Do Google Trends Indicate How the Market will Behave?

Financial Modeling | December 1, 2025

Motivation

Context

Rising inflation and cost-of-living pressures have increased economic uncertainty, creating a need for faster and more responsive indicators of public concern and market conditions.

Problem

Traditional market and economic indicators are published with delays, limiting the ability of investors and policymakers to detect shifts in sentiment or financial risk in real time.

Idea

Google search activity may serve as an immediate behavioral signal, revealing changes in consumer and investor attention to inflation before these shifts appear in official statistics.

Why It Matters

Real-time search data could provide an early forecasting edge for businesses, investors, and policymakers by highlighting emerging economic pressures and sentiment-driven market movements. It also demonstrates the potential value of integrating alternative data sources into financial prediction models.

Overview

Research Question: Do increases in Google searches for inflation-related terms predict changes in the market?

Approach

Run a regression to see whether people googling the word *'inflation' or inflation related words* tells us anything about what the stock market will do next month

Expected Finding

Higher Google search activity is associated with slightly higher future market returns, reflecting increased investor attention

Model Summary

Excess_Returnt_tt = S&P 500 return in month t minus the risk-free rate

SVIt−1_{t-1}t−1 = Lagged Google search index for "inflation" (previous month)

Mkt_RFt_tt = Fama–French market excess return factor

α = intercept

β_1 = effect of Google search interest on future returns

β_2 = effect of market risk factor on returns

εt_tt = error term

$$\text{Excess\_Return}_t = \alpha + \beta_1\,\text{SVI}_{t-1} + \beta_2\,\text{Mkt\_RF}_t + \varepsilon_t$$

Model: Key Assumptions

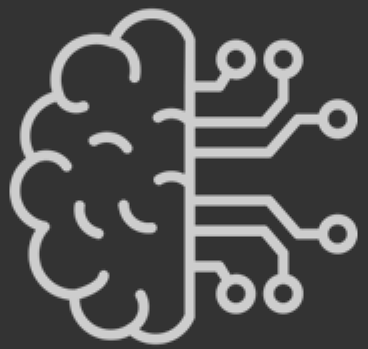
- Google search trends capture public attention before markets react.
  - Spikes in searches (e.g., “inflation”) reflect real-time concern or interest.
- Using a lag structure ($t-1 \rightarrow t$) lets search trends predict what happens next.
  - Last month's searches are compared to this month’s market returns to measure true predictive power.
- There is a testable relationship between SVI and future market performance.
  - Higher search activity may translate into investor attention, which can influence returns.

Model

Setup

- Ken French Market Risk Factor
- S&P 500 Index from FRED
- Google Trends data with key words: inflation, gas prices, rent inflation, grocery prices

Intuition



- People search for information on topics that they are curious or worried about (not random)
- Can add more terms (food, housing, gas prices) to see which words are most predictive

$$\text{Excess\_Return}_t = \alpha + \beta_1 \, \text{SVI}_{t-1} + \beta_2 \, \text{Mkt\_RF}_t + \varepsilon_t$$

Data: General

- Time Frame: 2004 to 2025
    - Google trends didn't exist until 2004

- Frequency: Monthly
- Inputs/Variables: Ken French Market Factor Data, S&P 500 Index
- How we did this...
    - Instead of using excel we used:
    - Python (Data Analysis + Modeling)
    - Pandas (Used for loading datasets)
    - NumPy
    - Matplotlib (visuals)
    - Statsmodels (ols regression)

