Fast Food, Fast Growth: The Health Implications of Processed Diets

Team 18

Abstract

The relationship between diet, health, and economics is becoming increasingly evident in today's data-driven world. In America, over 2 in 5 adults are obese, largely due to a diet high in processed meats and sugars. This epidemic not only affects individual health but also strains the US healthcare system, consuming nearly $173 billion in medical expenditures in 2019 [19]. Therefore, our report explores the relationships between processed food consumption, obesity, and financial market dynamics. Using rigorous statistical tests and quantitative models, we analyze obesity levels in relation to the production and consumption rates of processed food components. A key highlight of our study is our innovative pairs selection and trading algorithm, which substantiates our hypothesis that stocks of companies producing or using processed food correlated with obesity also show strong correlations with healthcare sector stocks, and it provides a method to capitalize on this pattern. We obtain 7 pairs using this strategy for a Sharpe ratio of 1.58. We find that the strongest correlations are between processed food companies and healthcare sector stocks, highlighting the financial significance of health trends. To refine our hypothesis and trading strategy, we propose gathering additional evidence using SARIMA forecasting models to enhance predictive accuracy, offering actionable insights for stakeholders in both the food and healthcare sectors.

Contents

I Introduction 3

II Non-Technical Summary 3
- 2.1 Main Questions . . . 3
- 2.2 Key Findings . . . 4

III Technical Summary 4
- 3.1 Assumptions . . . 5
- 3.2 Data Exploration and Preprocessing . . . 5
  - 3.2.1 Meat Dataset Exploration . . . 5
  - 3.2.2 Sugar Dataset Exploration . . . 8
- 3.3 Finding Strong Correlations to Obesity . . . 10
  - 3.3.1 Obesity in the US . . . 10
  - 3.3.2 Broiler Meat and Obesity . . . 11
  - 3.3.3 Sugar and Obesity . . . 12
- 3.4 Obesity and the Healthcare Market . . . 12
  - 3.4.1 Correlation between Obesity and Drug Manufacturers . . . 13
  - 3.4.2 Correlation between Obesity and Medical Instrument Suppliers . . . 14
  - 3.4.3 Correlation between Obesity and Healthcare Planning . . . 15
  - 3.4.4 Overall Healthcare Market Impact . . . 15

IV Real World Application 16
4.1 Methodology 16

V Pairs Selection 16
5.1 Dimensionality Reduction by Principal Component Analysis 16
5.2 OPTICS Clustering 17
5.3 Absolute Rules of Disqualification (ARODs) 18
5.3.1 Johansen's Cointegration Test 18
5.3.2 Hurst Exponent of Spread 18
5.3.3 Half-Life of Spread 18
5.3.4 Number of Crossings with Mean for Spread 18
5.4 Summary of Pairs Selection 18

VI Pairs Trading Strategy 19
6.1 Results 20
6.2 Conclusion 21
6.2.1 Limitations 22
6.2.2 Future Directions 22

I Introduction

One of the most critical health concerns in American society in recent years has been the rising prevalence of obesity. This epidemic has rapidly grown over the past few decades, with significant portions of the population being affected. Between the years of 1994 and 2018, the prevalence of obesity in adults has nearly doubled, with about 43% of the population being classified as obese [23] [15]. Researchers have identified various factors contributing to this crisis, with processed food consumption being among the most prominent [14]. The widespread availability and consumption of high-calorie, low-nutrient foods have made it increasingly difficult for individuals to maintain a healthy weight. This report focuses on the impact of processed foods on obesity rates and explores the broader implications of this issue.

The growing obesity epidemic has far-reaching consequences, particularly in the realm of healthcare. The increased prevalence of obesity-related conditions, such as diabetes, cardiovascular diseases, and certain cancers, has led to a surge in healthcare costs and a greater demand for medical services [13]. This strain on the healthcare system has prompted significant investments in drug manufacturing and the development of new treatments and medical instruments to combat obesity and its associated diseases. Additionally, the market for weight management and obesity-related chronic disease products and services has expanded rapidly, reflecting the urgent need for effective solutions to address this pervasive health issue.

This study examines these trends using datasets from the Citadel Datathon and Yahoo Finance, which provide comprehensive time series information such as stocks for food prices, sugar prices, and obesity. This study focuses on how meat and sugar consumption patterns influence obesity rates and, consequently, healthcare outcomes.

II Non-Technical Summary

2.1 Main Questions

This study investigates two main questions:

1. How do processed foods, especially meats and sugar, influence adult and adolescent obesity?
2. How does obesity then influence healthcare sales in terms of drug manufacturing/retail, medical materials, and healthcare planning?

Below is this study's overarching hypothesis regarding these two questions:

Hypothesis: Stocks of companies producing or using components in processed foods that strongly correlate with obesity will also exhibit strong correlations with healthcare sector stocks. This is based on the hypothesis that processed foods contribute to higher obesity rates and health issues, which in turn lead to increased healthcare expenditure.

2.2 Key Findings

This study identified a recent trend of increased production and cold storage weight of broiler meat compared to other meats. Suspecting a link between the industrial growth of broiler meat and rising obesity rates, we found a strong positive correlation between these factors. Additionally, we investigated the influence of sugar on childhood obesity. Contrary to the well-held notion that sugary drinks contribute significantly to childhood obesity, their national popularity has shown a consistent decline. Our analysis established a negative correlation between the decline in sugar consumption and obesity rates. This led us to conclude that broiler meat plays a more dominant role in the rise of nationwide obesity.

Once the links between different processed components and their relation to obesity were established, the correlation between obesity and the healthcare sector was then examined. It was found that the growing rates of obesity are most strongly correlated with the pharmaceutical manufacturing, healthcare planning, and medical instruments subsectors.

Given these findings, an opportunity for statistical arbitrage via pairs trading between these two sectors of the food and beverages industry and the healthcare industry was further investigated. Pairs selection was conducted by several mathematical and machine learning algorithms to find the most viable pairs for their spreads to be traded. This eventually led to 7 pairs being chosen and an overall Sharpe ratio of 1.58.

Comparing initial processed food and healthcare analysis expectations with the selected pairs, the conclusions were proven to be partially true. While particularly strong correlations between fast-food stocks (industries that rely heavily on broiler meat) and the previously identified healthcare sub-sectors of drug manufacturing, healthcare planning, and medical instruments were found, strong correlations between the same healthcare stocks and sugar beverage industries were also found. While this seems to contradict the initial findings, with the strong negative correlation between sugar drink intake and obesity not being reflected in the positive correlation between sugar beverage stocks and healthcare stock, the increase in the sugar beverage stock can be attributed to other factors.

Overall, our report serves as a valuable guide on how medical relationships involving processed food components can influence various financial markets and sectors. In addition, this report provides multiple statistically developed quantitative models, modeling obesity levels using production and consumption rates of processed food components. Our key contribution, however, lies in our innovative pairs selection and trading algorithm. This algorithm not only confirms our initial hypothesis that stocks of companies producing or using components in processed foods strongly correlated with obesity will also show strong correlations with healthcare sector stocks but also offers a method to capitalize on this observed pattern. Given more time, we would enhance our hypothesis and trading strategy by gathering additional evidence using SARIMA forecasting models.

III Technical Summary

We begin our technical summary by listing key assumptions and generalizations used in our analysis.

3.1 Assumptions

- We limit our analysis to only United States production, health, and financial statistics. This data is most readily available online and in the datasets provided by the Datathon. We acknowledge that our findings may deviate from what is expected on a global scale due to differences in culture, health regulations, and economic and technological advancements.
- We guide our analysis based on overall national statistics. Particularly when analyzing data from the CDC Behavioral Risk Factor Surveillance System [16], we only focus on overall nationwide data. While we acknowledge that health and consumption trends may vary on a state-by-state basis across gender, race, and age, generalized national trends are more easily understandable and useful when making general policy changes or financial decisions.
- When we discuss processed foods, we focus our investigation on industrial meats and processed sugar, as these are the dominating components of most processed foods in the American diet [9].

3.2 Data Exploration and Preprocessing

3.2.1 Meat Dataset Exploration

We began by analyzing meat and livestock datasets [22]. Our initial exploration analyzed the following datasets in the provided order:

1. Red Meat and Poultry Production
2. Cold Storage

We didn't consider the Slaughter Counts dataset, as they do not indicate anything novel about changes relevant to consumer demands and experiences with the meat processing industry. The Red Meat and Poultry Production dataset conveys similar data to the Slaughter Counts; however, as meat is sold by weight and not head, the Red Meat and Poultry Production dataset produced more accurate information regarding meats sold to consumers.

