Analysis of Stock Market Fluctuations and Fear Index Using Data-driven Methodologies

SEPIDEH FOROUZI

GITHUB: HTTPS://GITHUB.COM/SEPIDFS/DTSA-5509

Problem Description

Problem

- Predict the next month's value of the VIX index (market's fear gauge).
- Understand the economic, financial, and political factors influencing market volatility.

Approach

- Use **text mining** on news articles from over 1,000 U.S. newspapers.
- Extract signals using the **Equity Market Volatility (EMV) Tracker**.

Dataset

- Monthly VIX index values and EMV categories from **January 1990 to December 2022**.

Methods

- Apply four regression models:
  OLS, **Ridge**, **LASSO**, and **Elastic Net**.
- Perform segmented analysis to study **market regime shifts** over time.

VIX Index

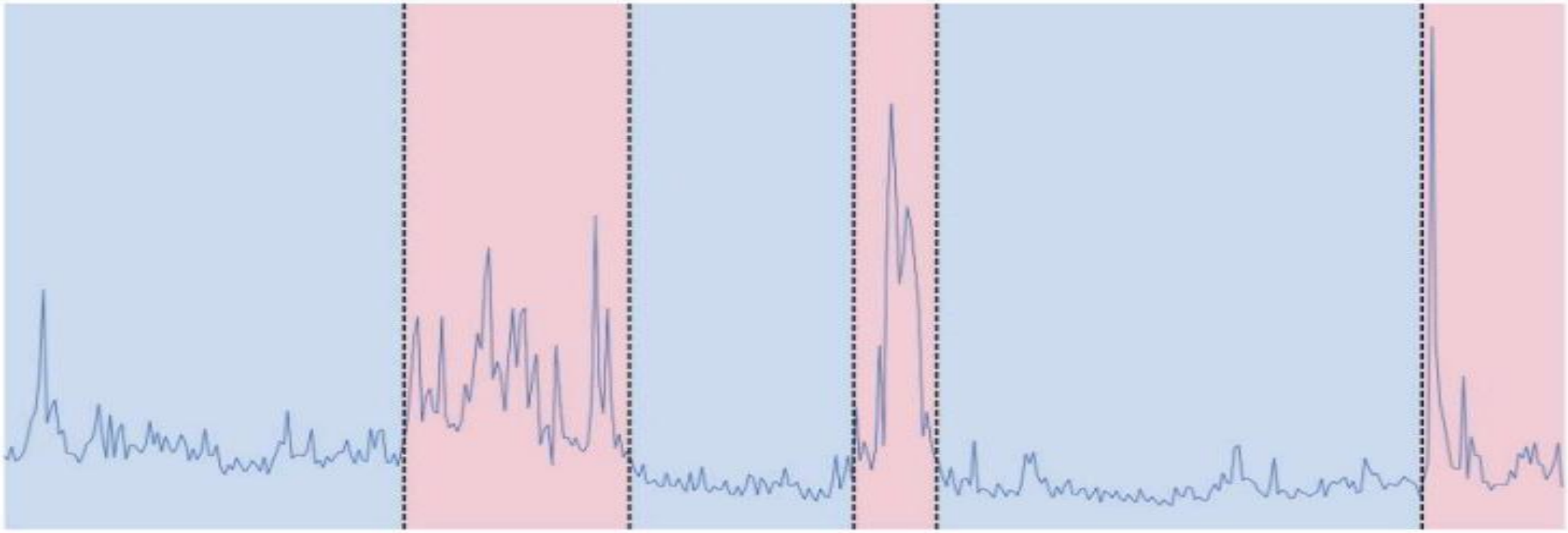


Figure 1: Figure 1: VIX value over time. Pink segments are more volatile, while blue segments are calmer.

