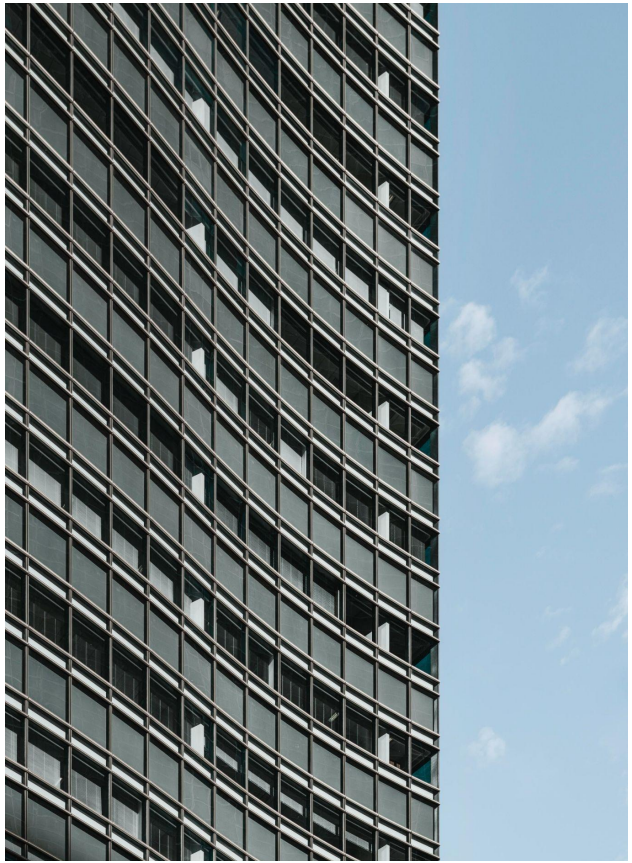
CMPT 733 Spring 2024 - Milestone Presentation

Integrating AI and Quantitative Analysis for Equity Investment and Portfolio Optimization



Group: HappyCNY

Shung Ho (Jonathan) Au
Hongying Yue
Sitong Zhai
Qin Duan

AGENDA

01 Motivation

02 Progress Report & Schedule

03 Future work

Introduction

Capital Markets:

- Buying and selling of **securities** (stocks, bonds etc)
- Mechanism for price **valuation** (Demand and Supply)
- The goal is to **outperform** the market baseline (eg: S&P500, other indexes etc)

The Truth:

- The Majority falls behind indexes in long run (individuals, professionals etc)
- Price movement affected by many forces (sentiment, news etc)
- Deep Dive Analysis is slow and time consuming

Quantitative Finance:

- A rising field to incorporate machine learning into investing decision
- Selecting securities based on historical data and trained models
- Targets specific drivers of return to create portfolios (**Factor Investing**)

The New York Times

Actively Managed Mutual Funds Consistently Fail to Beat Markets, Study Finds

It's very hard to beat the stock or bond markets with any regularity. Each year, some investors manage to do it, of course, but can they do...

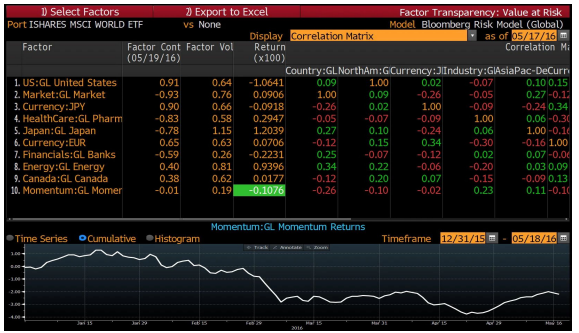
Dec 2, 2022

CNBC

New report finds almost 80% of active fund managers are falling behind the major indexes

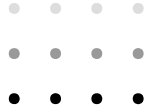
A majority of active mutual fund managers are underperforming the S&P 500 and the Dow, according to the newest S&P Indices versus Active...

Mar 27, 2022



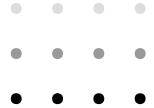
Bloomberg Markets Magazine

Key Motivations



1. **Establish a rational and consistent Investing Framework**
    a. Systematically process and **identify high return stocks**
    b. Not predicting price but a **classification** task (performance is relative to its peers)
    c. Eliminate emotional Bias
2. **Enhance Return Potential through Machine Learning**
    a. Construct **features** for the predictive model (multi-factors)
    b. Assess model effectiveness through **simulating** portfolio performance (experiments)
    c. Examine the usefulness of features (Insights, model interpretability)

Key Motivations



1. **Establish a rational and consistent Investing Framework**
    a. Systematically process and **identify high-return stocks**
    b. Not predicting price but a **classification** task (performance is relative to its peers)
    c. Eliminate emotional Bias

2. **Enhance Return Potential through Machine Learning**
    a. Construct **features** for the predictive model (multi-factors)
    b. Assess model effectiveness through **simulating** portfolio performance (experiments)
    c. Examine the usefulness of features (Insights, model interpretability)

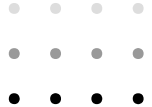
Challenges:

1. **Feature Selection and Engineering**
    a. Require **advanced knowledge in accounting & finance** to construct high-quality features
    b. Complex integration of trading and financial data
    c. Handling of missing financial data, gaps, outliers etc

2. **Data Quality, Granularity and Frequency**
    a. **Noisy** price movements may obscure trends
    b. Handling of stocks with **abnormal status** any point in time (suspension, delisting, severe distress)
    c. Aggregate into meaningful time windows to reveal time patterns (some information is lost)

