Stock Market Pre-Recession Analysis

Amogh Krishna N. & Shriyans Nellutla

Overview

This project dives into whether stock market behavior shows consistent patterns in the months leading up to major U.S. recessions. The financial data accessed through the `{quantmod}` and `{tidyquant}` packages in R, we explore trends, volatility shifts, and structural changes during key points in time. Our goal is to identify indicators that may signal recessions and to explore predictive insights.

We see two parts -

1. Main Stock Trends(S&P 500)
2. False indicators of a recession.

We have to take into consideration both the above as 1. helps us see trends and predict a stock market crash which is usually due to ONE factor. The problem arises when we see trends like the same in the stock market that caused a crash to not lead to one (ex.2021). Hence, we also need to look at the false indicators so that we can rule out recessions we think will happen but actually wont.

Interesting Insight

One of our insights comes from comparing the normalized S&P 500 trajectories across 2007, 2020, and 2025. Each series was normalized to 100 at its starting point to allow comparison across different crises. This helps find a correlation in whether the current market trends resemble past pre-recession environments.

Data Sources and Acknowledgements

We utilize data from FRED (CPI, Treasury Yields, Fed Funds) and Yahoo Finance (S&P 500). The data meets the **FAIR principles**:

- **Findable:** Available via FRED/Yahoo Finance API.

- **Accessible:** Downloadable using `tq_get`.

- **Interoperable:** Standard CSV/tibble formats.

- **Reusable:** Clearly documented sources(in code)

We acknowledge and credit the developers for enabling us to access financial market data.

Data Wrangling

```r
# Load libraries
library(tidyquant)
library(dplyr)
library(tidyr)
library(ggplot2)

# 1. Federal Funds Rate (Factor 1)
fed <- tq_get("FEDFUNDS", get = "economic.data") %>%
  filter(date >= as.Date("2010-01-01"))

# 2. S&P 500 Historical Data (1990 - Present)
sp_full <- tq_get("^GSPC", from = "1990-01-01", to = Sys.Date()) %>%
  mutate(norm = adjusted / adjusted[1] * 100)

# 3. Inflation Data (CPI)
cpi <- tq_get("CPIAUCSL", get = "economic.data") %>%
  filter(date >= as.Date("1990-01-01")) %>%
  mutate(norm = price / price[1] * 100)

# 4. Yield Curve Spread (Factor 3)
yc <- tq_get(c("DGS10","DGS2"), get = "economic.data") %>%
  spread(symbol, price) %>%
  mutate(spread = DGS10 - DGS2)

# 5. Normalising S&P 500
sp_2007 <- tq_get("^GSPC", from = "2005-01-01", to = "2011-05-01")
sp_2020 <- tq_get("^GSPC", from = "2019-01-01", to = "2021-01-01")
sp_now  <- tq_get("^GSPC", from = "2025-05-01")

# Normalize both to 100 at start for shape comparison
normalize <- function(df) {
  df %>% mutate(norm = adjusted / adjusted[1] * 100)
```

```
}

sp_2007_n <- normalize(sp_2007)
sp_2020_n  <- normalize(sp_2020)
sp_now_n  <- normalize(sp_now)

sp_2007_n <- sp_2007_n %>% mutate(period = "2005-2011")
sp_2020_n <- sp_2020_n %>% mutate(period = "2019-2021")
sp_now_n  <- sp_now_n  %>% mutate(period = "Present")

all_sp <- bind_rows(sp_2007_n, sp_2020_n, sp_now_n)
```
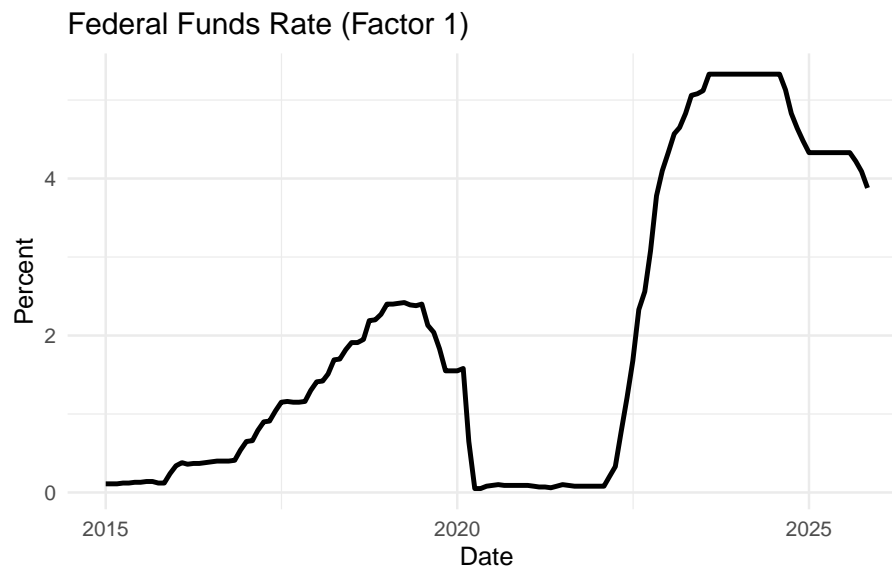
Visualizations