Dataset and Attributes

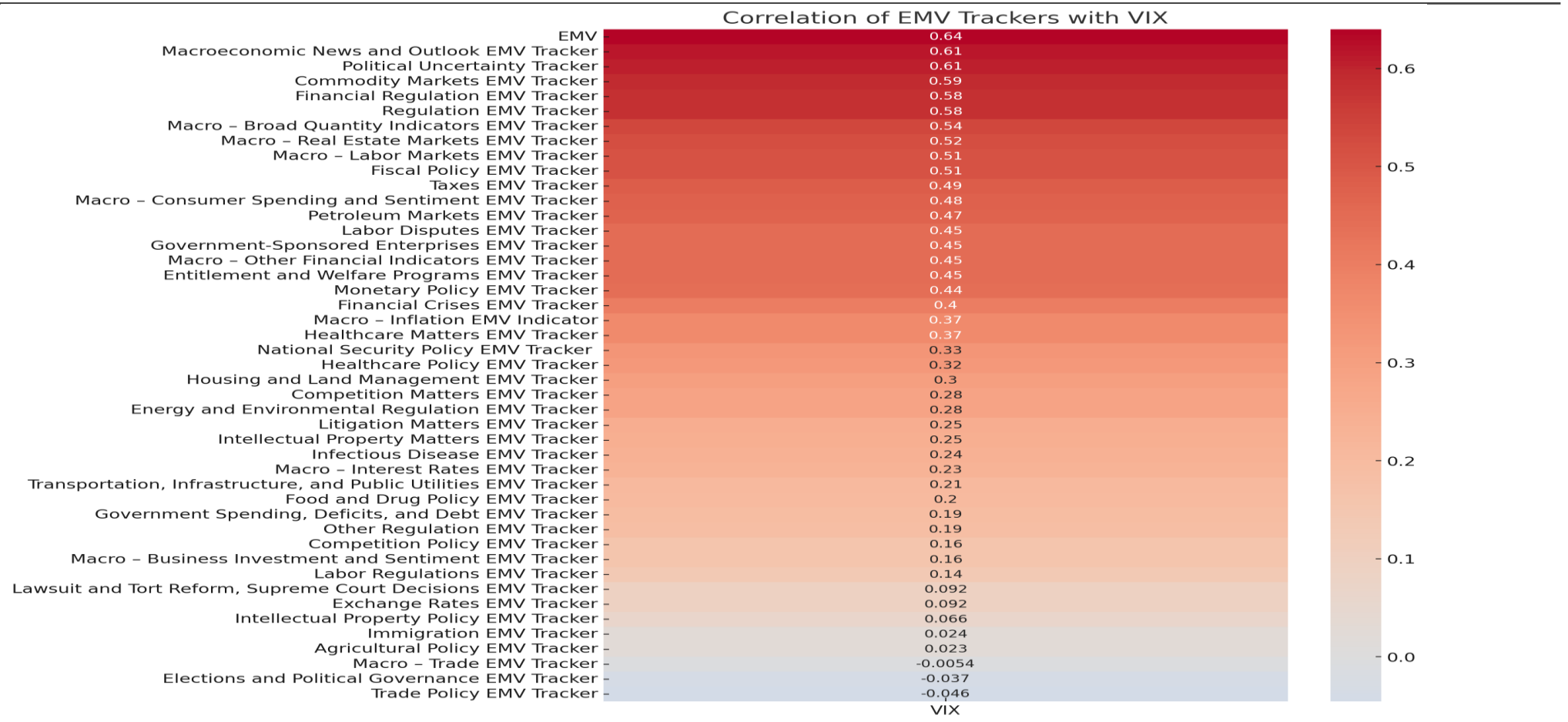
Main Attributes

- **Date:** Month and year of observation
- **VIX:** Monthly closing value of the VIX index (market volatility measure)
- **EMV Tracker:** General measure of economic and market uncertainty
- **45 Additional Trackers:**
    - Cover different topics such as:
        - Political Uncertainty
        - Infectious Disease News
        - Inflation and Interest Rates
        - Financial Crises and Real Estate
        - Labor Disputes, Financial Regulation, Trade Policy, Elections, and more

