

Aim/Materials

The aim of this study is to apply information theory to financial data in order to understand the magnitude and direction of information transfer between economies and industries. Datasets used includes daily price of stock market indices, exchange traded funds (ETFs), and stock, all of which were obtained from Yahoo! Finance. List of data used is described in tables 1, 2, and 3 below. Stock market indices will be used to represent respective markets, while ETFs will be used to represent industry sectors.

Table 1: Stock market indices representing different markets.

Ticker	Name	Market	n	Source
[^] GSPC	S&P 500	USA	500	Link
[^] N100	Euronext 100	Europe	100	Link
[^] N225	Nikkei 225	Japan	225	Link
[^] BVSP	Bovespa Index	Brazil	70	Link

Table 2: Exchange Traded Funds (ETFs) representing industry sectors.

Ticker	Name	Industry	Source
USO	United States Oil Fund	Oil	Link
ICLN	iShares Global Clean Energy ETF	Renewable Energy	Link
JETS	U.S. Global Jets ETF	Airline	Link
IYT	iShares Transportation Average ETF	Transportation	Link
XLP	Consumer Staples Select Sector SPDR Fund	Consumer staples	Link
SMH	VanEck Vectors Semiconductor ETF	Semiconductor	Link
IXP	iShares Global Comm Services ETF	Telecom	Link
VGT	Vanguard Information Technology Index Fund	Technology	Link
XPH	SPDR S&P Pharmaceuticals ETF	Pharmaceutical	Link

Table 3: Stocks

Index	Name	Industry	Source
XOM	ExxonMobil	Oil	Link
DAL	Delta Airline	Airline	Link

The primary motivation of this study is to explore whether information theory can be applied to uncover the dynamics between economies, as well as industries. In particular, we are interested in the amount and direction of information transfer. Through data analysis, we aim to test the hypothesis: The United States economy, represented by S&P 500, has more influence on Japanese economy, represented by Nikkei 225, than the other way around.

Method

To test the above hypothesis, transfer entropy will be used. Transfer entropy, $T_{X \rightarrow Y}$, measures the amount of information moving from a process or a random variable X to another process, Y . Transfer entropy is the most suitable metric to test our hypothesis, as interaction within financial markets is non-linear, and takes place over long periods. The closest metric to transfer entropy would be correlation. However, correlation is not optimal for this study, since correlation only extracts linear interactions, and is bidirectional. Another similar metric would be mutual information, which is another information-theoretic measure. Like correlation, mutual information does not give us the direction of information transfer. Since transfer entropy is asymmetric, we can calculate the transfer entropy for both directions, and subtract one from another to obtain $F_{X \rightarrow Y}$, the net information flow from X to Y .

Transfer entropy is calculated by:

$$T_{X \rightarrow Y} = \sum p(y_{t+1}, y_t, x_t) \log \frac{p(y_{t+1} | y_t, x_t)}{p(y_{t+1} | y_t)} \quad (1)$$

Net Information Flow $F_{X \rightarrow Y}$ is the quantified influence, or the net flow of information, from X to Y , and is defined as:

$$F_{X \rightarrow Y} = T_{X \rightarrow Y} - T_{Y \rightarrow X} \quad (2)$$

If the net information flow $F_{X \rightarrow Y}$ is positive, then X has more influence on Y than vice versa, and describes the level of influence X has on Y .

To calculate transfer entropy, instead of using the daily adjusted closing price, we first calculated the natural log of the daily return.

To calculate r_t , the natural log of daily return at time t ,

$$r_t = \ln p_t - \ln p_{t-1} = \ln\left(\frac{p_t}{p_{t-1}}\right) \quad (3)$$

There are two reasons why we use natural log of daily return instead of the closing price. First, price data has no upper bound, and therefore is not stationary. The second reason is that the price of a share of a company or an index being larger than another is not a function of influence or performance, but of number of shares issues by the company, at least initially when the stocks are issued. Daily return, however, is independent of the differences in price among companies or indices, and preserves information about the movement of price.

