



# Applying Functional Data Analysis to the Price Action of the Most Traded Stocks in the S&P 500 Index

Davide Giuseppe Griffon, Aurimas Bžėskis, Kasparas Rutkauskas

Vilnius University

18 march 2025

## 1 Introduction

## 2 Data and preprocessing

- Data set
- Data preprocessing
- Data clustering

## 3 Data smoothing

- Smoothing approach using B-Splines
- Smoothed data visualisation

## 4 FPCA of smoothed data

# Table of contents

## 1 Introduction

## 2 Data and preprocessing

- Data set
- Data preprocessing
- Data clustering

## 3 Data smoothing

- Smoothing approach using B-Splines
- Smoothed data visualisation

## 4 FPCA of smoothed data

- **Motivation.** Retail investors' share of U.S. equity trading has nearly doubled since 2010, reaching close to 20% by early 2023. However, **only 10% of traders achieve positive returns**, with even fewer remaining profitable long-term. Many **rely on intuition rather than data-driven strategies**, fueling debate over market predictability. While markets are largely efficient, intraday patterns and momentum may still offer predictive value despite the common disclaimer that "**past performance is not indicative of future results.**"
- **Research goal (objectives):**
  - ① to investigate whether intraday stock price movements exhibit persistent patterns that differentiate bull and bear market days;
  - ② to assess whether past intraday trends contain predictive information for future price action;
  - ③ to test which stock market sectors exhibit the highest degree of predictability.

# Table of contents

1 Introduction

2 Data and preprocessing

- Data set
- Data preprocessing
- Data clustering

3 Data smoothing

- Smoothing approach using B-Splines
- Smoothed data visualisation

4 FPCA of smoothed data

# Data set

- **Data description.** The SP 500 was selected as it represents the 500 largest U.S. companies, covering approximately 80% of the U.S. stock market and 50-60% of global market capitalization. As of 2025, it remains the largest and most widely followed index. This study focuses on the 100 most actively traded S&P 500 stocks.
- **Data source.** To ensure high-quality data, we sourced it from Polygon, accessing their database via a Python script with an authorized API key.
- **Timeframe.** For this study, we analyzed morning trading activity (9:30 AM–12:00 PM) from 2023–2024, yielding 10 data points per day-ticker. After restructuring, the dataset comprised 50,116 stock-day entries, each represented by 15-minute intervals.
- **Motives.** The first hours of the trading day were selected due to their high liquidity, volatility, and institutional activity, making them the most actively traded period.

# Stock price visualisation



(a) Apple 15-minute timeframe candlestick chart displaying price and volume for a single trading day



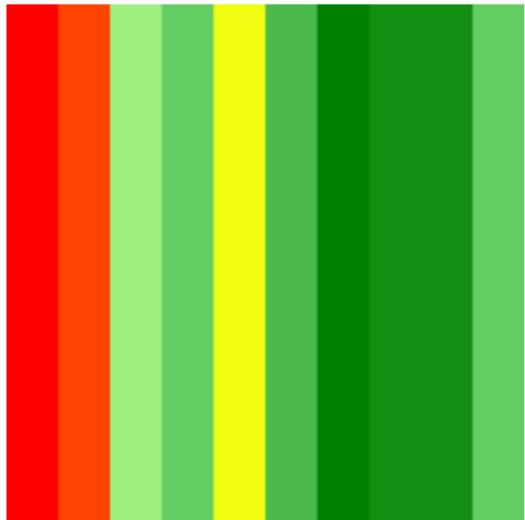
(b) Apple 15-minute timeframe line chart showing closing price and volume for a single trading day

Figure 1: Apple stock price charts - 1 of 100 stocks used in the study  
Source: Tradingview

# Data preprocessing

- **Data Normalization.** Normalized stock prices (0-1 scale) for cross-stock comparison: 0 → Lowest morning closing price and 1 → Highest morning closing price.
- **Visual Representation.** Each day-ticker was represented by a 10-strip color-coded image, where red indicated the lowest price (0), green represented the highest price (1), and gradient colors depicted intermediate values.
- **Clustering.** Initially, k-means clustering ( $k=2$ ) was used to identify bull and bear trends, but it resulted in "dirty" clusters, and further k-value adjustments did not yield satisfactory improvements. To enhance accuracy, a ResNet34 deep learning model was trained on five manually labeled classes, achieving 90% accuracy, and was applied to 50,116 day-tickers, with a confidence threshold of 0.8 ensuring high-confidence predictions were assigned to clusters while lower-confidence ones were placed in an "Unknown" category.
- **Results.** A total of 26,895 day-tickers were excluded as unstructured data, leaving four viable categories for analysis: Bear (5,838 day-tickers), Bull (7,230 day-tickers), Bull-Center (6,110 day-tickers), and Bear-Center (4,049 day-tickers).