3. **Experiment Design**
    a. Different **Time Horizon**, **labelling** techniques, **portfolio reconstruction interval**, may have different results

Progress Report



Completed Tasks

Stage 1: Research and Planning:

a. Conducted research about the possible ways to represent each stock from multiple perspectives
b. Identified 5 main categories of usable features, each with 3-4 metrics

Research Evidence:

Valuation factors: pe_ratio, pb_ratio, ps_ratio, pcf_ratio etc (stock **price to earnings** ratio is a relative measure of "**expensiveness**")

Growth factors: %_growth_in_net_profit, %_growth_in_total_revenue, %_growth_in_operating_profit (growing or declining **profits over time**)

Profitability factors: gross_profit_margin, net_profit_margin, cash_flow_to_profit, adjusted_profit (**profitability** of a business compared to **cost**)

Leverage factors: financial_leverage_ratio, debt_to_equity_ratio, current_ratio, cash_ratio (**debt and borrowings** composition etc)

Momentum Indicators: relative_strength_index (rsi), bias, psy, macd, hsl (Quantify **stock price movement trends**, shifts and momentum)

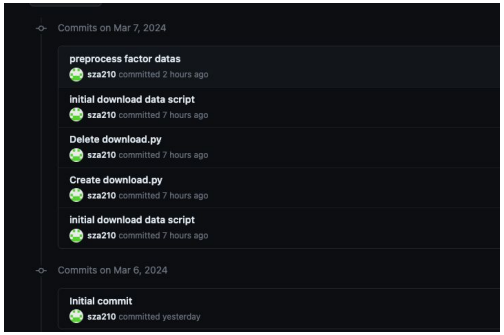
Progress Report

Completed Tasks

Stage 2: Data Collection

a. **Tools:** JQ Data API to gain access to stock and financial data
b. **Time Interval Covered:** 2016 - 2023 (Trading days)
c. **Stock Universe:** CSI500 China's mid and small-cap universe
d. **Data gathered**: Price data and scattered pieces of accounting information (for feature engineering)

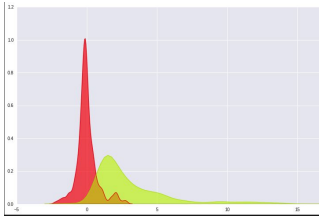
Evidence:



commits

```python
def preprocess_factors(trade_days_all):
    factors_all = ['pe_ratio', 'pb_ratio', 'ps_ratio', 'pcf_ratio',
                   'financial_leverage_ratio', 'debte_to_quity_ratio', 'cash_ratio', 'current_ratio',
                   'gross_profit_margin', 'net_profit_margin', 'adjusted_profit_to_profit', 'ocf_to_operating_profit',
                   'inc_total_revenue_year_on_year', 'inc_operation_profit_year_on_year', 'inc_net_profit_year_on_year',
                   'RSI', 'BIAS', 'PSY', 'DIF', 'DEA', 'MACD', 'ATR14', 'HSL',
                   'circulating_market_cap']
    for trade_day in trade_days_all:
        neutra_factor_datas = pd.DataFrame()
        for factor in factors_all:
            data = compute_win_stand_neutra(f'{factor}', trade_days_all[0])
            neutra_factor_datas[f'{factor}'] = data
        neutra_factor_datas.to_csv(f'neutra_{index}/{trade_day}.csv', index=True)
```

Code excerpt related to features and preprocessing



Distribution before vs after preprocessing

Project Schedule

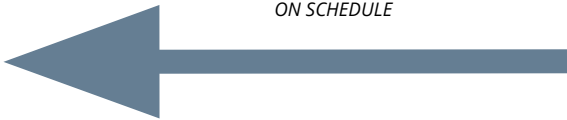
02/01 - 02/20: Research and Planning

- Studied and identified financial metrics for feature engineering, identified data source, stock list
- Planned for project scope, data granularity, procedures for model training and performance simulation

02/21 - 03/03: Data Collection

- Thoroughly **tested** the **JQData API**, identified limitations in data volume and breadth of information
- Finalized and **successfully acquired** the necessary data for further processing

ON SCHEDULE



03/04 - 03/10: Data Cleaning and Standardization, Experiment setup

- Implemented a script to address outliers and standardize the scale of various metrics
- Enriched our dataset with industry classification data for each stock
- **Feature bias reduction** and **normalization** to adjust for **industry influence** to enhance compatibility across stocks

03/11 - 03/25: Run Experiments, Compare with baseline, Gather Insights

03/25 - 04/03: Prepare final report, posters

Future Work

03/11 - 03/25: Run Experiments, Compare with baseline, Gather Insights

Examine Model Attributes:
- Mutual Information
- Feature Importance
- AUC over-time to inspect classification accuracy of “good stocks”

Examine Portfolio Performance:
- Rolling-forward testing methodology (train on 2016-2018 data, test on 2019-2023)
- Simulate “buying good stocks” on test months based on model classification, then compute for portfolio value change over-time
- Evaluate accumulated return % , maximum value decrease %, compared to CSI 500 index benchmark

Mitigate Risk:
- Outperforming the market is always hard, outperforming the index is not the only objective
- To evaluate effectiveness of a model, there can **additional benchmark** such as return comparison between “buying good stocks” and “buying bad stocks” based on model output, and inspect differences

03/25 - 04/03: Prepare final report, posters

Mitigate Risk: Start writing as we test.

Thank you