We exclude the Slaughter Weights dataset from our study because, as we will discuss, we ultimately focus our attention on broiler meat. The data on broiler meat provided by the Slaughter Weights dataset is both insufficient and uninteresting compared to the insights gained from the first two.

When analyzing the Red Meat and Poultry Production data, naive plotting revealed three key pieces of information:

1. Almost 95% of US meat production consists of broiler meat, beef, or pork
2. There is no clear difference in the production of federally inspected and commercial meat
3. Meat production contains seasonality

Identifying these properties, we apply several transformations to clearly analyze trends in meat production.

First, we noted that 94.4% of US meat production is comprised of broiler meat, beef, and pork.

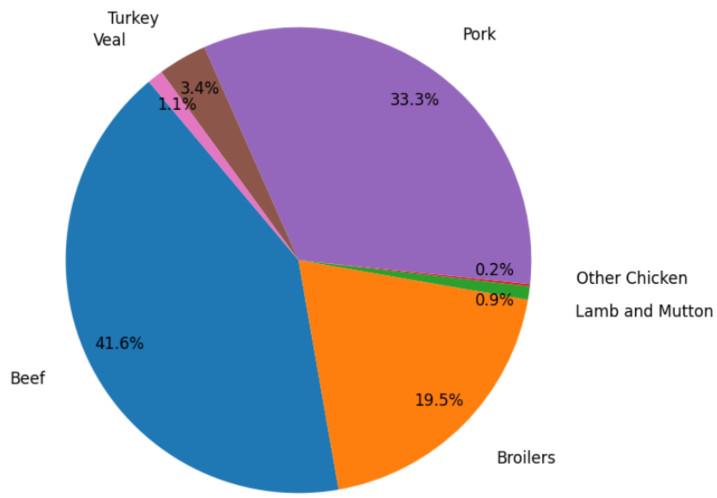

Figure 1: Cumulative Meat Production contribution by animal from 1985-2022

Therefore we omit less popular meats such as turkey, veal, lamb, mutton, and other chicken from our analysis. Next, we disregarded the distinction between "federally inspected" and "commercial" features, as this did not significantly affect meat production outcomes. Finally, we applied Seasonal-Trend Decomposition using Loess (STL) to break down the time series data into trend, seasonal, and residual components. We opted for STL due to its customizable smoothing and its ability to clearly and accurately visualize patterns in the data through decomposition into trend, seasonal, and residual components. When implementing STL, we set the seasonal period to be one year [1].

The trendline plot reveals a significant increase in broiler meat production. Broiler meat refers to chicken specifically raised for meat production, characterized by fast turnover and high yield, which boosts company profits. The data shows that broiler meat production surpassed beef between 1995 and 2000. This trend aligns with business reports indicating that chicken consumption has generally risen each year since 2000. Additionally, the broiler production industry is highly vertically integrated, allowing for rapid scaling of production [20]. Interestingly enough, this period was also when the World Health Organization first began to report obesity as a growing public health issue [24] .

When plotting and analyzing data, we only load the time series after 1985 as the data before then is incomplete. The seasonality and residual plots, while developed in code, didn't display any significant patterns and will not be mentioned in further discussion.

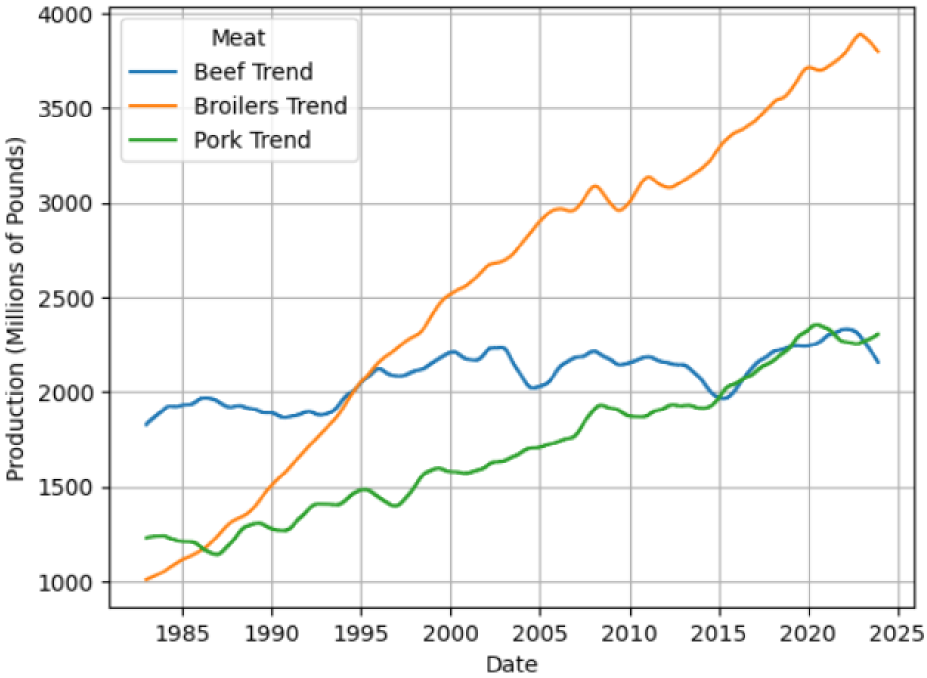

Figure 2: Smoothened Meat Production Data from 1985-2022

Next, we investigated the cold storage dataset. Similar to when analyzing meat production we identified a couple of key properties:

1. Most of cold storage consists of broiler meat, beef, and pork. Interestingly, there is also significant cold storage dedicated to turkey.
2. The cold storage exhibits seasonal trends.

Although turkey makes up a significant percentage of the cold storage data set, we chose to continue to focus primarily on broiler meat, beef, and pork data.

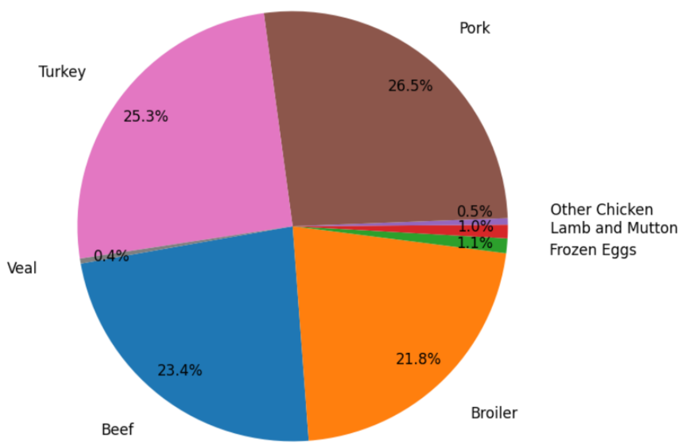

Figure 3: Cumulative Meat Production contribution by animal from 2004-2022

We also apply STL decomposition and plot the corresponding trendlines to identify patterns in the cold storage time series. When developing trendlines, we set the STL seasonal period to two years due to the fact that cold typically has longer seasonal periods than meat production [18].

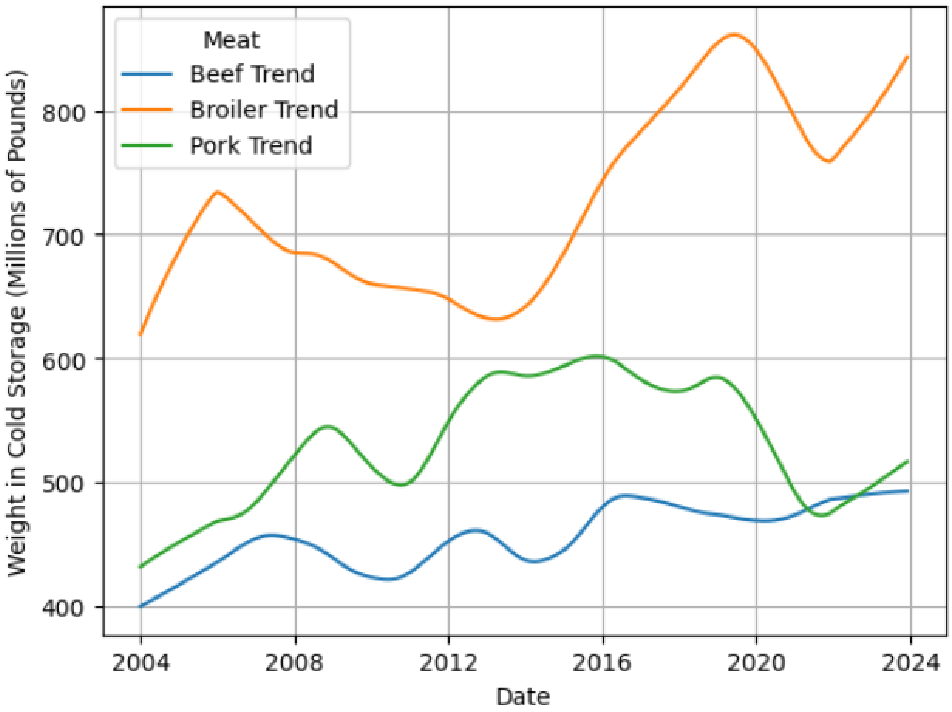

Figure 4: Smoothened Meat Storage Data from 2004-2022

When plotting and analyzing data, we only load the time series after 2004 as the data before then is incomplete. The seasonality and residual plots did not display any significant patterns and will not be discussed further.