Data

```
Cleaned Fama-French data:
        date  Mkt_RF      RF
0 1926-07-01  0.0289  0.0022
1 1926-08-01  0.0264  0.0025
2 1926-09-01  0.0038  0.0023
3 1926-10-01 -0.0327  0.0032
4 1926-11-01  0.0254  0.0031
           date  Mkt_RF      RF
1187 2025-06-01  0.0486  0.0034
1188 2025-07-01  0.0198  0.0034
1189 2025-08-01  0.0184  0.0038
1190 2025-09-01  0.0339  0.0033
1191 2025-10-01  0.0195  0.0037
Shape: (1192, 3)
```

```
S&P raw columns: Index(['Date', 'Open', 'High', 'Low', 'Close', 'Volume'], dtype='object')
         Date     Open     High      Low    Close        Volume
0  2004-01-31  1111.92  1155.38  1105.08  1131.13  1.823333e+10
1  2004-02-29  1131.13  1158.98  1124.44  1144.94  1.554756e+10
2  2004-03-31  1144.94  1163.23  1087.16  1126.21  1.866550e+10
3  2004-04-30  1126.21  1150.57  1107.23  1107.30  1.756217e+10
4  2004-05-31  1107.30  1127.74  1076.32  1120.68  1.629244e+10
Using S&P price column: Close

Cleaned S&P:
        date  SP500_price  SP500_ret
0 2004-01-01      1131.13        NaN
1 2004-02-01      1144.94   1.220903
2 2004-03-01      1126.21  -1.635894
3 2004-04-01      1107.30  -1.679083
4 2004-05-01      1120.68   1.208345
```

Data

```
Google Trends columns: Index(['Month', 'gas prices: (United States)', 'inflation: (United States)',
       'rent inflation: (United States)', 'grocery prices: (United States)'],
      dtype='object')
     Month  gas prices: (United States)  inflation: (United States)  \
0  2004-01                            5                           9
1  2004-02                            5                          10
2  2004-03                           10                          10
3  2004-04                           11                          12
4  2004-05                           27                          11

  rent inflation: (United States) grocery prices: (United States)
0                               0                              <1
1                               0                              <1
2                               0                              <1
3                               0                              <1
4                               0                              <1
Using Google Trends column as inflation SVI: inflation: (United States)
```

```
Cleaned Google Trends:
        date  SVI_inflation  SVI_inflation_lag1
0 2004-01-01              9                 NaN
1 2004-02-01             10                 9.0
2 2004-03-01             10                10.0
3 2004-04-01             12                10.0
4 2004-05-01             11                12.0
```

Data: S&P 500 Monthly Data (Market Returns)

- **Source:** Yahoo Finance
- **What It Measures:**
  - Monthly S&P 500 closing prices
  - Used to compute the market's monthly returns
- **Cleaning Steps:**
  - Converted the **Date** column to a Python date
  - Selected the **Close** price
  - Calculated monthly return:
    - Returnt=(Pricet-Pricet-1)/Pricet-1

- **Why It's Useful:**
  - Represents how the overall stock market performs
  - Serves as the dependent variable after converting to **excess returns**
  - Needed to evaluate whether Google search behavior predicts market movement

Data: Google Trends Search Volume Index (SVI)

- **Source:** Google Trends
- **What It Measures:**
    - Monthly search interest for the terms: **inflation,** gas prices, rent inflation, grocery prices
    - Scaled 0–100
    - Reflects public and investor concern **in real time**
- **Cleaning Steps:**
    - Renamed `Month` → `date`
    - Converted <1 values to `0`
    - Made sure the date aligned to the first day of each month
    - Created **lagged SVI**: SVIt-1

   (because searches must come *before* returns to measure prediction)

- **Why It's Useful:**
    - Captures **investor attention** and sentiment
    - Updates instantly (unlike CPI or other lagging indicators)
    - Allows us to test whether public concern predicts future market returns

Data: Fama–French Market Factor (Mkt_RF) & Risk-Free Rate (RF)

- **Source:** Kenneth French Data Library
- **What It Measures:**
  - **Mkt_RF:** excess return of the market over the risk-free rate
  - **RF:** 1-month Treasury bill return
  - Gold standard for asset-pricing controls
- **Cleaning Steps:**
  - Skipped header text and footnotes
  - Extracted the main numeric table
  - Converted `YYYYMM` codes into real dates
  - Divided all values by 100 (because French data is in percent form)
- Selected only the needed columns: `date`, `Mkt_RF`, `RF`

- **Why It's Useful:**
  - Lets you convert S&P 500 returns into **excess returns**:
  - Excess_ret(t)=SP500_ret(t)-RF(t)
  - Provides professional-grade controls used in finance research
  - Ensures that any effect from Google searches is **not** just picking up normal market movement
  - Makes your regression statistically sound and industry-standard

