EXPLORING THE ABILITY OF MEDICARE SALES DATA TO PREDICT FUTURE STOCK MOVES IN HEALTH INSURANCE CONGLOMERATES

Jessica Chipera

Dr. Amlan Chatterjee

Northcentral University

TIM 8550

Week 8

Overview

Data provided by a Medicare brokerage firm contains sales from the past five Medicare Annual Election Periods. According to (Medicare.gov, n.d.), the Annual Election Period occurs between October 15 and December 7 every year. During this time, a Medicare beneficiary should choose what Parts C and D coverage they want for the following year. Parts C and D are products of health insurance companies and created to supplement Parts A and B, which are provided by the government (Medicare.gov, n.d.). Needless to say, Medicare contributes a significant portion to the revenues of large health insurance companies.

The purpose of this experiment is to determine if sales data from the Medicare Annual Election Period could be used to predict future stock movements in these publicly traded health insurance companies. My hypothesis is that this data can in fact be used to predict future stock moves. Therefore, the null hypothesis is that the data cannot be used for that purpose. Placing trades based on such information would likely be considered insider trading if it yielded a profit, but this experiment has applications in looking for broader patterns that could be legal to trade or for identifying to what extent similar information could be used to predict stock price once it is publicly available and therefore legal to trade on.

The sample was placed into to a closed vector space, ξ over K where $K \subseteq \mathbb{R}^+$. $\left|\left|\cdot\right|\right|$ is the norm of ξ because $\left|\left|\cdot\right|\right|: \xi \to \mathbb{R}^+$ is a function according to the properties of a vector space.

The sample contains both discrete and continuous data. A number of vectors bisect ξ. Each of these bisecting vectors represent the stock price data of health insurance companies

that are publicly traded on the New York Stock Exchange: Humana (trading under ticker "HUM"), United Health Group (ticker "UNH"), CVS Health (ticker "CVS"), and Cigna Group (ticker "CI"). We will call these vectors $\vec{h}, \vec{u}, \vec{a}, and\ \vec{c}$ respectively, and they live inside corresponding closed vector spaces H, U, A, and C. The Universe completely contains all five vector spaces. Note that $\vec{h}, \vec{u}, \vec{a}, \vec{c} \in \mathbb{R}$. Figure 1 shows the vectors otherwise known as stock prices over the past 5 years.

Figure 1

Stock Prices of Publicly Traded Health Insurance Companies, Past 5 Years

In order to do exploratory data analysis, we will define a closed unit epsilon ball $\mathcal{B}$ to be the set:

$\mathcal{B}_n = \{x \in \xi: ||x|| \leq 1\}$, where ball $\mathcal{B}_n$ exists inside Universe Q.

And $H, U, A, C \cap \xi =$ Universe Q.

Therefore, in addition to cleaning the data, we must delete any data that does not fall inside Universe Q. Some of the data from Vector Space A will fall outside Universe Q because CVS did not finalize its purchase of Aetna Health Insurance until late 2019 (CVS Health Investor Relations, 2018). Therefore, the data from Space A must be culled to be completely inside Q.