We have not included the 3 plots of S&P 500 in visualizations since that was for our reference. To access those images, they will be in our Final Project Git
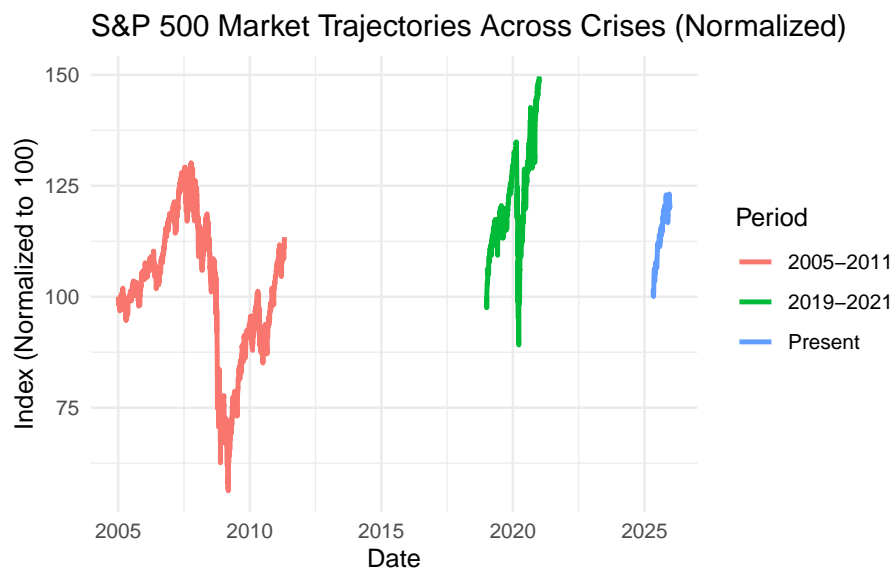
```
# 1. Federal Funds Rate
ggplot(fed, aes(date, price)) +
  geom_line(linewidth=1) +
  labs(title="Federal Funds Rate (Factor 1)", x="Date", y="Percent") +
  theme_minimal()
```
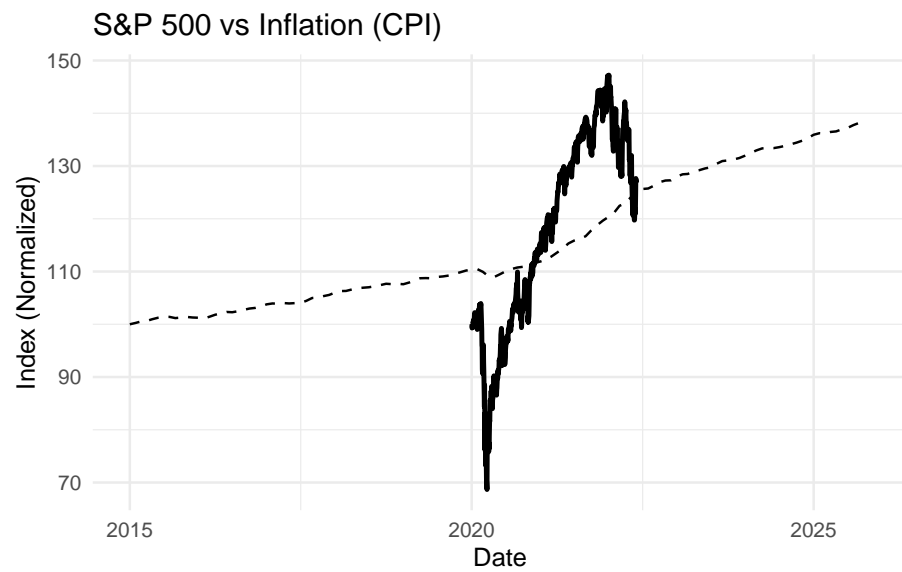