The key pattern observed was a significant increase in the volume of broiler meat in cold storage, indicating a rise in supply. According to macroeconomic principles, higher cold storage volumes suggest increased meat supply, which can lead to lower prices and higher consumption of broiler meat. This increase in consumption could potentially be linked to the growing obesity rates. While this observation does not definitively prove a correlation between broiler meat consumption and obesity, it, combined with our previous finding that obesity was recognized as a health crisis around the same time broiler meat became the dominant meat source, warrants further investigation into the relationship between broiler meat consumption and obesity.

3.2.2 Sugar Dataset Exploration

When analyzing the sugar-related data, we considered two datasets:

1. The Commodities dataset which includes the historical cents per pound price of sugar
2. The Nutrition Physical Activity and Obesity Data which includes the national historical data for youth consumption of sugar drinks

Given that the Nutrition, Physical Activity, and Obesity Data provides annual information on youth sugar beverage consumption from 2007 to 2019, while the Commodities dataset offers monthly data from 1990 to 2024, we need to filter and smooth the Commodities dataset to enable accurate comparisons.

When plotting the sugar commodity time series data, there are no apparent trends. Like most commodities, it has seasonal highs and lows.

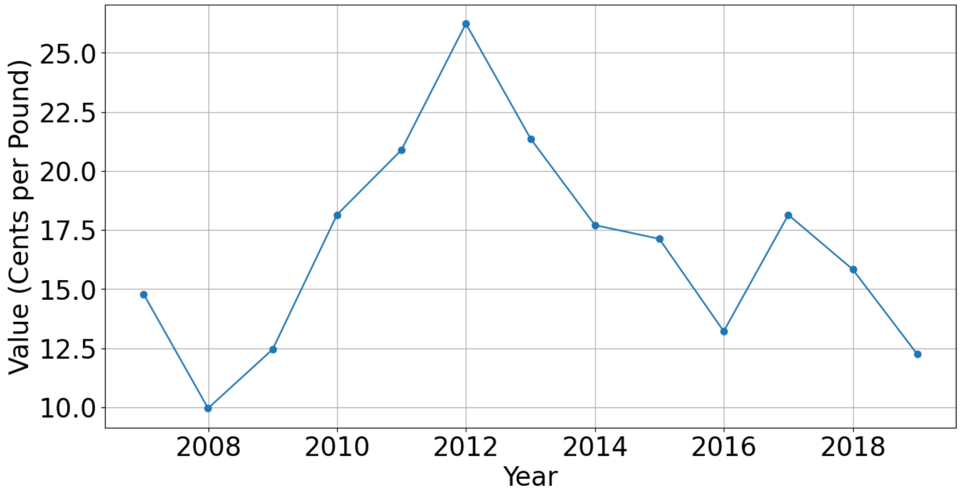

Figure 5: Sugar Commodity Prices from 2007 to 2019

We similarly plot the student sugar beverage consumption time series data.

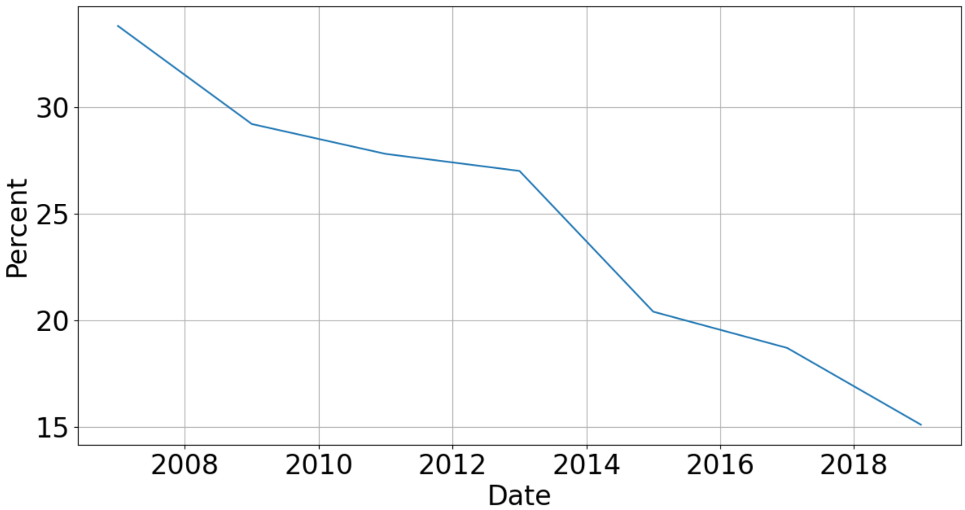

Figure 6: Percent of students in grades 9-12 who drank regular soda/pop at least one time per day from 2007 to 2019

The student sugar beverage consumption plot shows a very clear decline in the consumption of sugary drinks. This finding is surprising due to the fact that sugary drinks have been frequently blamed by the media for the growth of obesity [6] and have scientific links to illnesses like diabetes [3]. Although shocking, Harvard's School of Public Health confirms this finding reporting that the percentage of children who were heavy consumers declined from 11% to 3% from 2003 to 2016. [11]

While this decline in consumption could be attributed to the variable cost of the sugar commodity, we statistically disprove this claim by computing the Pearson correlation coefficient between the sugar commodity price and the decline in sugary soft drink consumption. We obtain a Pearson correlation coefficient of -0.276, indicating a weak linear correlation, making it unlikely that the commodity price is the driving influence for the lower soft drink consumption rates.

Given these preliminary findings, we decide to further investigate the correlations between broiler meats and sugar with rising obesity levels.

We develop the following sub-hypotheses:

1. **Sub-Hypothesis 1:** There is a positive correlation between broiler meat production and rising obesity levels.
2. **Sub-Hypothesis 2:** There is a negative correlation between sugar beverage consumption and childhood obesity.

3.3 Finding Strong Correlations to Obesity

3.3.1 Obesity in the US

We begin by plotting overall nationwide obesity rates. The Nutrition, Physical Activity, and Obesity Data categorizes obesity based on whether the subjects are adolescents (in grades 9-12) or adults (above 18 years), so we make a similar distinction in our analysis. Additionally, the time ranges available for each subset of data (adolescent or adult obesity) differ, necessitating careful selection of appropriate intervals for comparison with broiler meat production and sugar consumption data.

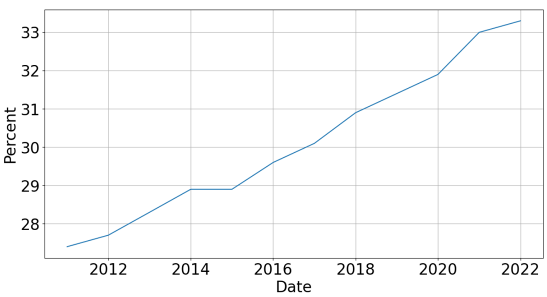

Figure 7: Percent of Adults who are Obese from 2011 to 2022

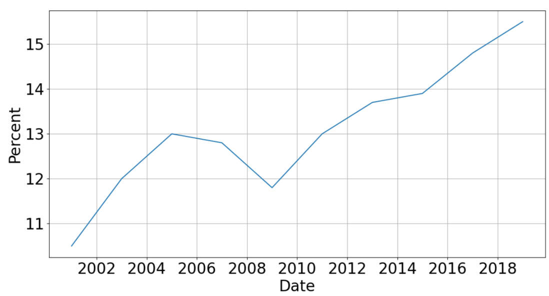

Figure 8: Percent of Adolescents who are Obese from 2001 to 2019

Figure 9: Overall National Obesity Rates split across Age

We truncate the adult obesity plot to include data from 2011 to 2022 and the child obesity plot to include data from 2001 to 2019 because these are the ranges with usable data.

