Information Theory Project Proposal

Macroeconomic Indicators and the U.S. Stock Market: An Information-Theoretic Analysis of Dynamic Information Flows

1. Motivation, Aim, and Materials

1.1 Motivation and Significance

Understanding the dynamic relationship between macroeconomic conditions and the stock market is a central topic in both financial economics and investment practice. Traditional econometric methods, such as Granger causality or regression analysis, often assume linear relationships and may fail to detect more complex, nonlinear dependencies. Information theory offers a more general framework for exploring these relationships. Measures such as Transfer Entropy (TE) can capture the direction and strength of information flow between time series, regardless of whether the dependencies are linear or nonlinear.

The main goal of this project is to use information-theoretic measures to examine how key U.S. macroeconomic variables (such as interest rates and inflation) and the S&P 500 index exchange information over time. Specifically, the study will explore whether macroeconomic indicators drive stock market movements, or if the stock market anticipates and predicts changes in the economy. Understanding these dynamics has clear implications for policymakers (interpreting market signals) and investors (anticipating economic cycles and managing risk).

1.2 Research Questions and Hypotheses

- **Question 1:** What is the dominant direction of information flow between macroeconomic variables and the stock market? How does this change over time?
  Hypothesis 1: The stock market usually leads macroeconomic indicators (reflecting market expectations), but during major policy shifts such as interest rate hikes, macroeconomic variables exert stronger influence on market movements.
- **Question 2:** After controlling for other factors, how much unique information does a single macroeconomic variable (e.g., the Federal Funds Rate) contribute to predicting stock market movements?
  Hypothesis 2: The interest rate has a significant net effect on the stock market even after controlling for inflation.
- **Question 3:** Does the pattern of information flow change significantly across different economic periods (e.g., the 2008 financial crisis vs. stable growth periods)?

1.3 Data Sources

(a) Stock Market Data:

- S&P 500 Index (monthly average of daily closing prices), retrieved using the yfinance Python package.
- Time range: 1990–present (covering multiple economic cycles).

(b) Macroeconomic Indicators:

- Effective Federal Funds Rate (interest rate)
- Consumer Price Index (CPI) (used to calculate inflation rate)
- Unemployment Rate
- (Optional) Industrial Production Index, PMI (Purchasing Managers Index)

All macroeconomic data will be obtained from the Federal Reserve Economic Data (FRED) database, which provides open and reliable time series.

1.4 Relation to Existing Literature

Previous studies have applied Transfer Entropy to financial systems (e.g., (Marschinski & Kantz, 2002) and to explore causality between macroeconomic factors and markets (e.g., Sensoy et al., 2014). However, most of these studies either focused on single variables or used static, full-sample analysis.
This project extends prior work in two main ways:

(a) Multivariate and Conditional Analysis: By including several key macroeconomic indicators, conditional TE can be used to isolate the "net" information contribution of each variable while controlling for others.

(b) Dynamic Perspective: Using a rolling-window approach allows exploration of how information flow changes across time, capturing the effects of major economic events such as the 2008 crisis or the COVID-19 pandemic.

2. Method

2.1 Data Preprocessing

(a) Time Alignment: Since macroeconomic data are monthly and stock market data are daily, the S&P 500 series will be aggregated to monthly averages. All time series will be aligned by date.

(b) Transformation:

- Stock market data will be converted to log returns: $r_t = \log(P_t) - \log(P_{t-1})$.
- Inflation rate will be calculated from year-on-year CPI changes
- Interest rate and unemployment rate will be used in levels or first differences.

(c) Stationarity Check: Each series will be tested for stationarity (ADF test). Differencing or detrending will be applied if necessary.

(d) Normalization: Data will be normalized to zero mean and unit variance for comparability.

(e) (Optional) Data may also be discretized into bins to compare results between discrete and continuous estimators for robustness.

2.2 Information-Theoretic Measures

All measures will be implemented using the JIDT (Java Information Dynamics Toolkit).

(a) Transfer Entropy (TE):

- Used to quantify the direction and magnitude of information flow, e.g., TE(InterestRate → StockMarket) vs. TE(StockMarket → InterestRate).
- Estimator: TransferEntropyCalculatorKraskov (for continuous data).
- Parameters:
  - History length $k = 1, 2, 3$ (chosen via AIC or validation).
  - Prediction length $l = 1$.
  - Nearest neighbours $k_{NN} = 4$ (default).
- Statistical significance: Determined via surrogate (permutation) testing with 1000 shuffles, keeping only results with ($p < 0.05$).

(b) Conditional Transfer Entropy (CTE):

- Used to identify the net directional influence of one variable while controlling for others, e.g., CTE(InterestRate → StockMarket | Inflation) removes indirect effects of inflation.
- Estimator: ConditionalTransferEntropyCalculatorKraskov.

(c) Mutual Information (MI):

- Used as a baseline to assess the overall (non-directional) statistical dependence between variables.

2.3 Analysis Plan and Interpretation

(a) Full-sample Baseline: Compute average TE and CTE for the full sample to establish general information flow patterns.

(b) Rolling-window Analysis: Apply a 60-month (5-year) rolling window to compute TE dynamics over time.

(c) Event Overlay: Compare temporal changes in TE with key macroeconomic events (e.g., 2008 crisis, COVID-19 onset, major Fed rate changes).

(d) Interpretation:

- A higher TE value from X→Y indicates stronger predictive information flow.
- Comparing TE(Macro→Market) and TE(Market→Macro) identifies the dominant direction.
- Sudden spikes or drops in TE may reflect shifts in market expectations or policy impacts.

2.4 Expected Outcomes and Implications

- Stock markets often lead macroeconomic indicators under normal conditions, showing anticipatory behaviour.
- During monetary tightening cycles or crises, macroeconomic factors (like interest rate shocks) may dominate the information flow.
- Conditional TE will reveal which indicators (e.g., interest rate vs. inflation) have unique predictive roles after controlling for others.

2.5 Plan for Final Presentation

- Show the data preprocessing steps and brief code implementation in JIDT.
- Present time-varying information flow graphs and discuss how they align with economic events.
- Critically evaluate the approach, noting limitations such as parameter sensitivity, omitted variables, and potential common drivers.

AI Acknowledgment: Draft developed with assistance from ChatGPT (OpenAI GPT-5) for editing, structure, and clarity improvement. Analytical design, research ideas, and interpretation remain my own.

Reference

FRED. "FRED Economic Data." *Stlouisfed.org*, Federal Reserve Bank of St. Louis, 2025, fred.stlouisfed.org/.

Lizier, Joseph T. "JIDT: An Information-Theoretic Toolkit for Studying the Dynamics of Complex Systems." *Frontiers in Robotics and AI*, vol. 1, 2 Dec. 2014, https://doi.org/10.3389/frobt.2014.00011.

Marschinski, R., and H. Kantz. "Analysing the Information Flow between Financial Time Series." *The European Physical Journal B*, vol. 30, no. 2, Nov. 2002, pp. 275–281, https://doi.org/10.1140/epjb/e2002-00379-2. Accessed 25 June 2022.

Sensoy, Ahmet, et al. "Effective Transfer Entropy Approach to Information Flow between Exchange Rates and Stock Markets." *Chaos, Solitons & Fractals*, vol. 68, Nov. 2014, pp. 180–185, https://doi.org/10.1016/j.chaos.2014.08.007. Accessed 20 Aug. 2021.