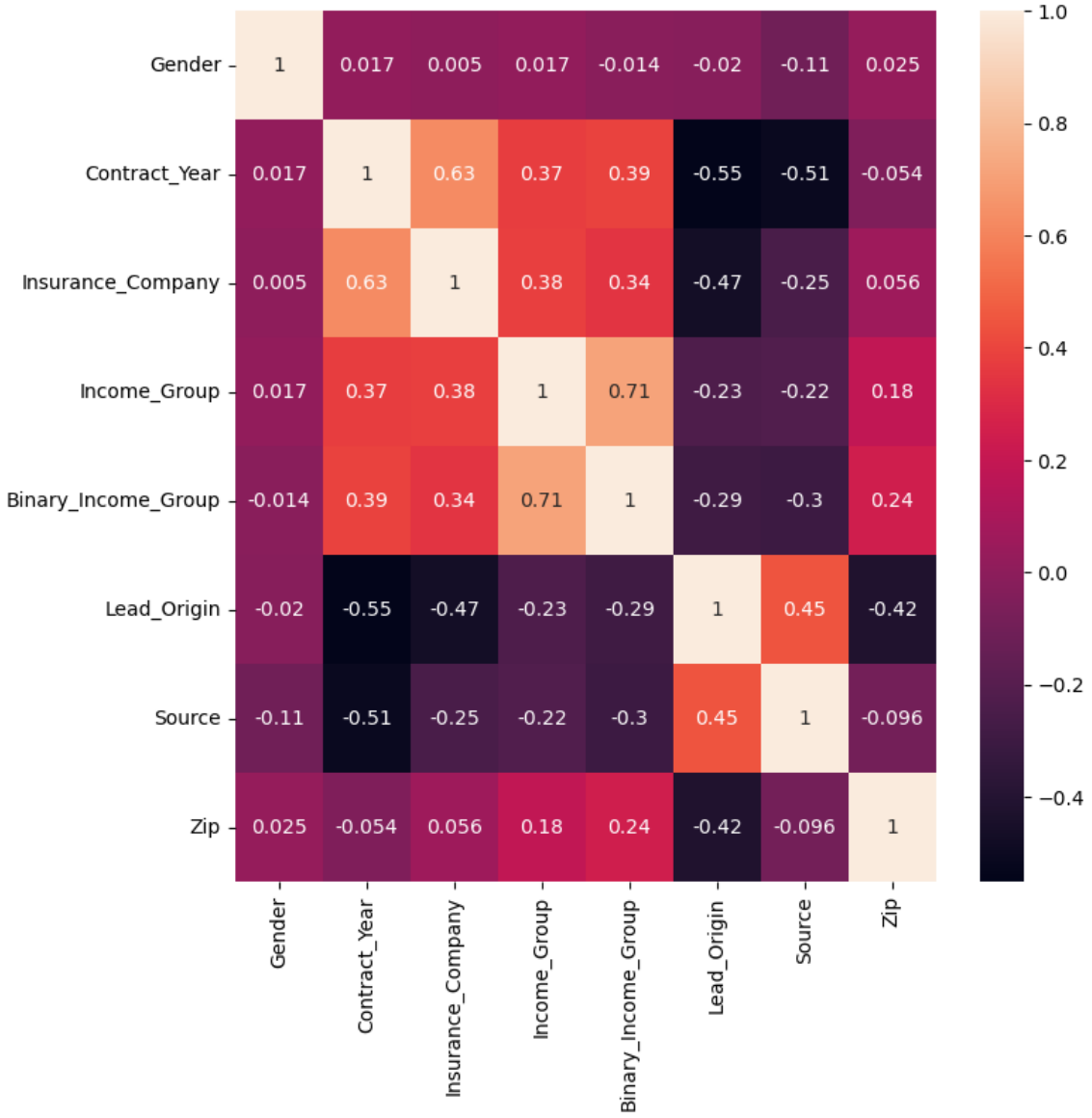
To achieve this, I deleted rows containing pricing data of CVS Health stock before it completed its purchase of Aetna. In addition, I deleted rows from the sales spreadsheet provided by the brokerage house that included sales of Aetna insurance before it was owned by CVS Health.

Looking for Correlations in the Data

With a newly cleaned dataset, I looked for potential correlations in ξ that could help me understand what I am analyzing. For this, I used the heatmap shown in Figure 2. The evidence of correlations could be useful to the health insurance brokerage that provided the data, as well as in answering my research question.