For calculating transfer entropy, we used the Java Information Dynamics Toolkit ¹, which is an open-source tool for computing information-theoretic measures available to many programming languages. Since we are dealing with time-series, we used Kraskov estimator, instead of using discretization, gaussian model or box-kernel estimation. The reason is that box-kernel estimation and discretization are sensitive to the size of the box or bin used, and gaussian model only detects linear component of interactions between the variables. In other words, the analysis will be based on the assumption that there is nonlinear dependency between the variables X and Y . For parameters of the Kraskov estimator, history length of $k = 1$ and nearest neighbors of 4 was used. Nearest neighbor was set to 4 to prevent undersampling, and we didn't use a larger value because there are only 364 data points of log return values for each year that is considered. History length of $k = 1$ was used because by using log return instead of closing price, we expect little or no autoregression. Since log daily return is a continuous variable, information theoretic measures will be expressed in nats, using base e .

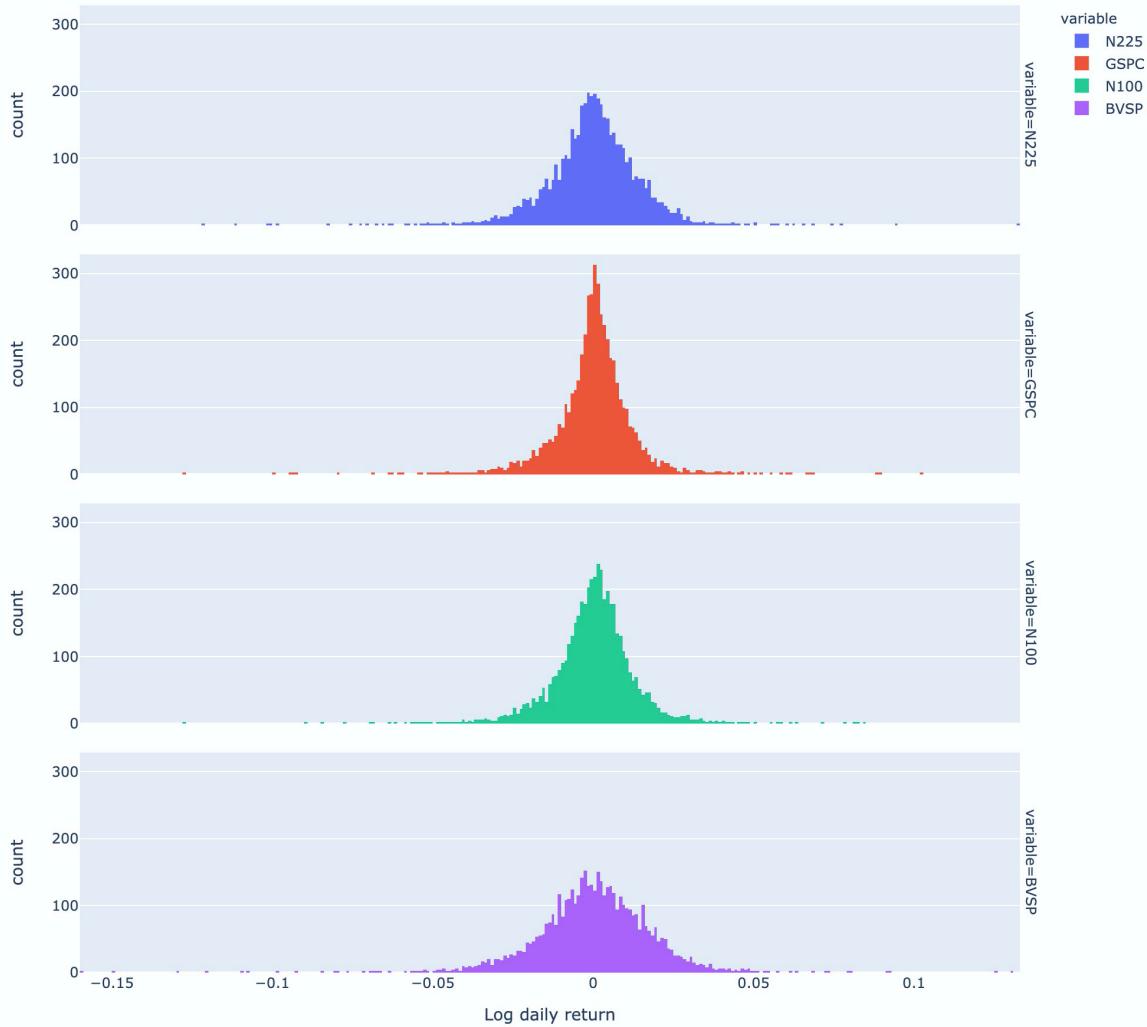
Results

Market Index

First step of the analysis is converting daily adjusted closing price of each market index to log daily return. Histograms in figure 1 below shows the distribution of daily return for each of the markets. We can see that log daily return is approximately normal and symmetrically distributed with different tail lengths. If we consider Brazil as an "emerging economy," the longer tails makes sense, as emerging economies are less predictable and tends to be more volatile than developed economies like the United States.

