z-scores: Instead of traditional Z-scores, compute Z-scores based on the actual rank of stocks.

- Benefits:
  - Addresses the outlier issue by focusing on relative ranks.

- Trade-off:
  - Loses the information on the absolute distance between stocks' factor values. Only retains the relative Z-score values between stocks.

- Ranking Z-score Formula:

$$Z_{it}^{\text{rank}} = \frac{\text{rank}(f_{it}) - \frac{\sum_{i=1}^{N} \text{rank}(f_{it})}{N}}{\sigma_{\text{rank}(f_{it})}}$$

Note: "rank" refers to the cross-sectional numerical rank of each stock from high (better attribute) to low.

The Percentile Ranking method

An alternative way to handle outliers, building upon the ranking method. This method emphasizes the relative positioning of stocks within the universe based on percentiles rather than absolute values.

- Procedure:
  1. Rank Stocks: Rank all N stocks based on their factor value from highest to lowest.
  2. Compute Percentile Z-scores:

$$Z_{it}^{\text{per}} = \frac{N - \text{rank}(f_{it})}{N}$$

Disclaimer

This course is for educational purposes only and does not offer investment advice or pre-packaged trading algorithms. The views expressed herein are not representative of any affiliated organizations or agencies. The main objective is to explore the specific challenges that arise when applying Data Science and Machine Learning techniques to financial data. Such challenges include, but are not limited to, issues like short historical data, non-stationarity, regime changes, and low signal-to-noise ratios, all of which contribute to the difficulty in achieving consistently robust results. The topics covered aim to provide a framework for making more informed investment decisions through a systematic and scientifically-grounded approach.