3.3.2 Broiler Meat and Obesity

Both plots display a steady linear increase in obesity starting from the early 2000s. To identify the correlation between the broiler meat time series and each obesity time series, we first filter the broiler meat data to match the appropriate time periods and then smooth it to include only yearly data. Next, we calculate the Pearson correlation coefficient between broiler meat production and each obesity time series.

From these calculations, we obtain Pearson correlation coefficients of 0.872 and 0.919 for the correlations between broiler meat production and adolescent and adult obesity rates, respectively. These high correlations indicate a strong positive relationship, allowing us to display the best-fit regression lines relating broiler meat production to each obesity rate.

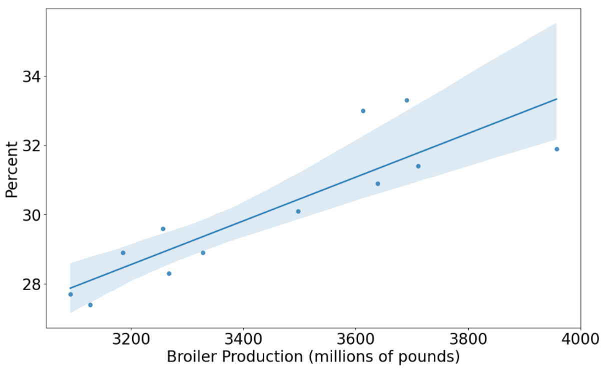

Figure 10: Relationship between Broiler Production and Adult Obesity

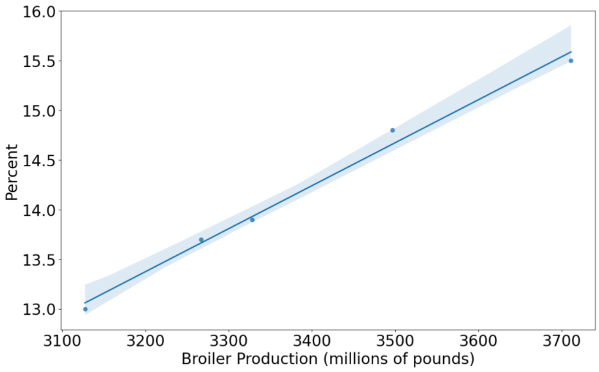

Figure 11: Relationship between Broiler Production and Child Obesity

Figure 12: Regression Lines plotting the Relationship between Broiler Production and different types of Obesity

As we show, there is a very strong positive linear correlation between broiler production and obesity rates. This correlation can be captured using the following equations:

$$\text{Adult Obesity:} \quad y = 0.00630x + 8.39409 \tag{1}$$
$$\text{Child Obesity:} \quad y = 0.00433x - 0.48121 \tag{2}$$

Figure 13: Regression Lines between Broiler Production and Obesity

Note that in figures 1 and 2 above x is production weight in millions of pounds, and y is the national percentage. This finding confirms our first sub-hypothesis that broiler meat production is positively correlated with obesity. This is also confirmed by current research which finds that broiler meat has very high fat levels, and processed chicken products often contain high levels of sodium and other additives that can negatively impact health [21].

3.3.3 Sugar and Obesity

Similar to when trying to find correlations between broiler meat and obesity, we must consider the format of the sugar consumption data from the Nutrition, Physical Activity, and Obesity Data. The sugar consumption data is formatted to only biannual intervals from 2007 to 2019. Furthermore, this data only considers adolescent participants. Thus to appropriately find the correlation between sugar consumption and obesity, we must filter our obesity data to only consider adolescent obesity and smooth it to biannual intervals. Applying these transformations, we observe a Pearson correlation coefficient of -0.878 which indicates a strong negative correlation.

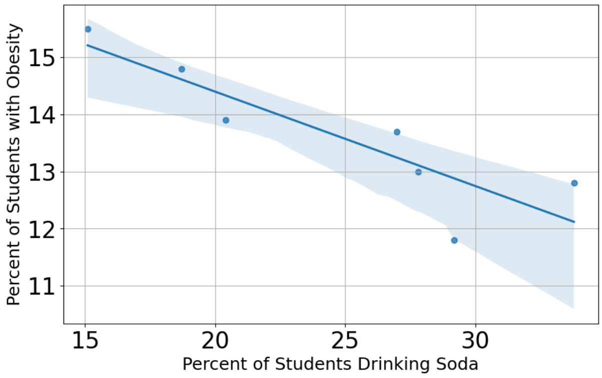

Figure 14: Relationship between Sugar Consumption and Child Obesity

This finding confirms our second hypothesis that there is a strong negative correlation between sugar beverage consumption and childhood obesity. We once again can quantify the relation with the following equation:

Drinking Soda with Obesity:

$$y = -4.66042x + 88.15288$$

Figure 15: Regression Line between Soft Drink Consumption and Obesity

In this case, x is the percent of students drinking soda or other soft drinks. This result is also supported by research, which consistently identifies sugary drinks as a major contributor to obesity [10]. Therefore, a reduction in sugary drink consumption is expected to lead to a decrease in obesity rates.

Now that we have confirmed our two sub-hypotheses, we will confirm whether these trends are present in financial markets.

3.4 Obesity and the Healthcare Market

Given that there have been correlations between broiler meat and sugary drinks with obesity, now we would like to see if there are correlations between obesity and the healthcare market. This is all to achieve our end goal of identifying food stocks utilizing processed meats (particularly broiler meat) and processed sugar that is strongly correlated with healthcare stocks.

To conduct this analysis, we load the obesity data from the mentioned in Nutrition Physical Activity and Obesity Data and we load healthcare financial data from Yahoo Finance. As the obesity data was split amongst adolescents and adults, we chose to solely focus on the adult obesity data. This was for two reasons:

1. In many healthcare markets, adults are the primary consumers due to the prevalence of chronic diseases, preventive care, and the need for routine medical services
2. There was more recent data regarding adult obesity rates

To make appropriate comparisons with the obesity data, we first smooth the daily healthcare financial data by aggregating it into annual intervals. We then account for a standard three-year response lag between changes in obesity rates and healthcare responses. Finally, we filter the financial data to match the obesity 11-year time window. While we evaluated Pearson correlation coefficients between obesity and all sectors of healthcare, we only found moderate correlations in the following sectors. We will focus on these sectors in our discussion and exclude the others.

Table 1: Pearson Correlation with Each Sector

Sector	Correlation Coefficient
Drug Manufacturing	0.620
Medical Instruments	0.534
Healthcare Planning	0.511

While these correlations are not as strong as those observed between processed ingredients and obesity, they still indicate a moderate relationship. This is understandable, given that obesity-related treatments and responses are only a portion of the products and services within each sector. We will discuss the research findings and conclusions related to these sector-obesity correlations in the following subsections.