Figure 1: Histogram of log daily return: N225 = Japan, GSPC = United States, N100 = Europe, BVSP = Brazil.

Distribution of daily return



To test the hypothesis that the United States economy has more influence on Japanese economy, transfer entropy was calculated for both directions, GSPC to N225, and N225 to GSPC, for each year since 2000. As shown in figure 2 below, transfer entropy from GSPC to N225 is greater than N225 to GSPC for every time period. The transfer entropy in both directions and net information flow from GSPC to N225 is shown in figure 3. From these two figures, it is evident that while there's information transfer in both directions, the United States economy has far more influence on the Japanese economy than vice versa.

Figure 2: Transfer Entropy between the United States (GSPC) and Japan (N225), from 2000 to 2020.

Transfer Entropy between GSPC and N225

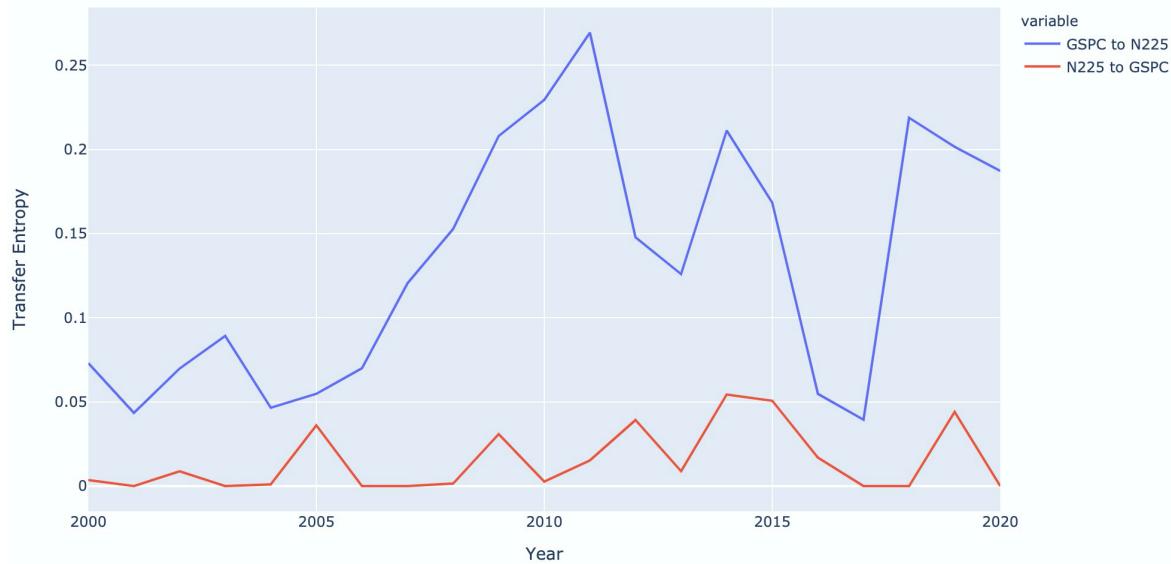


Figure 3: Transfer Entropy between GSPC (USA) and N225 (Japan), and Net Information Flow from GSPC to N225.

year	GSPC to N225	N225 to GSPC	net_info_flow
2000	0.072982	0.003568	0.069414
2001	0.043448	0.000000	0.043448
2002	0.069751	0.008767	0.060984
2003	0.089150	0.000000	0.089150
2004	0.046556	0.000993	0.045563
2005	0.054815	0.036079	0.018736
2006	0.069950	0.000000	0.069950
2007	0.120493	0.000000	0.120493
2008	0.152673	0.001503	0.151170
2009	0.207931	0.030833	0.177098
2010	0.229401	0.002642	0.226759
2011	0.269289	0.015190	0.254099
2012	0.147744	0.039211	0.108533
2013	0.125985	0.008871	0.117115
2014	0.211213	0.054382	0.156831
2015	0.168373	0.050655	0.117718
2016	0.054729	0.016947	0.037783
2017	0.039395	0.000000	0.039395
2018	0.218760	0.000000	0.218760
2019	0.201544	0.044058	0.157486
2020	0.187151	0.000000	0.187151

