

## 0.1 Data collection

The dataset, **sourced from Bloomberg at a daily frequency**, integrates multiple dimensions of financial information, including price dynamics, risk factors, volatility, and valuation measures, all of which are essential for stock return forecasting. Together, these variables enable a comprehensive evaluation of predictive models across both statistical and machine learning approaches.

### 0.1.1 OHLCV

The OHLCV variables provide the fundamental basis of price and volume movements.

- Open: The price at which the asset opened during the trading session.
- High: The highest price reached during the session.
- Low: The lowest price recorded during the session.
- Close: The final price at the end of the trading session (referred to as `PX_LAST` in this paper, following Bloomberg's convention).
- Volume: The total number of shares/contracts traded during the session.

### 0.1.2 Technical Indicators

Technical indicators capture short-term trading behavior and reflect underlying market microstructure effects.

- Returns: Daily returns, often calculated as the percentage change from the previous close.
- Vwap: Volume Weighted Average Price, calculated as the ratio of traded value to volume.
- Ask Price: Current best asking price (sell quote).
- Bid Price: Current best bidding price (buy quote).
- Spread: Difference between ask and bid prices, indicating liquidity and transaction cost.

### 0.1.3 Volatility measures

Volatility measures across different horizons capture both forward-looking expectations and historical uncertainty, both of which are central to understanding financial market behavior.

- Vix: Implied volatility index, often referred to as the "fear index."
- Historical Call Implied Volatility: Implied volatility from call options, reflecting market expectations of future price fluctuations.

- Volatility 30 days: Realized/historical volatility over the last 30 days.
- Volatility 60 days: Realized/historical volatility over the last 60 days.
- Volatility 90 days: Realized/historical volatility over the last 90 days.

#### **0.1.4 Fundamentals Data**

Fundamental ratios provide valuation perspectives that can influence stock price movements over longer horizons.

- Price Earnings Ratio (P/E): Ratio of the company's current share price to its earnings per share.
- Best P/E Ratio: Bloomberg estimate of the future P/E ratio based on forecasted earnings.
- Price/Cash Flow: Ratio of share price to cash flow per share, indicating valuation based on cash generation.

#### **0.1.5 Fama-French and Related Factors**

The Fama-French and related factors serve as well-established proxies for systematic risk, thereby situating the dataset within the broader framework of asset pricing theory.

- Mkt-RF: Market risk premium (market return minus the risk-free rate).
- SMB: Small Minus Big — the return difference between small-cap and large-cap stocks.
- HML: High Minus Low — the return spread between value and growth stocks.
- RMW: Robust Minus Weak — profitability factor in Fama-French five-factor model.
- CMA: Conservative Minus Aggressive — investment style factor (low vs. high asset growth).
- RF: Risk-free rate, often proxied by short-term government bonds.