3.4.1 Correlation between Obesity and Drug Manufacturers

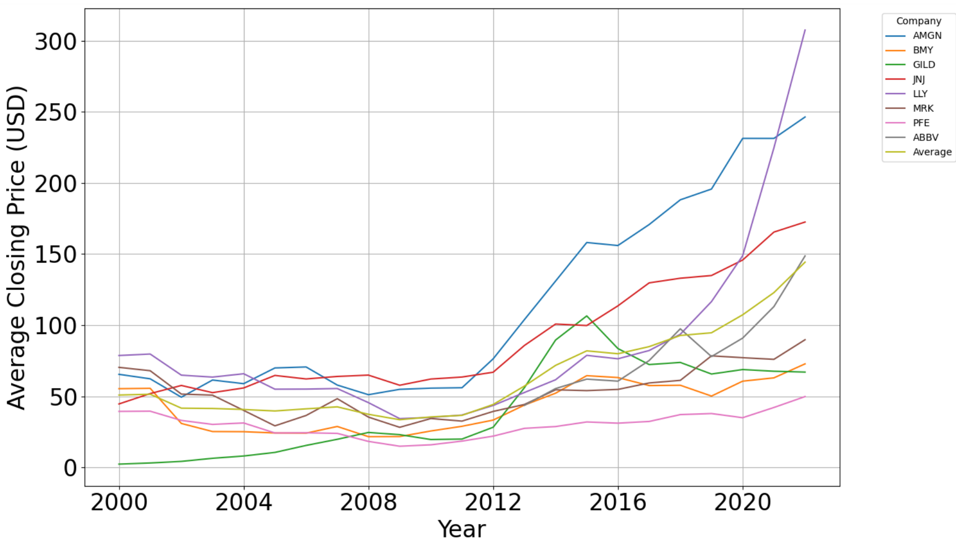

Figure 16: Annually Average Drug Manufacturer Stock Prices

The rise in obesity rates has significantly influenced the pharmaceutical industry, leading to an increase in drug manufacturing, specifically targeting obesity-related conditions. According to a study by the World Obesity Federation, the pharmaceutical industry has seen a surge in the development of medications aimed at treating comorbidities associated with obesity, such as type 2 diabetes, hypertension, and hyperlipidemia [8]. This trend emphasizes the demand for medical interventions that can manage the complex health issues arising from obesity, prompting pharmaceutical companies to invest heavily in the research and production of relevant drugs. In addition, the correlation between obesity and drug manufacturing is evident as the rising prevalence of obesity drives demand for medications addressing related health complications. A study by the Canadian Medical Association Journal highlights that obesity-related diseases account for a substantial portion of the global pharmaceutical market, with an increasing number of drugs being developed and marketed for these conditions [2]. This demand incentivizes drug manufacturers to focus their research and development efforts on creating effective treatments for obesity and its associated ailments.

3.4.2 Correlation between Obesity and Medical Instrument Suppliers

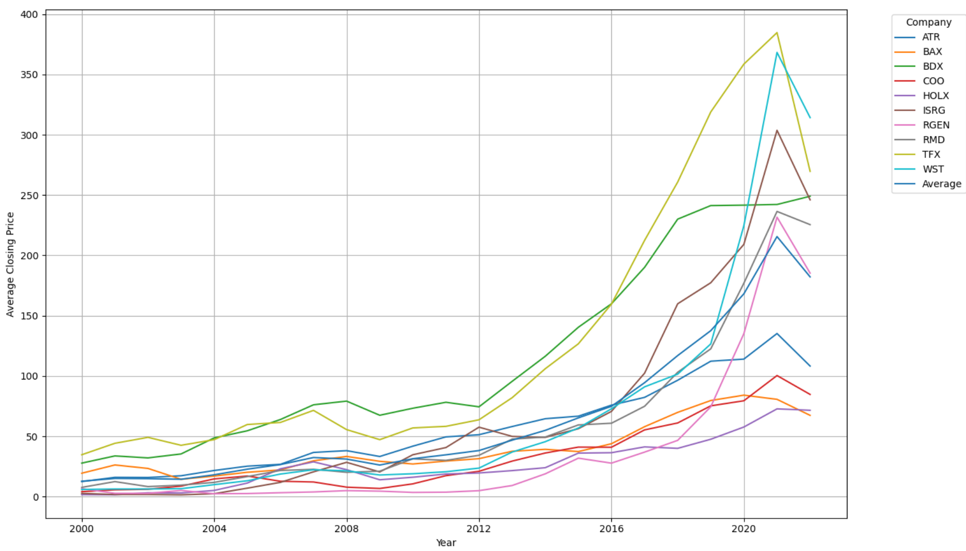

Figure 17: Annually Average Medical Instrument Supplier Stock Prices

The surge in obesity rates also affects the design and utilization of medical instruments. A study by Boston University School of Medicine and Boston Medical Center emphasizes the need for specialized medical equipment to accommodate obese patients, such as bariatric surgery tools, reinforced hospital beds, and imaging devices capable of handling higher body weights [7]. The increasing demand for these instruments reflects the growing recognition of obesity as a critical factor in healthcare provision, requiring adaptations in medical technology to ensure accurate diagnosis and effective treatment of obese patients.

3.4.3 Correlation between Obesity and Healthcare Planning

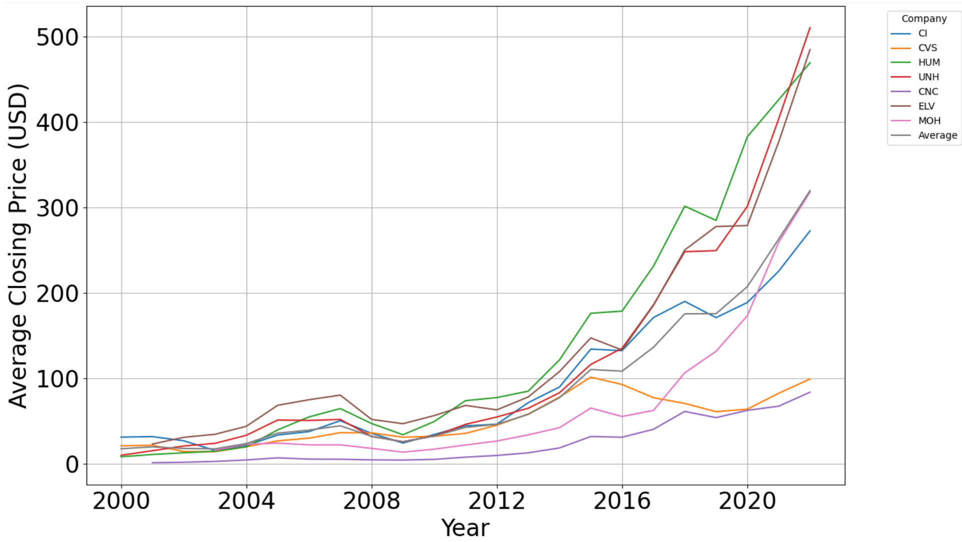

Figure 18: Annually Average Healthcare Planning Stock Prices

Obesity significantly impacts healthcare planning, necessitating strategic adjustments to address the growing burden of obesity-related health issues. According to the World Health Organization, healthcare systems must adapt to manage the increased prevalence of chronic diseases linked to obesity, such as cardiovascular disease and diabetes [23]. This requires comprehensive planning, including resource allocation, development of prevention programs, and integration of obesity management into primary care practices. Healthcare planners must consider the long-term implications of obesity on healthcare infrastructure and ensure that adequate services and support systems are in place.

3.4.4 Overall Healthcare Market Impact

Overall the rise in obesity rates has led to a significant increase in healthcare expenditure, with obesity-related conditions accounting for a substantial portion of overall healthcare costs. According to a study by the Public Health Economics Program, medical spending attributable to obesity accounted for 9.1% of total annual medical expenditures in the United States, with obese individuals incurring significantly higher medical costs compared to their non-obese counterparts [4]. This is because obesity is associated with higher medical expenses due to the need for frequent and intensive medical interventions to manage comorbidities such as diabetes, cardiovascular diseases, and certain cancers. This increase in expenditure highlights the economic burden that obesity places on healthcare systems, necessitating comprehensive strategies to mitigate these costs and reduce obesity among our populations.

Given the scientifically and statistically established link between obesity and the healthcare market, we have confidence in our core hypothesis: stocks of companies producing or using components in processed foods that strongly correlate with obesity will also show strong correlations with healthcare sector stocks. To operationalize these observations quantitatively, we have developed a pairs trading strategy.

IV Real World Application

It is a common strategy to execute pairs trading on correlated or cointegrated stocks that are in the same industry due to their tendency to trend in the same direction. However, our earlier analyses have shown how there are correlations between processed food and underlying health effects as well as correlations between these health effects to healthcare needs and expenditure. Therefore, there is a possibility that these two markets are closely related and thus, stocks from both industries could be considered for pairs trading. It will also be more likely to be profitable if pairs are not commonly traded. This section conducts pairs selection and implements a simple pairs trading strategy.

4.1 Methodology

The stock universe used were some stocks from the healthcare sector (drug manufacturers, healthcare plans companies, medical devices, medical instruments and supplies, medical care facilities, medical distribution) obtained from Yahoo Finance as well as the stocks from the food and beverages sector as provided.

The trading period was from 2002-12-01 to 2024-02-29. Stock splits were accounted for and non-trading days were dropped. A fixed size sizing method (100 exposure size) was employed for simplicity.

V Pairs Selection

For the purpose of this study, pairs were only considered if they had one ticker from the healthcare industry (obtained from Yahoo Finance) and one from the foods and beverages industry (provided dataset).

In order to prevent high computational costs, this study suggests an effective pre-partitioning of the considered asset universe to reduce the number of feasible combinations. This dramatically reduces the number of necessary statistical tests, consequently decreasing the likelihood of finding spurious relations.