Figure 2

Correlations in ξ Exposed in the Data

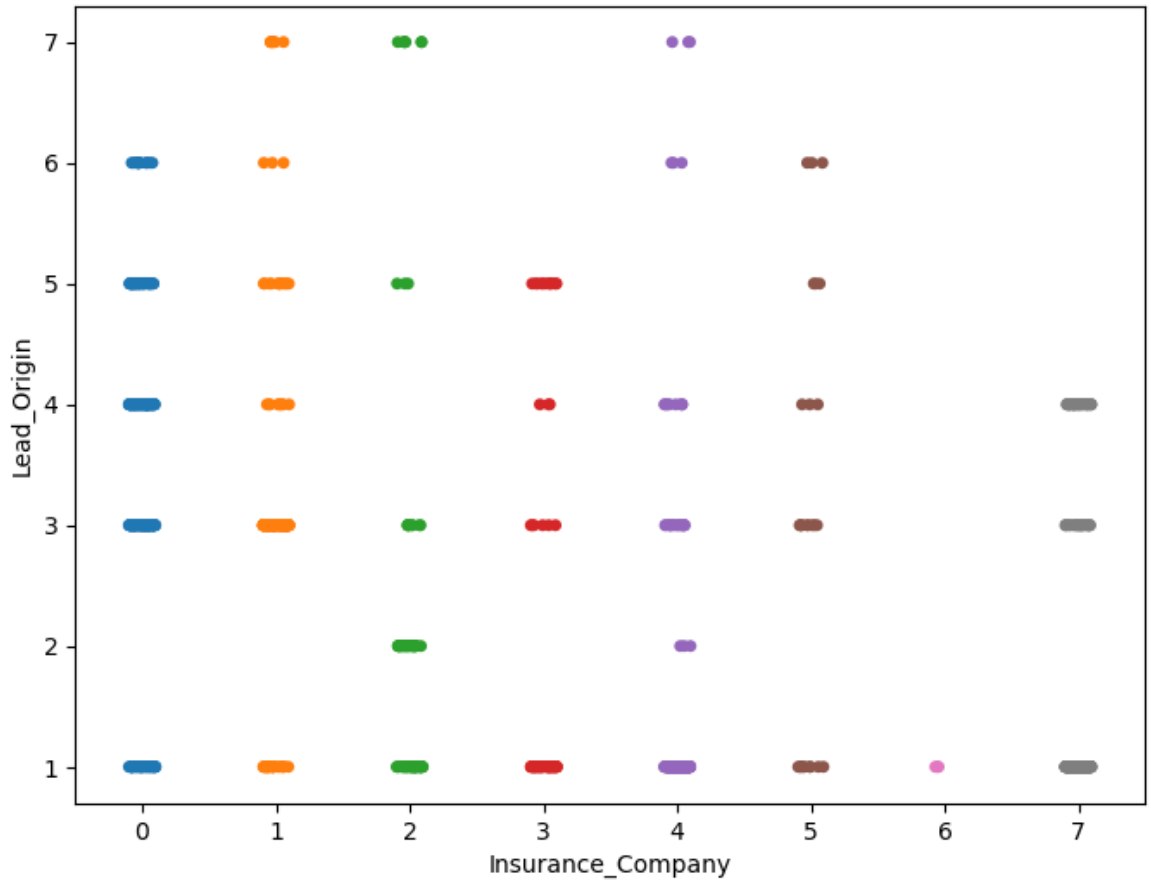


In addition to a heatmap, a strip plot was used to discover correlations in ξ. The strip plot provided in Figure 3 shows correlations in the insurance company whose product was sold by the brokerage and the origin of the referral. Specifically, the lead origin marked number 2 yielded sales for only two insurance companies, marked number 2 and number 4. According to

the data dictionary, lead origin number 2 is referrals from Kroger Pharmacies. Insurance company number 2 is Aetna, and insurance company number 4 is Humana.

Figure 3

A Comparison of Lead Origin and the Branding of the Insurance Policy, for Space ξ



In addition, insurance company number 6 originated only at the lead origin marked number 1 and nowhere else. Lead origin number 1 is referrals from pharmacists affiliated with Albertson's, and insurance company number 6 is Amerigroup, an affiliate of Anthem.