3

```
# 2. Normalized S&P 500 Trends
ggplot(all_sp, aes(x = date, y = norm, color = period)) +
  geom_line(linewidth = 1) +
  labs(
    title = "S&P 500 Market Trajectories Across Crises (Normalized)",
    y = "Index (Normalized to 100)",
    x = "Date",
    color = "Period"
  ) +
  theme_minimal()
```


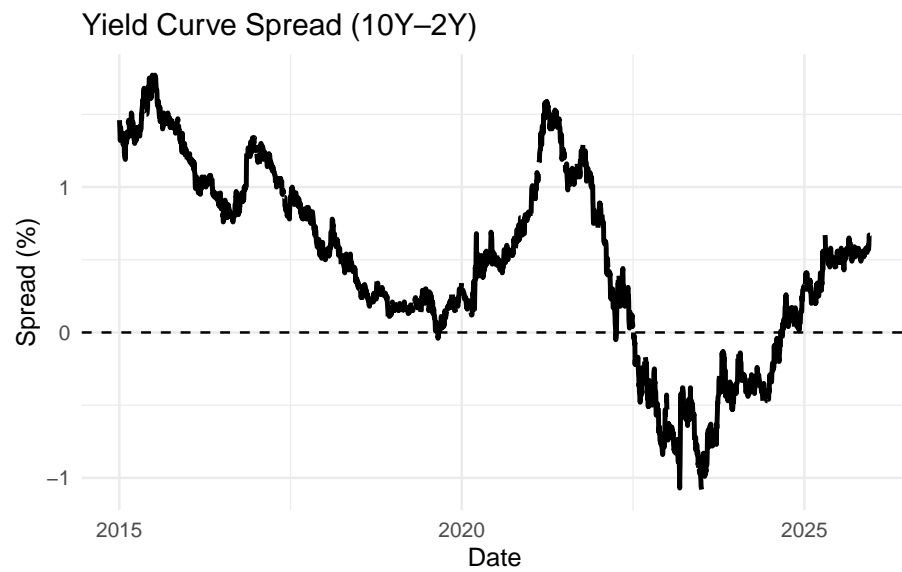S&P 500 Market Trajectories Across Crises (Normalized)

```
# 3. Inflation vs S&P 500
sp <- tq_get("^GSPC", from = "2020-01-01", to = "2022-06-01")
cpi <- tq_get("CPIAUCSL", get = "economic.data")
ggplot() +
  geom_line(data = sp, aes(date, adjusted / adjusted[1] * 100),
            linewidth = 1) +
  geom_line(data = cpi, aes(date, price / price[1] * 100),
            linetype = "dashed") +
  labs(
    title = "S&P 500 vs Inflation (CPI)",
    y = "Index (Normalized)",
    x = "Date"
  ) +
  theme_minimal()
```