# Bull & Bear classes



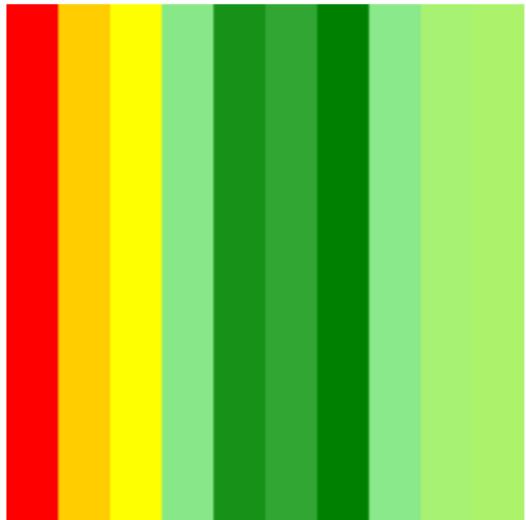
(a) **Bull day.** Day-tickers exhibiting a clear bullish morning trend, with prices generally rising throughout the morning session.



(b) **Bear day.** Day-tickers showing a distinct bearish morning trend, with prices predominantly falling during the morning session.

**Figure 2:** Classified bull&bear data example  
Source: Made by author

# Bull-Center & Bear-Center classes



(a) **Bull center day.** A price pattern characterized by an initial price increase followed by a downtrend.

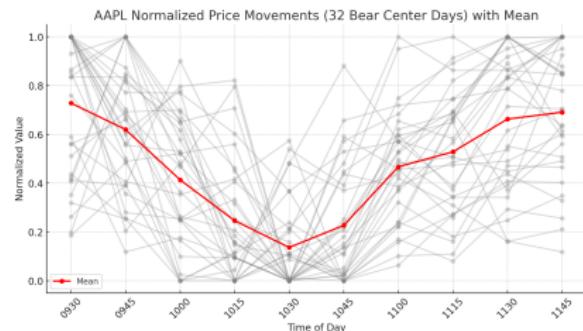
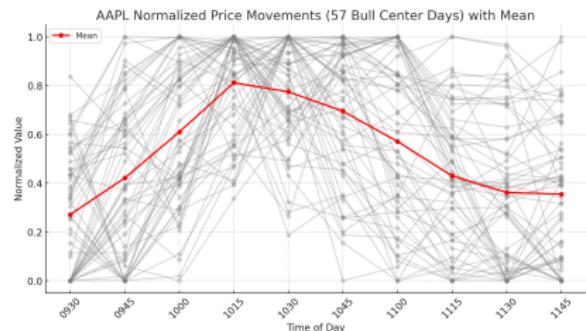
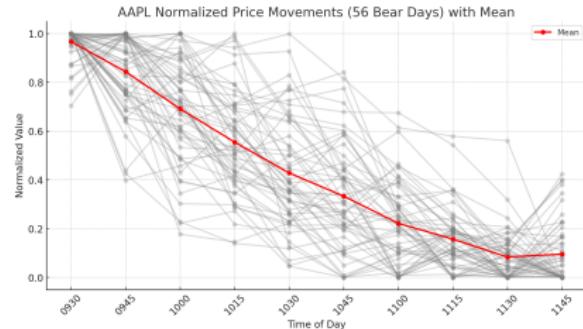
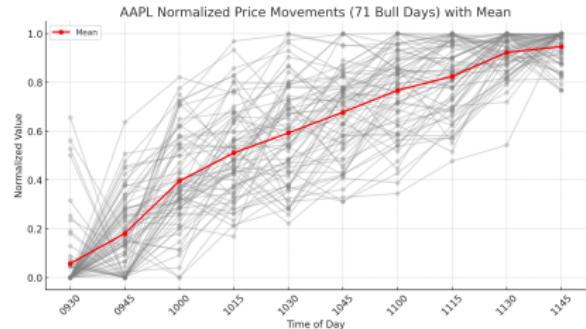


(b) **Bear center day.** A price pattern showing an initial downtrend followed by a price recovery.