Evaluating Data Quality

In general, the data appears to be a representative sample despite that it originates in only one state: New Mexico. The data contains a uniform number of points representing females and males. It also contains people of various income groups, the majority of which are regular middle-class individuals. Some of the data represents poor individuals, and some represents rich individuals. The sample creates a different category for collected from veterans because veterans on Medicare can get health coverage from both Medicare and the Veterans Affairs system (U.S. Department of Veterans Affairs, 2023; AARP, 2023).

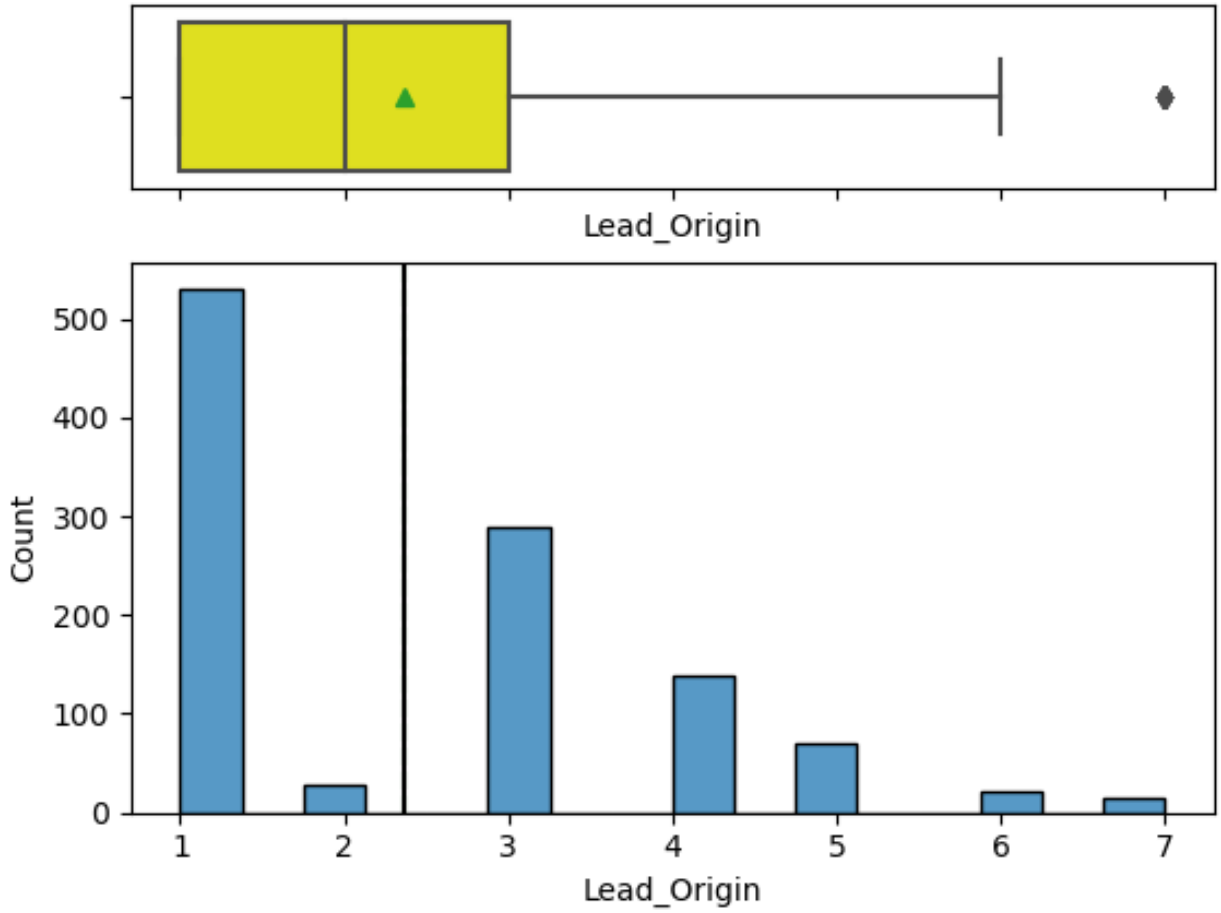
At first, I was concerned that retrieving data from only one state might prove problematic, since insurance is offered based on what state a person lives in. However, it appears that the larger publicly traded companies that are the subject of this research are available more-or-less nationwide, so one state's data could be an appropriate sample (Humana, 2023; Cigna, 2023; United Health, 2023; Aetna Medicare, 2023).