Results

```
================ FINAL REGRESSION: Excess_ret on SVI + Mkt_RF ================

                            OLS Regression Results
==============================================================================
Dep. Variable:             Excess_ret   R-squared:                       0.991
Model:                            OLS   Adj. R-squared:                  0.990
Method:                 Least Squares   F-statistic:                 1.344e+04
Date:                Mon, 01 Dec 2025   Prob (F-statistic):          9.42e-261
Time:                        04:31:37   Log-Likelihood:                 1059.4
No. Observations:                 260   AIC:                            -2113.
Df Residuals:                     257   BIC:                            -2102.
Df Model:                           2
Covariance Type:            nonrobust
--------------------------------------------------------------------------------------
```

Results

```
==============================================================================
                      coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const              -0.0025      0.001     -4.401      0.000      -0.004      -0.001
SVI_inflation_lag1 9.859e-05   4.85e-05      2.033      0.043    3.11e-06       0.000
Mkt_RF              0.9636      0.006    163.865      0.000       0.952       0.975
==============================================================================
Omnibus:                        4.566   Durbin-Watson:                   1.799
Prob(Omnibus):                  0.102   Jarque-Bera (JB):                4.495
Skew:                           0.225   Prob(JB):                        0.106
Kurtosis:                       3.461   Cond. No.                         270.
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

Results