S&P 500 vs Inflation (CPI)


```r
# 4. Yield Curve Spread (10Y - 2Y)
ggplot(yc, aes(date, spread)) +
  geom_line(linewidth = 1) +
  geom_hline(yintercept = 0, linetype = "dashed") +
  labs(
    title = "Yield Curve Spread (10Y-2Y)",
    y = "Spread (%)",
    x = "Date"
  ) +

  theme_minimal()
```
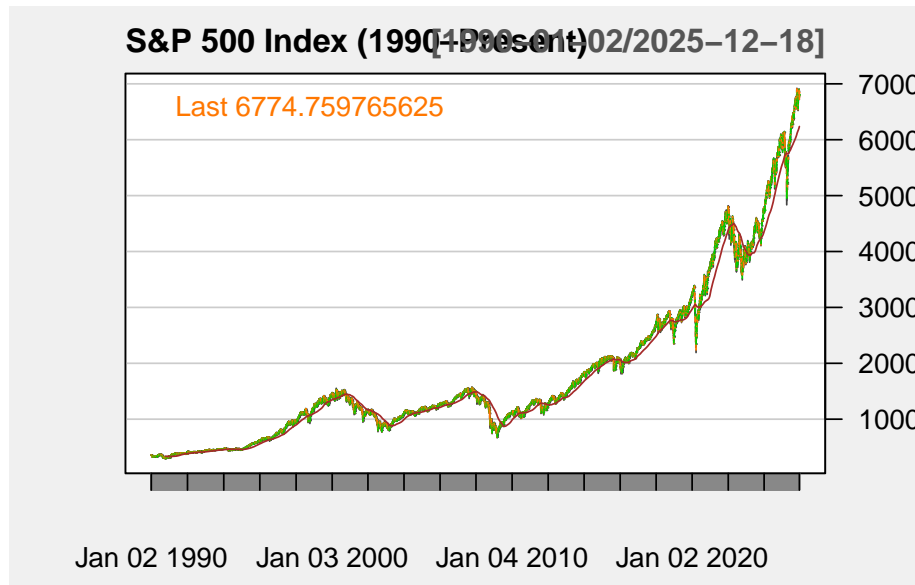
Yield Curve Spread (10Y–2Y)


```
# 5. 1990 to Present S&P500
getSymbols("^GSPC", src = "yahoo", from = "1990-01-01")
```

```
[1] "GSPC"
```

```
chartSeries(GSPC,
            name = "S&P 500 Index (1990-Present)",
            theme = chartTheme("white"),
            TA = "addSMA(200)")
```

S&P 500 Index (1990–Present) [1990-01-02/2025–12–18]

Last 6774.759765625

Jan 02 1990   Jan 03 2000   Jan 04 2010   Jan 02 2020

Analysis and Interpretation

- **Federal Funds Rate:** - The graph shows how the Federal Rates lowered - This is not due to a potential recession but due to stresses in the economic system.
- **Normalized S&P 500:** Provides a clear visualization of market trajectories before major recessions allowing us to show a comparison across multiple crises.
- **S&P 500 vs Inflation:** - This graph shows us that the S&P 500 rise causes the CPI to increase - meaning "INFLATION". This is a major false indicator of a recession. We see similar trends before a recession so we have to be careful to take these into consideration.
- **Yield Curve Spread:** Negative spreads precede recessions, confirming its value as an early indicator. For 2021, it shows that it is not as significant of a negative spread. We can use this as a tool to find the accuracy of the prediction of a recession.

Each visualization is discussed in context, linking patterns to potential recession signals and structural market changes. We see that the coming trend is not a false indication and we also see a very high leading cause for there to be a recession.

Reproducibility and Open Science

- All data processing and analysis steps are included in this QMD.

- Version control is applied via GitHub with multiple commits and PRs.

- Inline comments, consistent coding style, and naming conventions ensure readability and reproducibility.

- FAIR/CARE principles are considered in data sourcing and handling.

- Code can be re-run on other machines to reproduce all results.