No dataset is perfect. This data contains some outlier points as well as some potential biases. Figure 4 box plot shows that there are outliers in column number 7, which is online sales from the independent brokerage company's website (so, not the website of the health insurance company or from Medicare.gov).

There was an outsized number of online sales made in 2023 compared to all other years combined. While the outliers appear clearly in the box plot, they are less evident in the histogram. I chose to leave this data in the sample instead of deleting it. Since our dataset contains discrete variables, a normal distribution would not be expected. I made no attempts to normalize the frequency distribution of the data.

Figure 4

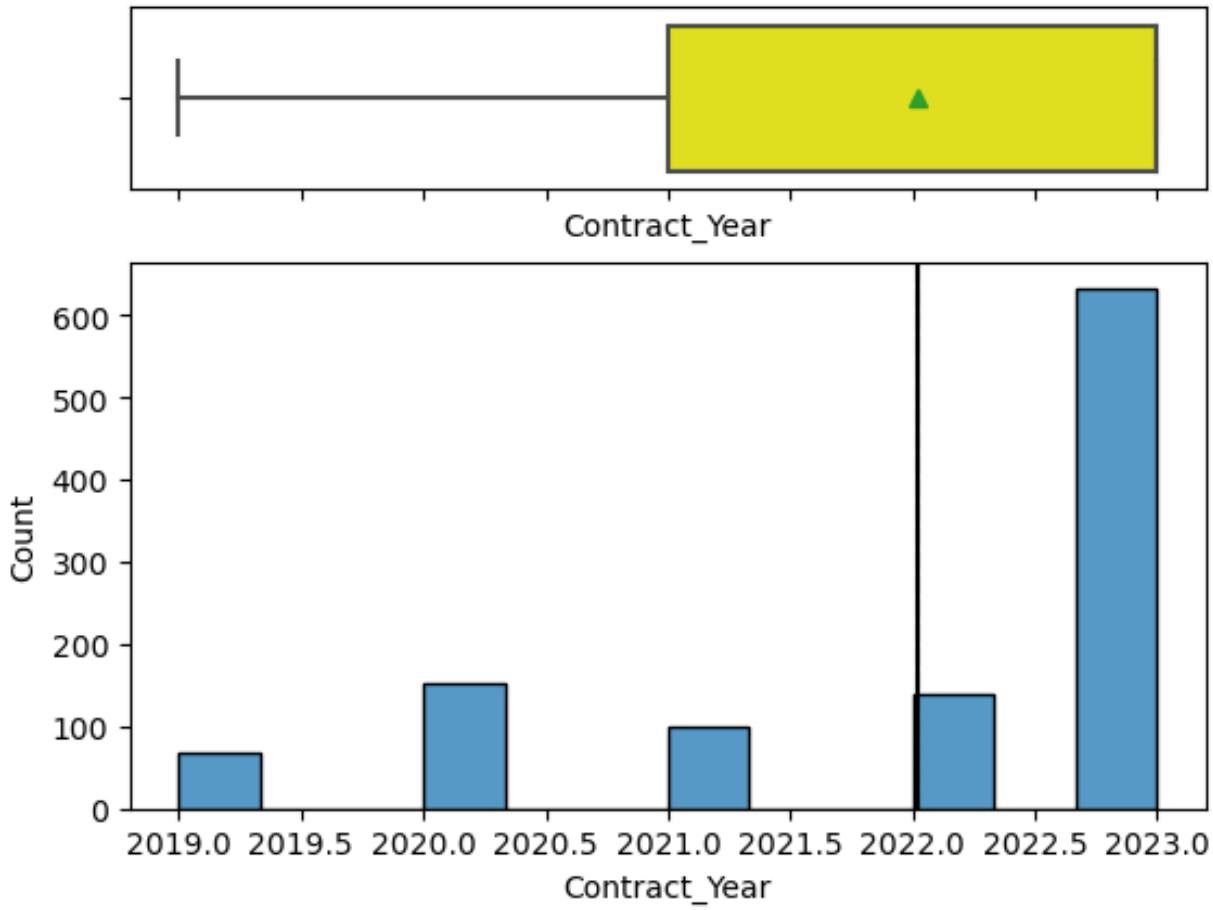
A Histogram and Box Plot of Referral Origins over the Past 5 Years, for Space ξ