Figure 3: Classified bull&bear center data example

Source: Made by author

# Classified data of individual stock



**Figure 4: Apple Stock classified data – 1 of 100 stocks in the study**  
Source: Made by author

# Table of contents

1 Introduction

2 Data and preprocessing

- Data set
- Data preprocessing
- Data clustering

3 Data smoothing

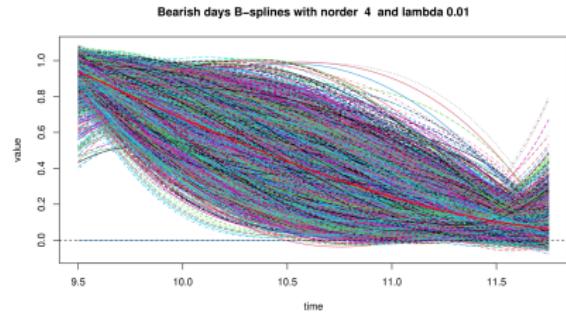
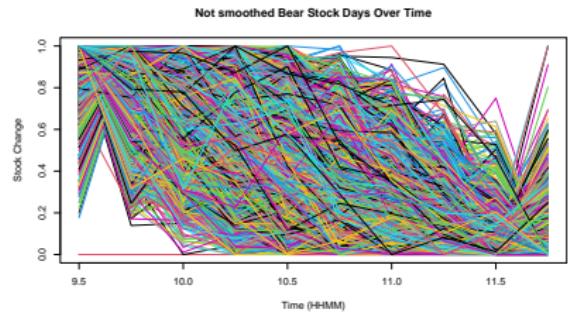
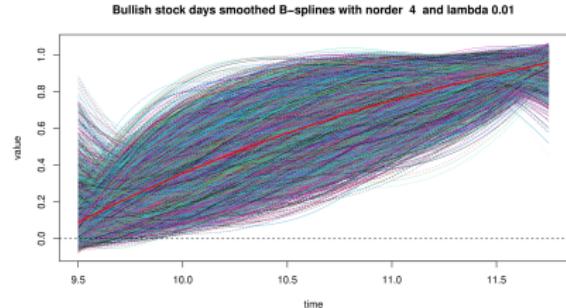
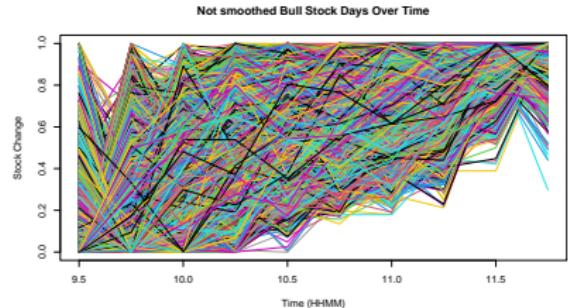
- Smoothing approach using B-Splines
- Smoothed data visualisation

4 FPCA of smoothed data

# Smoothing approach using B-Splines

- **Data Integrity.** Removed all stock-day entries have at least one missing value.
- **B-Spline Setup.** Defined range 9:30 AM – 11:45 AM, norder (4) and nbasis (5) for smooth representation.
- **Lambda Optimization.** Created a sequence of 10 lambda values between  $10^{-4}$  and  $10^{-1}$  (for GCV). Optimized lambda by minimizing the GCV value.
- **Functional Analysis.** Computed Mean Function and plotted smoothed functions to visualise stock-day trends.