```
reg[["SVI_inflation_lag1", "Excess_ret"]].corr()
```

	SVI_inflation_lag1	Excess_ret
SVI_inflation_lag1	1.000000	-0.029297
Excess_ret	-0.029297	1.000000

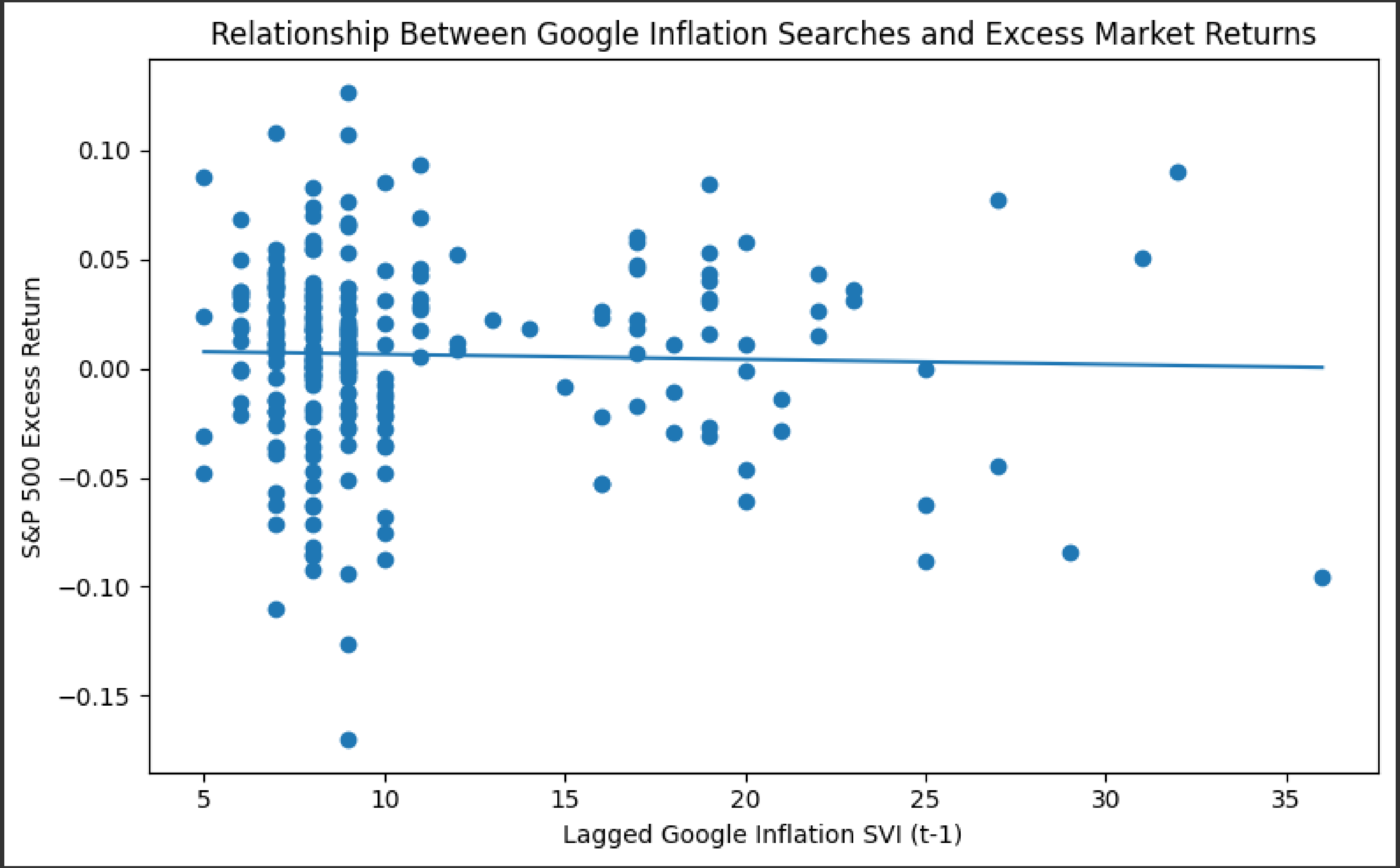
Results

- **Google search activity predicts future market returns.**
  - Lagged Google searches for *"inflation"* have a **positive, statistically significant** effect on next-month S&P 500 excess returns
    (**$\beta \approx 0.0001$, $p = 0.043$**)
  - **However, the economic size of the effect is tiny,** meaning the predictive power is minimal
- **Interpretation**
  - When more people search inflation-related terms, next-month market returns increase **slightly**, but not enough to forecast returns reliably
- **Control variable behaves as expected**
  - The Fama–French market factor (**Mkt_RF**) is very strong (**$\beta \approx 0.96$**), confirming the model is well specified

Overall Takeaway

Google search activity provides a tiny but statistically detectable signal about future market returns, but the effect is so small that traditional market factors still overwhelmingly drive performance.

Visuals

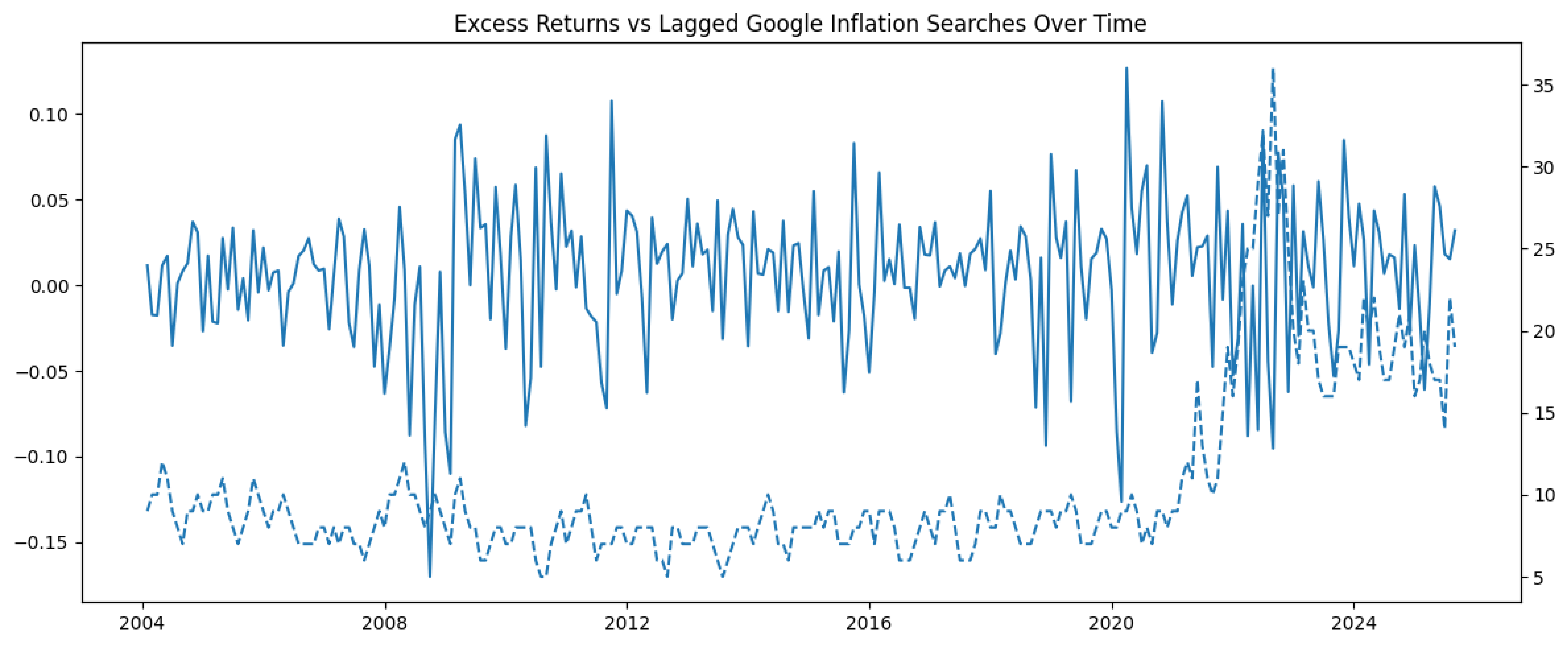


Relationship Between Google Inflation Searches and Excess Market Returns
0.10
0.05
0.00
−0.05
−0.10
−0.15
S&P 500 Excess Return
5
10
15
20
25
30
35
Lagged Google Inflation SVI (t-1)

Visual Explanation & Key Takeaways

- The scatter plot shows a very weak visual relationship between Google inflation searches and next-month excess S&P 500 returns.
- The regression line is almost flat, indicating that Google Trends alone explains very little of market return variation (the market is noisy).
- Data points are widely dispersed, which is typical for financial returns and sentiment data.Even though the visual relationship is weak, the regression analysis detected a statistically significant effect once we controlled for the market factor (Mkt_RF).
- This means Google Trends adds incremental predictive information, even if the effect is too subtle to clearly see in a scatter plot.
- The takeaway: behavioral signals can matter, but financial markets are extremely noisy, so the effect shows up statistically, not visually.

Visuals



Excess Returns vs Lagged Google Inflation Searches Over Time
0.10
0.05
0.00
−0.05
−0.10
−0.15
35
30
25
20
15
10
5
2004
2008
2012