One potential bias or issue with the data is that significantly more of it was collected in 2023 than in any other year in the sample, as shown in Figure 5. Data that doesn't contain a similar sample of sales data for each year may not be successful in predicting stock movements. Further exploration will show us if this dataset is problematic in this specific way or not. For right now, it doesn't appear to be a problem to have most of the datapoints originating in one single year, especially since CVS had already finalized its purchase of Aetna.

Figure 5

A Histogram and Box Plot of Sales Data from Past 5 Years, for Space ξ



Transforming the Data

Common sense says that many factors would influence a stock's movement over five years. Thus, in order to test the Medicare sales data specifically for predictability, I had to reduce the noise in the stock charts over that five-year timespan. Ideally, one would attempt to collect a larger sample to reduce noise in Universe Q. However, a bigger dataset isn't available.

Thus, I finally reveal the reason why I annoyingly decided to treat stock data like vectors or waves inside vector spaces: to identify noisy portions using wavelet transform. Unlike Fourier

transform, which provides information about frequencies of a signal but does not tell us when in time the signal occurred, wavelet transform reveals both the frequency of the oscillation and when it occurred. Since these vector spaces are not comprised of stationary waves and are in fact containing sales and stock price data, wavelet transform was applied to this problem (Boggess and Narcowich, 2001).

Doing so revealed what data was noisy in addition to when the noisy data occurred. Using a number of Haar scaling functions, I was able to define, attribute, and filter out noisy data using the following family of functions (Boggess and Narcowich, 2001; Nicoll, 2020):

$$\Phi(x) \in \mathbb{R} = \begin{cases} 2, & if\ 1 \leq x \\ 1, & if\ 0 \leq x < 1 \\ 0, & elsewhere \end{cases}$$

And

$$F_{(\tau,\frac{1}{\omega})} = \frac{1}{\sqrt{|\frac{1}{\omega}|}} \int_{-\infty}^{\infty} f(t)\Psi^* \left(\frac{t-\tau}{\frac{1}{\omega}} \right) dt$$

Where Φ is the scaling function, Ψ is the wavelet or basis function which acts as a window function, and F is the wave's function.

This function family helps identify different waves that could influence the stock price without identifying which frequency is associated specifically with Medicare sales data. (In addition, note that any waves with complex-valued frequencies will be dropped.) Based on the time period in which the experiment revealed abnormal price jumps, some frequencies were

identified to be most likely associated with the Covid-19 pandemic and were therefore labelled "noise" and dropped. February to May of 2020 was the noisiest portion for all the vector spaces studied. Of all four companies, CVS Health (Vector Space A) had the least noisy sample over the year 2020, but there was insufficient non-noisy data from before that portion to compare it with. Thus, that data could not be successfully transformed using this method.

After looking up when each insurance company reported their quarterly earnings, I could eliminate another frequency from each subset that I suspected was associated with earnings. By eliminating enough frequencies, I was left with some that could be associated with Medicare sales data.

Last, I had to measure the amplitude of any waves found to be affiliated with Medicare sales data to determine if they had enough amplitude to overcome the noise and change the stock price.