This employs a three-pronged approach as mentioned in an existing paper by Sarmento and Horta in 2020 [12] which comprises:

1. Dimensionality reduction - find a compact representation for each stock
2. Unsupervised Learning - apply an appropriate clustering algorithm (OPTICS)
3. Select pairs - define Absolute Rules of Disqualification (ARODS) to select pairs for trading

5.1 Dimensionality Reduction by Principal Component Analysis

The stock data of high dimensionality is first reduced into a meaningful representation of reduced dimensionality that still corresponds to the intrinsic dimensionality of the data. This is done so by the means of Principal Component Analysis (PCA) and the minimum number of parameters is chosen such that it accounts for the observed properties of the data.

Profitable pairs chosen should have the same risk exposure and any deviations from the expected return can be regarded as mispricings and an opportunity to place trades. Using PCA on the return series to extract the common underlying risk factors for each stock has been used in past research by Avellaneda and Lee in 2010 [5].

PCA is a statistical procedure that transforms a set of observations of possible correlated variables orthogonally into a set of linearly uncorrelated variables, called principal components. Principal components account for a decreasing amount of variance and each succeeding component has the highest variance possible under the constraint that it must be orthogonal to preceding components.

This study chose the number of principal components to be the minimum possible number that accounted for at least 95% of the properties of the data.

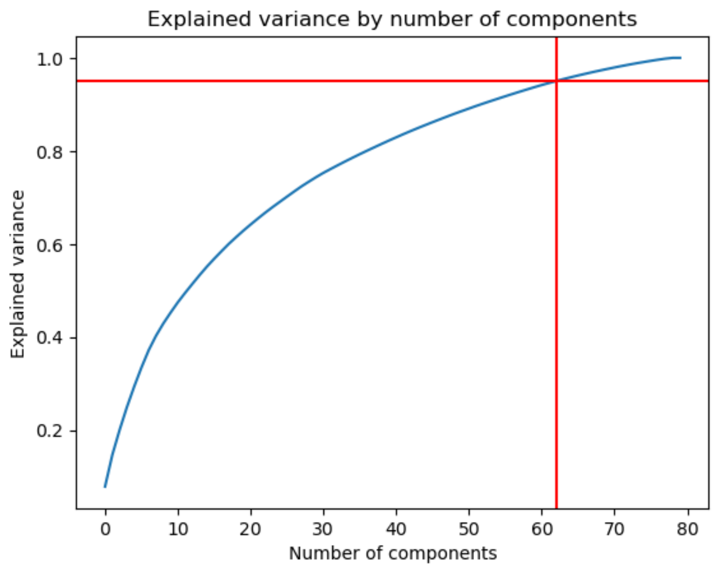

Figure 19: Explained Variance against Number of Principal Components

5.2 OPTICS Clustering

To find closely related stocks, the principal components obtained from the PCA above were then used in an unsupervised learning clustering technique, OPTICS (Ordering Points to Identify the Clustering Structure). This technique finds core samples of high density and expands clusters from them.

Compared to another clustering technique DBSCAN, it can detect meaningful clusters in data of varying density. It does so by linearly ordering the points in the data such that spatially closest points are neighbors in the ordering. Additionally, a special distance is stored for each point that represents the density that must be accepted for a cluster so that both points belong to the same cluster.

After the closely related stocks were produced, pairs were generated where one ticker from the healthcare industry and one from the foods and beverages industry.

5.3 Absolute Rules of Disqualification (ARODs)

In order to find appropriate candidate pairs and select those that will be most suitable for pairs trading, a set of conditions are defined and ran on price series of stock data. The most common approaches to select pairs are distance, correlation and cointegration. The cointegration approach was selected due to existing literature that suggests it performs best.

5.3.1 Johansen's Cointegration Test

First, the pairs were put through Johansen's cointegration test at a 90% significance level. This tests for cointegration of the price time series data and if there is a long-run equilibrium relationship between them. Using the trace statistic, the test evaluates the null hypothesis that there are at most r cointegrating relationships against the alternative of more than r. This study sets r to 0 as it aims to accept pairs that have any number of cointegrating relationships between them.

5.3.2 Hurst Exponent of Spread

Second, a validation step was implemented to ensure the mean-reversion characteristics of each pair's spread. This study imposes a condition that the Hurst exponent associated with the spread is less than 0.5, indicating that the process is mean-reverting.

5.3.3 Half-Life of Spread

Third, the pair's spread movement is put through another constraint relating to the half-life of the mean-reverting process. Half-life is the time that the spread will take to mean-revert half of its distance after having diverged from the mean of the spread, given a historical window of data. This study aims to capture medium-term price movements and hence spreads that have half-lives of either very short ($<$ 1 day) or very long ($>$ 252 days) periods are not suitable.

5.3.4 Number of Crossings with Mean for Spread

Lastly, every spread is checked to make sure it crosses its mean at least twelve times a year, to provide enough liquidity and thus providing enough opportunities to exit a position.

5.4 Summary of Pairs Selection

After these clustering techniques and ARODs, 7 pairs were chosen. The following pairs trading strategy focuses on the following pairs:

('JNJ', 'PEP'), ('JNJ', 'SBUX'), ('JNJ', 'YUM'), ('ABT', 'SBUX'), ('ABT', 'YUM'), ('SYK', 'MCD'), ('ZBH', 'SBUX')

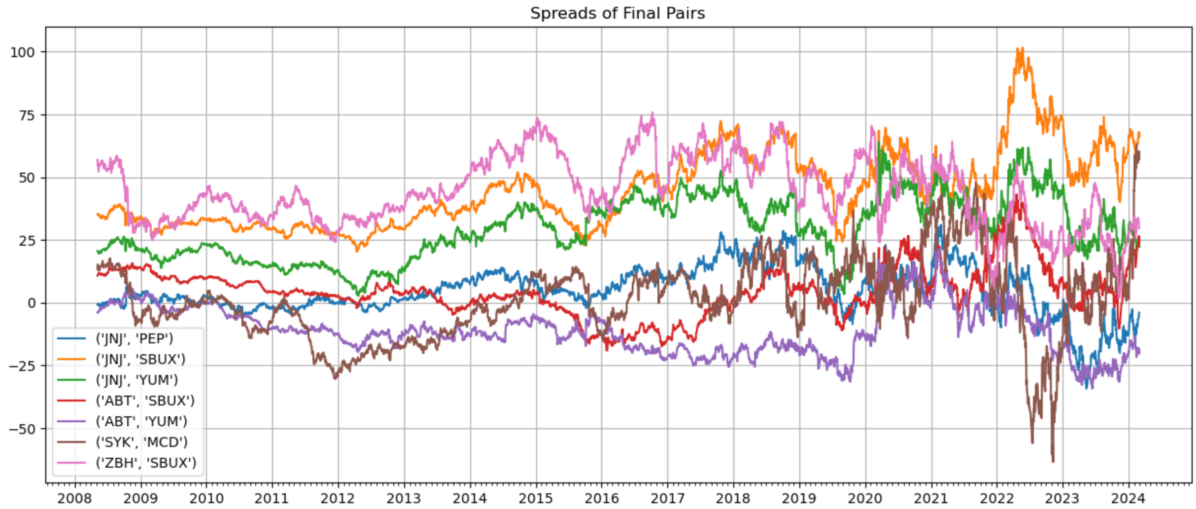

Figure 20: Spreads of Final Pairs Used in Trading Strategy

We can observe that these stock pairs somewhat confirm our hypothesis. Note that all the healthcare stocks mentioned fall into specific categories: drug manufacturers (e.g. JNJ), medical instrument suppliers (e.g. ABT, SYK), or healthcare planners (e.g. ZBH). Additionally, we find significant correlations with stocks and ETFs related to processed meats (e.g. MCD and YUM).

A surprising finding is the success and correlation of sugar beverage stocks like PEP and SBUX, despite the overall decline in soft drink consumption. This can be explained by the fact that these key sugar beverage brands, such as Pepsi (PEP) and Starbucks (SBUX), are large global players not solely influenced by American consumer dietary habits. Moreover, many of these major brands have diversified their product lines to include lower sugar and sugar-free options, which could account for their continued market expansion despite a reduction in sugar consumption [17].

VI Pairs Trading Strategy

Given the pairs obtained above, this study implements a simple pairs trading strategy where the spread between the two stocks in each pair is traded.