Using all of the data available from the year 2000, we also calculated the overall transfer entropy. The transfer entropy from the United States to Japan is 0.1327 nats with p-value of 0, while the transfer entropy from Japan to the United States is 0.018 nats with p-value of 0.01. Using a threshold of 0.05, we fail to reject the null hypothesis that the United States has more influence on Japan, with a result that is statistically significant. The result of the data analysis suggests that the log return of N225 can be predicted to an extent from the log return of S&P500 (GSPC).

Transfer entropy was also calculated in both directions for every pair of markets: Japan, the United States, Europe, and Brazil. Figure 4 below shows each of the transfer entropy values. The United States (GSPC) indeed has the highest influence on both Japan and Europe, but Japan seems to have more influence on Brazilian economy than the United States.

Performing row-wise summation and normalizing the sum by dividing by the total transfer

entropy also gives us the transfer entropy contribution, which tells us the percentage of contribution each of the source market made to the overall transfer entropy (table 4). From the transfer entropy contribution, we can see that the United States has the most influence on the rest of the markets, followed by Europe and Japan. With contribution of 0.474, the United States has almost as much influence as Europe, Japan, and Brazil combined.

Figure 4: Transfer Entropy Matrix

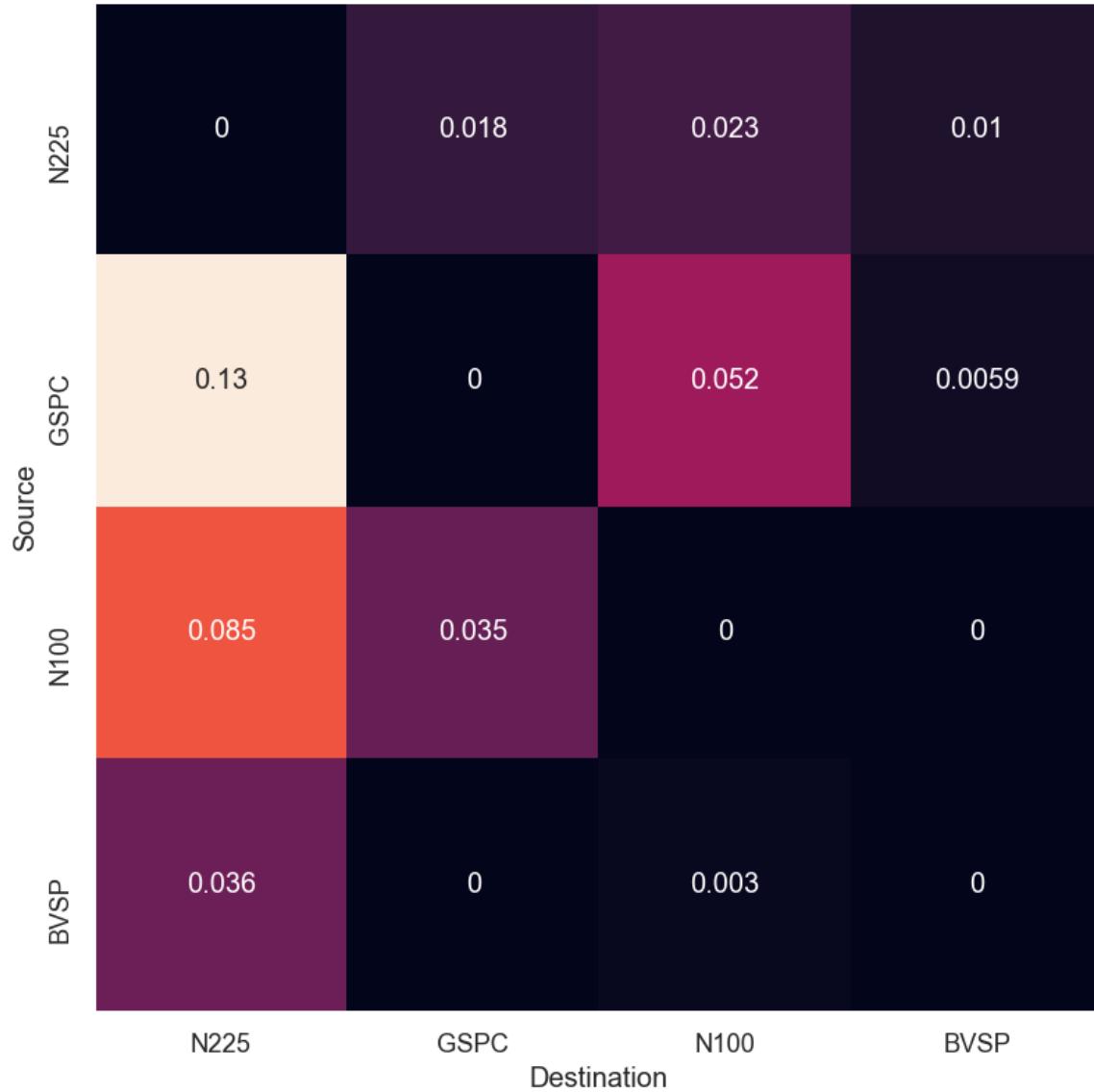


Table 4: Transfer Entropy Contribution of US, Europe, Japan, and Brazil.

Source Index	Transfer Entropy Contribution
GSPC (USA)	0.4740
N100 (Europe)	0.2995
N225 (Japan)	0.1279
BVSP (Brazil)	0.0987

Industry Index (ETFs)

Using selected ETFs to represent various industries, transfer entropy contribution was calculated for each year (figure 5). Since some ETFs have existed for longer than others, consecutive years starting from the year 2000 with transfer entropy contribution of 0 indicate newer ETFs, such as the Airline ETF, which has only existed from 2016. For each year, color scale was applied to emphasize influential industries within each year. Some of the emphasized values were selected and validated by looking at relevant news articles from that year relating to the industry.

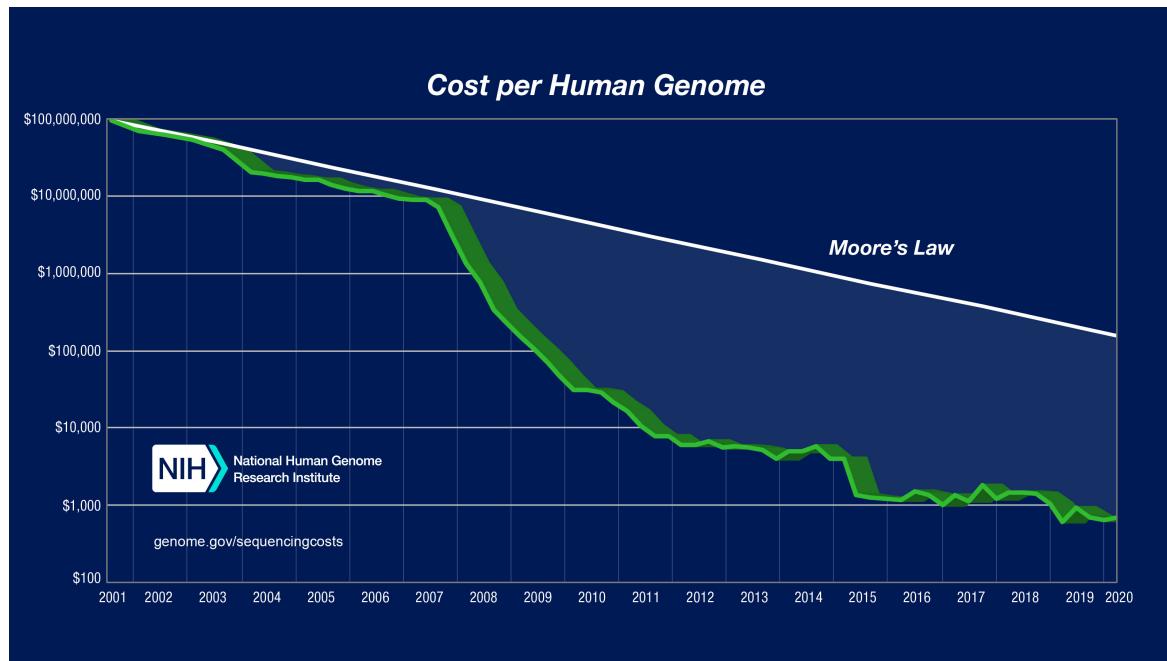
Figure 5: Yearly Transfer Entropy Contribution by Industry Sector (Color scale applied row-wise. Darkest color indicate the maximum TE contribution for that year.)