# Smoothed data visualisation I



**Figure 5:** Raw vs. B-Spline smoothed stock data for Bull and Bear Days  
Source: Made by author

# Smoothed data visualisation II

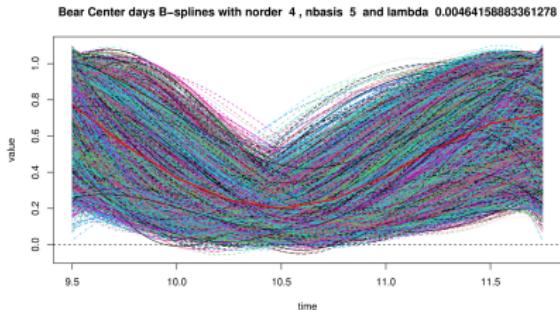
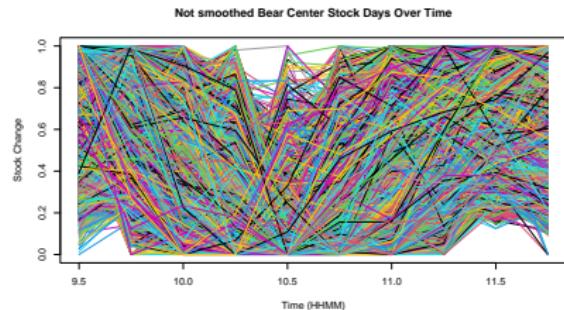
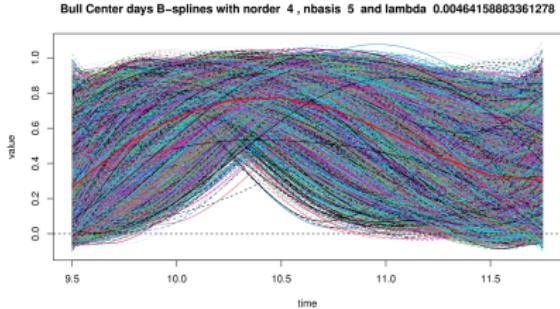
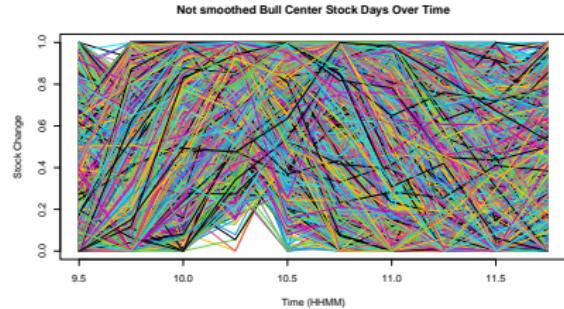


Figure 6: Raw vs. B-Spline smoothed stock data for Bull-Center and Bear-Center Days  
Source: Made by author

# Table of contents

## 1 Introduction

## 2 Data and preprocessing

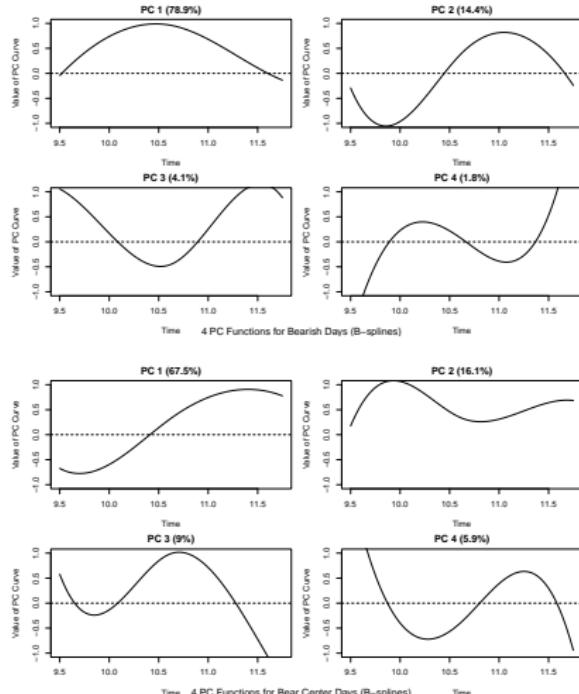
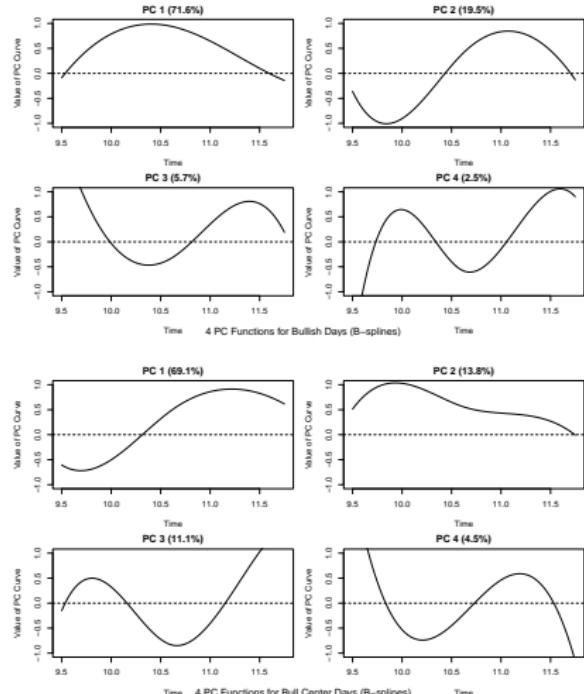
- Data set
- Data preprocessing
- Data clustering

## 3 Data smoothing

- Smoothing approach using B-Splines
- Smoothed data visualisation

## 4 FPCA of smoothed data

# FPCA of smoothed data



**Figure 7: 4 PC functions for Bull, Bear, Bull-center, Bear-center days**  
Source: Made by author

Thank you for your attention