First, a linear regression was ran between the price series of the two stocks in each pair and their residuals were obtained. These residuals, which represent the spread between the two stocks, can be modeled as a Ornstein-Uhlenbeck (OU) process, which is a type of continuous-time stochastic process used to model mean-reverting behavior, as the spread was proven to drift towards a long-term mean. These residuals represent the divergence between the two stocks from their mean relationship.

Second, these residuals are normalized to identify significant deviations. A rolling Z-score of a 100 day window was calculated. The Z-score series generated indicates the number of standard deviations the residuals are from their rolling mean, which will be used as a signal for trading.

Lastly, to exploit the mean-reverting opportunities of the spreads, the strategy takes a long position on the spread when the Z-score series falls below −2 and when the Z-score series exceeds +2. Positions are closed when the Z-score reverts to zero.

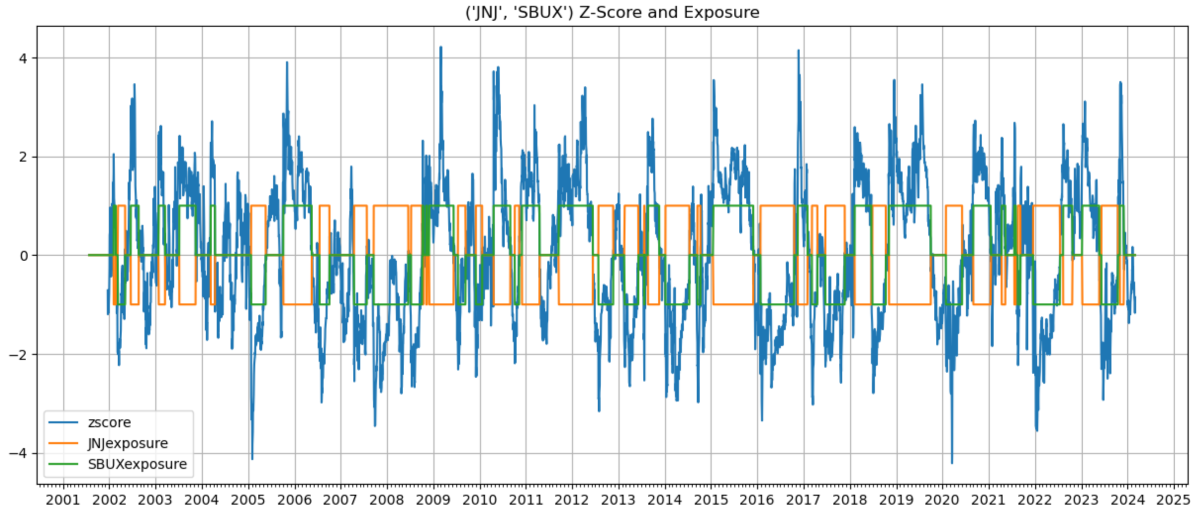

Figure 21: Z-Scores and Pairs Trading Exposure for JNJ / SBUX Pair

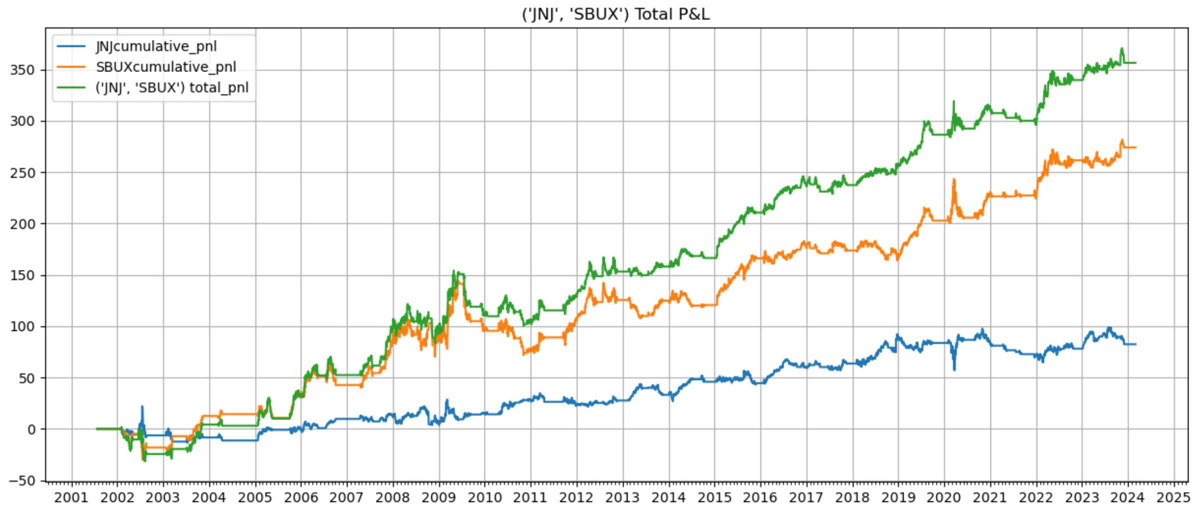

Figure 22: P&L for JNJ / SBUX Pair

6.1 Results

While this approach is very simple, it was still able to generate profitable returns ($1706.34) and an overall Sharpe ratio of 1.58 from 2002-12-01 to 2024-02-29.

Individual Pairs Performance:

('JNJ', 'PEP') P&L: 65.79, Sharpe Ratio: 1.70
('JNJ', 'SBUX') P&L: 356.38, Sharpe Ratio: 1.38

('JNJ', 'YUM') P&L: 215.79, Sharpe Ratio: 1.18
('ABT', 'SBUX') P&L: 490.72, Sharpe Ratio: 1.67
('ABT', 'YUM') P&L: 103.79, Sharpe Ratio: 1.96
('SYK', 'MCD') P&L: 275.39, Sharpe Ratio: 1.92
('ZBH', 'SBUX') P&L: 198.49, Sharpe Ratio: 1.23