Year	Pharmaceutical	Semiconductor	Renewable Energy	Airline	Consumer Staples	Oil	Telecom	Transportation	Technology
2002	0.000000	0.351504	0.000000	0.000000	0.546312	0.000000	0.102184	0.000000	0.000000
2003	0.000000	0.268880	0.000000	0.000000	0.502318	0.000000	0.228802	0.000000	0.000000
2004	0.000000	0.057117	0.000000	0.000000	0.131527	0.000000	0.255277	0.425826	0.130253
2005	0.000000	0.201834	0.000000	0.000000	0.109531	0.000000	0.087163	0.359444	0.242029
2006	0.000000	0.138424	0.000000	0.000000	0.245103	0.000000	0.227399	0.298876	0.090199
2007	0.023105	0.097225	0.000000	0.000000	0.039366	0.213104	0.068295	0.280837	0.278068
2008	0.206011	0.027874	0.000000	0.000000	0.047243	0.252743	0.199062	0.106525	0.160542
2009	0.025382	0.081823	0.067746	0.000000	0.241051	0.085381	0.103899	0.252535	0.142484
2010	0.074009	0.036683	0.164379	0.000000	0.004832	0.246538	0.268585	0.075735	0.129240
2011	0.153696	0.022067	0.002933	0.000000	0.245540	0.196389	0.127495	0.063833	0.188046
2012	0.106558	0.046596	0.126051	0.000000	0.035346	0.160731	0.157176	0.110552	0.256990
2013	0.157566	0.136117	0.131832	0.000000	0.108710	0.081415	0.163197	0.123647	0.097515
2014	0.156866	0.071775	0.175108	0.000000	0.015930	0.143599	0.145213	0.117638	0.173871
2015	0.134925	0.066612	0.121931	0.000000	0.081826	0.123902	0.240245	0.119352	0.111208
2016	0.030367	0.086001	0.202885	0.169737	0.093977	0.078701	0.168365	0.065023	0.104944
2017	0.044076	0.074593	0.086967	0.092864	0.125465	0.136620	0.105720	0.232160	0.101535
2018	0.086298	0.173048	0.016955	0.082027	0.115597	0.097943	0.194383	0.097919	0.135830
2019	0.106301	0.121157	0.093962	0.223178	0.104455	0.117606	0.083892	0.057465	0.091983
2020	0.137117	0.132593	0.118680	0.019373	0.235498	0.023934	0.156946	0.095072	0.080786

Pharmaceutical

2007 to 2008 was an important year for pharmaceutical industry, and represented an important milestone in human genome sequencing. Between 2007 and 2008, the cost of human genome sequencing dropped sharply from \$10 million USD to less than \$500,000 USD² (figure 6).