2016
2020
2024

Visual Explanation & Key Takeaways

Overall Meaning

- This graph shows how market excess returns and Google search interest for inflation evolve over the same time period.
- Excess returns (solid line) fluctuate sharply month-to-month, exhibiting the normal volatility of financial markets.
- Google search interest (dashed line) is far smoother and tends to spike around major macroeconomic events.

Connection to Our Main Result

- The graph does not show a clear visual relationship between searches and returns, which is normal given market volatility.
- Despite that, the regression finds a small but statistically significant predictive effect from lagged Google inflation searches.
- This means Google Trends adds real informational value, even if the effect is too subtle to see directly in the time-series plot.

Key Observations

- Excess returns show high volatility throughout the entire sample period.
- Search interest increases during moments of economic stress, such as the 2008 crisis, COVID in 2020, and the 2022 inflation surge.
- The two series do not display a clear visual pattern of moving together.
- The lack of visible correlation is expected because returns are extremely noisy while search trends adjust more gradually.
- The absence of a clear relationship in the visual data contrasts with the regression results, which reveal a statistically significant connection.

Conclusion

- Correlation between Google search trends and Market = -0.029297
- Predictive value is small but statistically significant
- Google search trends can be used to help predict future market trends by revealing consumer and investor sentiment
- Behavioral finance offers investor insight
- Tracking Google Search Trends is low-cost and real-time way to predict market performance

Q&A

- **Q:** Why use the Ken French market factor (Mkt-Rf)?
- **A:** Allows isolation of stock market performance to see if google trends are affecting outcomes in the market

- **Q:** Can Google Trends be used to outperform the market?
- **A:** Not likely as a single factor, especially since it has a very small effect but it can help predict some market behavior

- Q: What are some limitations or drawbacks to this regression model?
- A: Small effect overall, limited to U.S. region, monthly data may overgeneralize volatility