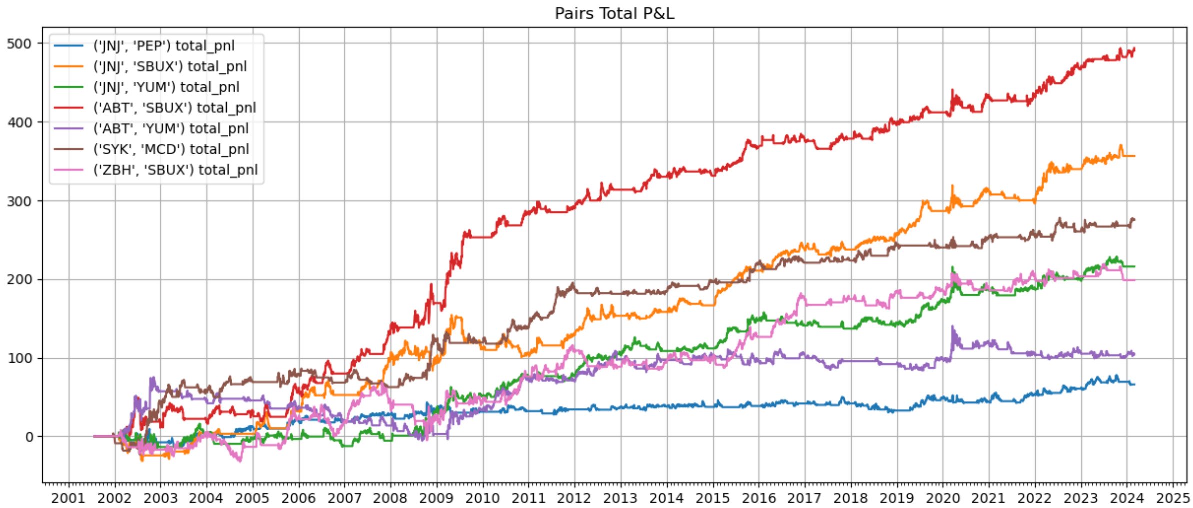

Figure 23: Individual P&L for All Pairs

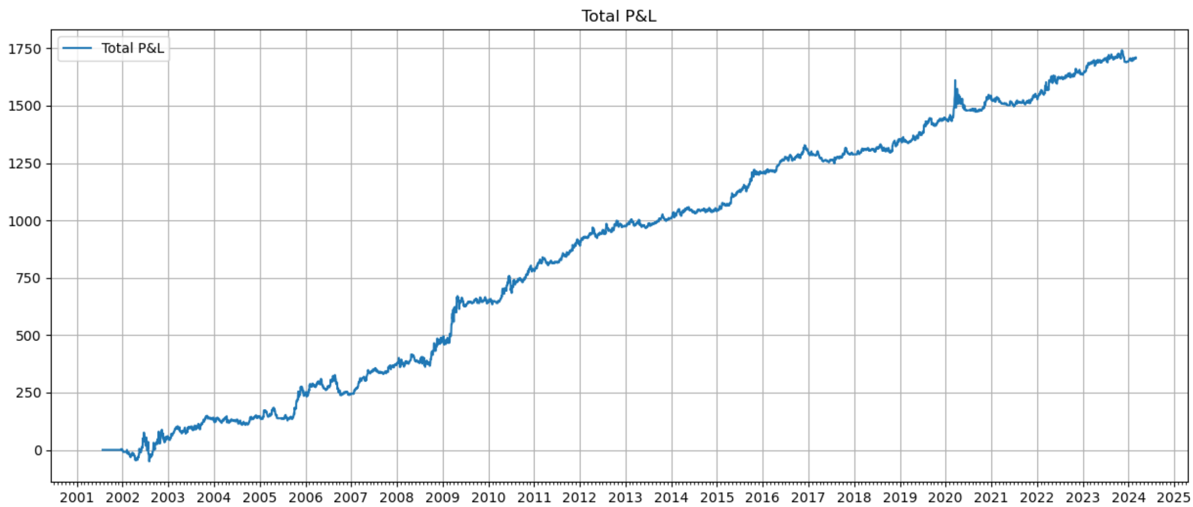

Figure 24: Total P&L for All Pairs

6.2 Conclusion

Overall, this report shows how medical relationships involving processed food components can influence financial markets. We utilized data processing and analysis, external research, presented quantitative models for modeling obesity levels, and developed an innovative pairs selection and trading algorithm. This algorithm confirms that stocks of companies producing or using processed foods correlated with obesity also correlate with healthcare sector stocks, offering a method to capitalize on this pattern.

6.2.1 Limitations

Despite the strategy being simplistic with the use of linear regression, it was still profitable, which illustrates how correlations between different markets could be exploited for pairs trading instead of only performing pairs trading on stocks within the same sectors.

Due to our focus on industrial meats and processed sugars, we may have overlooked other significant categories of processed foods that also impact health and economic outcomes. These limitations underscore the need for a more detailed and comprehensive analysis to fully understand the complexities of food consumption and its broader implications.

One limitation of the pairs trading strategy is that it relies solely on pairs consisting of one ticker from the healthcare industry and one from the food and beverages industry, potentially overlooking valuable correlations within a broader range of sectors. This narrow focus, dictated by data availability from Yahoo Finance and a provided dataset, may miss out on more nuanced relationships and opportunities for arbitrage. Additionally, while the pre-partitioning method effectively reduces computational costs and the likelihood of spurious relations, it may also exclude potentially significant pairs that do not fit the initial criteria, thereby limiting the scope and comprehensiveness of the analysis.

6.2.2 Future Directions

Future research should explore the specific mechanisms by which broiler meat contributes to obesity, including an examination of its nutritional content, consumption patterns, and metabolic effects. Additionally, the impact of multinational corporations and their diversified product lines on stock correlations and public health outcomes deserves further investigation. Understanding these dynamics could enhance the development of more effective public health strategies and financial models in the food and healthcare sectors.

Moreover, employing advanced machine learning techniques or mathematical models, such as copula approaches and the Vector Error Correction Model, could improve pairs trading strategies. We additionally plan to backtest our trading strategy using SARIMA forecasting. Incorporating macroeconomic indicators or momentum signals could also be beneficial for improving profit and loss (P&L) and the Sharpe ratio.

References

[1] P. Chemineau et al. "Seasonality of reproduction and production in farm fishes, birds and mammals". In: *Animal* 1.3 (2007), pp. 419–432. ISSN: 1751-7311. DOI: `https://doi.org/10.1017/S1751731107691873`. URL: `https://www.sciencedirect.com/science/article/pii/S1751731107691873`.

[2] David C.W. Lau et al. "2006 Canadian clinical practice guidelines on the management and prevention of obesity in adults and children [summary]". In: *CMAJ* 176.8 (2007). Ed. by, S1–S13. ISSN: 0820-3946. DOI: `10.1503/cmaj.061409`. eprint: `https://www.cmaj.ca/content/176/8/S1.full.pdf`. URL: `https://www.cmaj.ca/content/176/8/S1`.

[3] Lenny R Vartanian, Marlene B Schwartz, and Kelly D Brownell. *Effects of soft drink consumption on nutrition and health: A systematic review and meta-analysis*. Apr. 2007. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1829363/.

[4] Eric A. Finkelstein et al. "Annual medical spending attributable to obesity: Payer- and service-specific estimates". In: *Health Affairs* 28.Supplement 1 (Jan. 2009). DOI: 10.1377/hlthaff.28.5.w822.

[5] M. Avellaneda and J.H. Lee. "Statistical Arbitrage in the U.S. Equities Market". In: *Quantitative Finance* 10.7 (Jan. 2010), pp. 761–782. DOI: 10.1080/14697680903124632.

[6] Michael. *New Yorks ban on Big Sodas is rejected by Final Court*. June 2014. URL: https://www.nytimes.com/2014/06/27/nyregion/city-loses-final-appeal-on-limiting-sales-of-large-sodas.html.

[7] Caroline M. Apovian. "The obesity epidemic understanding the disease and the treatment". In: *New England Journal of Medicine* 374.2 (Jan. 2016), pp. 177–179. DOI: 10.1056/nejme1514957.

[8] G.A. Bray, K.K. Kim, and J.P.H. Wilding. "Obesity: A chronic relapsing progressive disease process. A position statement of the World Obesity Federation". In: *Obesity Reviews* 18.7 (May 2017), pp. 715–723. DOI: 10.1111/obr.12551.

[9] Michael J Gibney et al. "Ultra-processed foods in human health: a critical appraisal". In: *The American Journal of Clinical Nutrition* 106.3 (2017), pp. 717–724. ISSN: 0002-9165. DOI: https://doi.org/10.3945/ajcn.117.160440. URL: https://www.sciencedirect.com/science/article/pii/S0002916522026235.

[10] Samir Faruque et al. *The dose makes the poison: Sugar and obesity in the United States - A Review*. 2019. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6959843/.

[11] Oct. 2020. URL: https://www.hsph.harvard.edu/news/hsph-in-the-news/heavy-consumption-sugary-beverages-declining/#:~:text=Heavy%20consumption%20of%20sugary%20beverages%20declining%20in%20U.S.,from%20Harvard%20T.H.%20Chan%20School%20of%20Public%20Health..

[12] S.M. Sarmento and N. Horta. "Enhancing a Pairs Trading strategy with the application of Machine Learning". In: *Expert Systems with Applications* 158 (Nov. 2020), p. 113490. DOI: 10.1016/j.eswa.2020.113490.

[13] *Obesity: Causes, consequences, treatments, and challenges*. Oct. 2021. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8530518/.

[14] Anthony Crimarco, Matthew J Landry, and Christopher D Gardner. *Ultra-processed foods, weight gain, and co-morbidity risk*. Sept. 2022. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8532572/.

[15] USAFacts. *US obesity rates have tripled over the last 60 years*. Mar. 2023. URL: https://usafacts.org/articles/obesity-rate-nearly-triples-united-states-over-last-50-years/.

[16] May 2024. URL: https://www.cdc.gov/brfss/index.html.

[17] Elmaz Sabovic. *Coca-Cola vs. Pepsi: Which is a better stock pick for 2024?* Apr. 2024. URL: https://finbold.com/coca-cola-vs-pepsi-which-is-a-better-stock-pick-for-2024/.

[18] Verticlemycoldadmin. *An overview of cold storage in the food industry*. Feb. 2024. URL: https://www.verticalcold.com/an-overview-of-cold-storage-in-the-food-industry/.

[19] URL: https://www.cdc.gov/obesity/php/data-research/adult-obesity-facts.html.

[20] URL: https://www.nationalchickencouncil.org/about-the-industry/statistics/.

[21] URL: https://www.pcrm.org/good-nutrition/nutrition-information/chicken.

[22] URL: https://www.ers.usda.gov/data-products/livestock-and-meat-domestic-data/livestock-and-meat-domestic-data/#All%20Meat%20Statistics.

[23] URL: https://www.who.int/publications/i/item/9789241510066.

[24] URL: https://pubmed.ncbi.nlm.nih.gov/11234459/.