Conclusions

The data revealed some interesting results. I was unable to find any frequencies in Cigna's data that could be positively attributed to the past five Medicare enrollment periods, so the null hypothesis holds true for Vector Space C.

There were five frequencies present in United Healthcare's data over the past five enrollment periods that could potentially be attributed to Medicare open enrollment sales. I was unable to specifically target one of them over any of the others. It's also possible that none of them were associated with enrollment data. In any case, none of them had a large enough

amplitude to influence the stock price by themselves. Thus, the null hypothesis is true for Vector Space U.

Contrastingly, there was a frequency in Humana's data that was undeniably connected to Medicare enrollment data. Therefore, there is a strong predictability with the change in Humana stock price based on the success of the enrollment period due to a wave with a large amplitude. My hypothesis is true for Vector Space H. When Medicare enrollments were net positive for Humana (meaning that they had more enrollments than disenrollment), the stock price was observed to rise by at least 4 percent in the first quarter, and it remained net positive in a year-on-year basis for the remainder of the year. In the year when Medicare enrollments were net negative for Humana, the stock stayed more or less flat in price for the first three months of the year and fell slightly on a year-over-year basis. In other words, it was as if half of the wave function was missing (in other words, the function appeared to be partially continuous). This was interesting.

Further Research

If I were to continue research on this project, I would like to draw a larger sample and try the experiment again. I would want the new sample to contain data from multiple states as well as a longer period of time. An analysis of the marketing mix of the four insurance companies reveals that the dataset should be expanded in another way as well.

Humana, the only company in which an insider would come out net positive after placing a trade based on Medicare sales information, has significantly more Medicare customers than any other type (Humana Inc, 2023). Conversely, Cigna and United Healthcare

have more customers on worksite plans than Medicare plans (Cigna Group, 2023; United Health Group, 2023).

If I were to run this experiment again, I would include worksite insurance plans in the sample to see if a combination of enrollments when considering all of the insurance offerings could show predictability of future stock movement in more companies than just Humana. In other words, those many waves found in United Healthcare's stock data could be grouped if found to be appropriate, and wavelet transform could be used to see if the grouped waves could predict a future stock movement.

I would also be interested in expanding this research to look for any tradable correlations in publicly available pharmacy sales revenue from Kroger, Albertson's and other pharmacies in relation to predictability of price movements of health insurance stocks. As Figure 3 showed, the origin of the referral mattered in Vector Space ξ, which implied that pharmacy network and market share could potentially influence stock price.

References

AARP Organization. (February, 2023). *AARP Medicare Questions and Answers Tool.* Do I need to sign up for Medicare if I'm a veteran with VA health care? www.aarp.org/health/medicare-qa-tool/does-medicare-work-with-veterans-coverage/

Boggess, A., and Narcowich, F. (2001). A First Course in Wavelets with Fourier Analysis. Prentice Hall.

Cigna Group. (February, 2023). *United States Securities and Exchange Commission.* Form 8-K.

CVS Health Investor Relations (2018). *CVS Health Completes Acquisition of Aetna, Marking Start of Transforming Consumer Health Experience.* CVS Health Investor Relations. https://www.cvshealth.com/news/company-news/cvs-health-completes-acquisition-of-aetna-marking-start-of.html

Humana, Inc. (February, 2023). *United States Securities and Exchange Commission*. Form 8-K.

Medicare.gov. (No date). *Parts of Medicare.* www.medicare.gov/basics/get-started-with-medicare/medicare-basics/parts-of-medicare

Nicoll, A. (July, 2020). *The Wavelet Transform for Beginners: Signal Denoising and Quantum Magnetometry.* Warwick University, Department of Physics. Youtube.

U.S. Department of Veterans Affairs. (October, 2022). *VA Health Benefits.* VA health care and other insurance. www.va.gov/health-care/about-va-health-benefits/va-health-care-and-other-insurance/