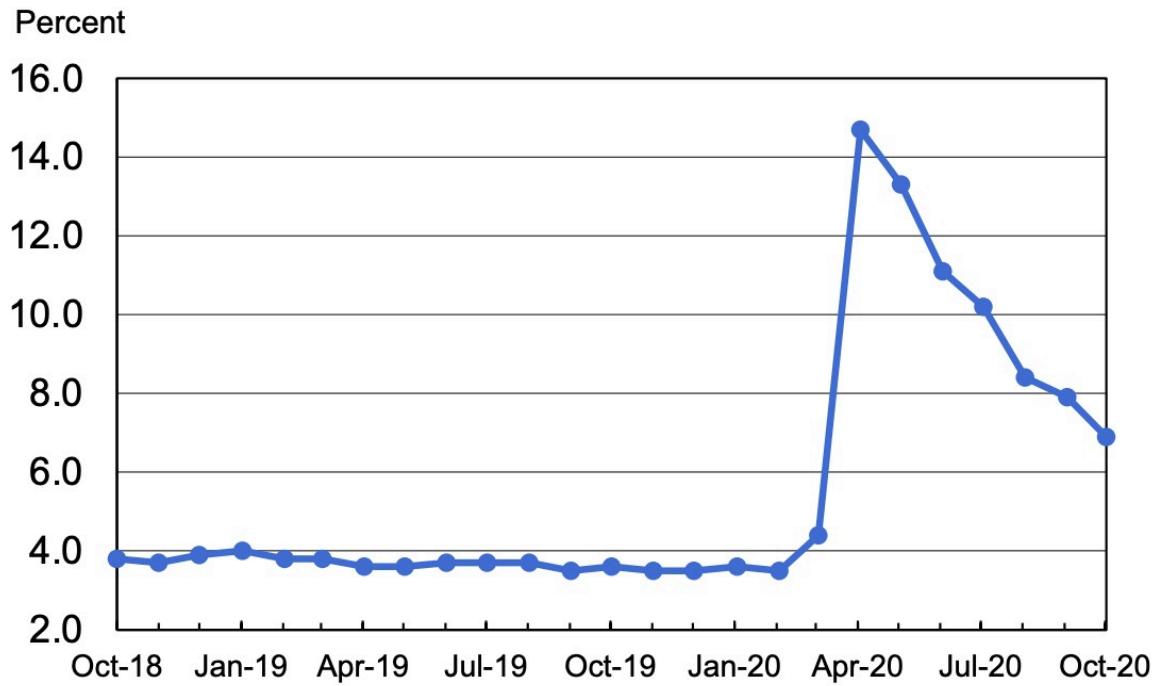
Figure 6: Cost of Human Genome Sequencing



Consumer Staples

Consumer staples refers to essential products such as food, beverage, household goods, and hygiene products³. The transfer entropy contribution in 2020 being significantly higher than the preceding years makes sense, as unemployment rate in the United States surged in 2020 due to COVID-19 pandemic causing consumers to refrain from leaving their homes. The high unemployment rate likely led to cutback in consumer spending on nonessential goods. In other words, consumer staples is one of the industries that are most resilient to economic downturn.

Figure 7: United States unemployment rate, seasonally adjusted, October 2018 - October 2020



Source: United States Bureau of Labor Statistics ⁴

Oil

In the year 2008, price of crude oil reached an all time high of 147.5 USD per barrel before a massive crash. This was partially caused by the weakening of the USD currency after \$70 billion USD Wallstreet bailout following the financial crisis of 2008. There was also a significant increase in unemployment and cutback on consumer spending due to the recession, and demand for energy fell drastically as a result ⁵.

Figure 8: Adjusted Daily Closing Price of USO (United States Oil Fund)

USO (Oil) Adjusted Closing Price



Discussion

Based on the findings discussed in the Results section, transfer entropy has shown to be an effective measure of the direction and magnitude of information transfer between markets and industries, which can be used to distinguish influential markets and industries. Transfer entropy also seems to capture the dynamics regardless of whether the correlation is inverse or not between two processes. The analysis done in this research can also be extended or improved in different ways. First, choice of using ETFs to represent industries should be avoided, as the ETFs are diversified to different extents, and some of the ETFs were not mutually exclusive. The benefit of using transfer entropy is that it removes active information storage. However, if the two processes contain the same stock, then the resulting transfer entropy also contains active information storage. Instead of ETFs, a number of stocks with the largest market capitalization should be selected from each industry to ensure disjointness of stocks between industries. Second, the input data can be discretized so that entropy estimators are not necessary. The financial data can be discretized by whether the daily return was positive or negative, or whether the return increased or decreased from the previous time step.

Appendix

Reference

1. <https://github.com/jlizier/jidt> ↵
2. <https://www.genome.gov/about-genomics/fact-sheets/Sequencing-Human-Genome-cost> ↵
3. <https://www.investopedia.com/terms/c/consumerstaples.asp#:~:text=Consumer%20staples%20are%20essential%20products,items%20as%20alcohol%20and%20tobacco> ↵
4. <https://www.bls.gov/news.release/pdf/empsit.pdf> ↵
5. <https://www.investopedia.com/ask/answers/052715/how-did-financial-crisis-affect-oil-and-gas-sector.asp> ↵