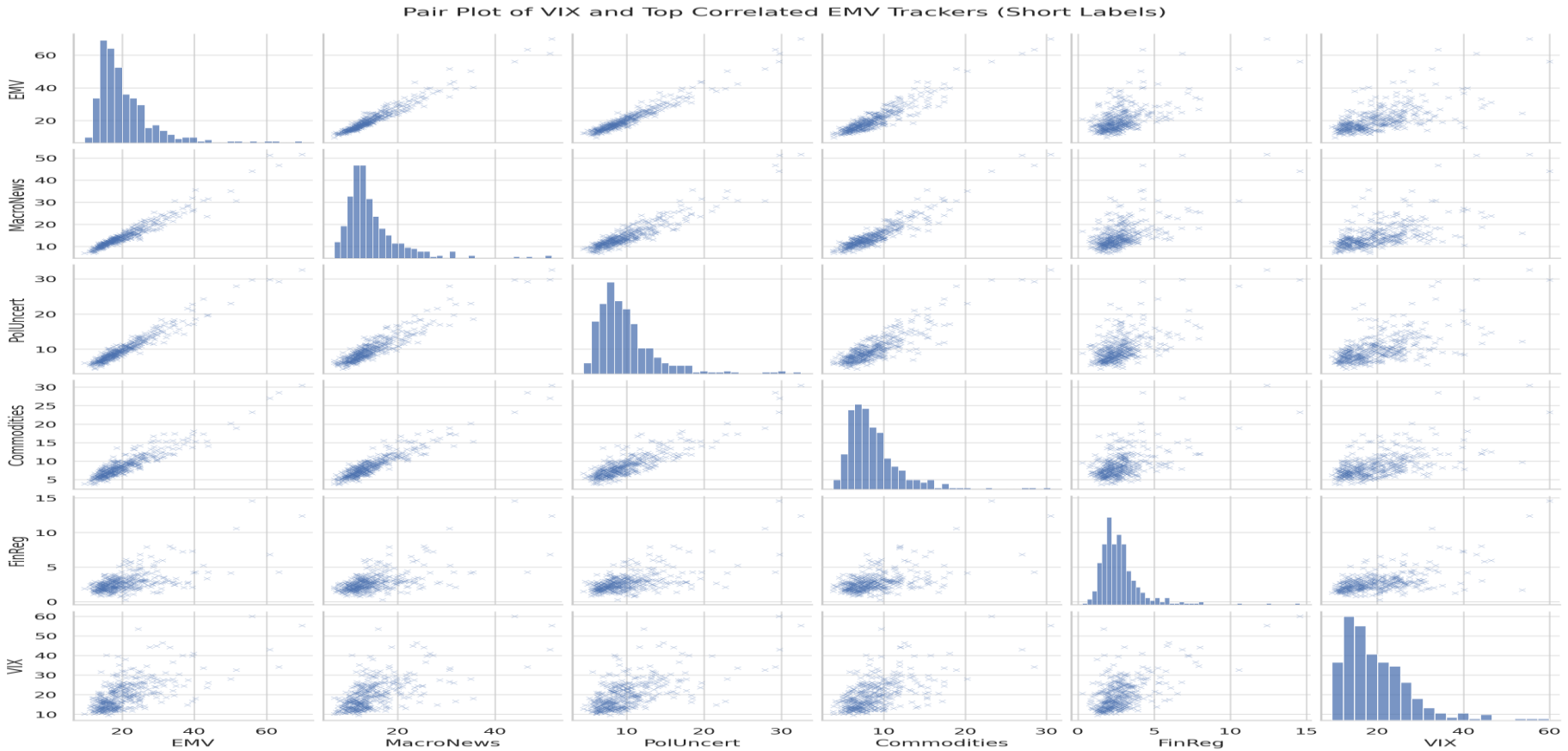
Target Variable

- **Next Month's VIX Value** (one-step-ahead prediction)

Explanatory Data Analysis



Explanatory Data Analysis



Elastic Net Model

Why Elastic Net?

- Combines the strengths of **Ridge** and **LASSO** regression.
- Helps handle **multicollinearity** and **feature selection** together.

Objective Function

$$\min_{\beta_0,\beta} \frac{1}{N}\sum_{i=1}^{N} w_i L(y_i, \beta_0 + \beta^T x_i)] + \lambda[(1-\alpha)||\beta||^2{}_2 + \alpha||\beta||_1]$$

- **λ (lambda):** Controls the amount of regularization (penalty size).
- **α (alpha):** Balances between Ridge (α = 0) and LASSO (α = 1).

Why Use It Here?

- Market data shows strong correlations (multicollinearity).
- Elastic Net finds a better balance between bias and variance.
- Helps select important predictors while keeping model stability.

Results and Discussions

Segment	Time Period	α	R^2	MSE	RMSE
1	1990–1993	0.10	0.709	0.0216	0.1469
2	1993–1998	0.60	0.638	0.0222	0.1490
3	1998–2002	0.75	0.691	0.0296	0.1721
4	2003–2007	0.85	0.696	0.0194	0.1393
5	2007–2009	0.80	0.917	0.0154	0.1242
6	2010–2019	0.45	0.702	0.0185	0.1359
7	2020–2023	0.70	0.854	0.0243	0.1559

Results and Discussions

Segment	Predictor 1	Predictor 2	Predictor 3	Predictor 4	Predictor 5
1	Disease (-0.63)	Labor (-0.61)	Commodities (+0.38)	Macro News (-0.26)	Regulation (-0.16)
2	Macro News (+0.62)	Regulation (-0.48)	Policy (-0.47)	Monetary (+0.27)	Labor (+0.16)
3	Labor (-0.86)	Macro News (-0.47)	Fiscal (+0.45)	Elections (+0.44)	Policy (-0.34)
4	Macro News (-0.65)	Trade (+0.46)	Elections (-0.35)	Regulation (-0.33)	Policy (-0.24)
5	Policy (-0.84)	Macro News (-0.76)	Fiscal (+0.56)	Trade (+0.55)	Labor (+0.41)
6	Labor (-0.61)	Fiscal (-0.42)	Policy (+0.35)	Macro News (-0.27)	Elections (+0.21)
7	Disease (+1.18)	Fiscal (+0.75)	Elections (+0.65)	Macro News (-0.38)	Policy (-0.28)

Thank You!