Citadel Datathon 2024

Team 18 Submission

Choi Yat Long
National University of Singapore

Lihua Guo
TsingHua University

Mulan Qin
National University of Singapore

Ruichen Li
University of Sydney

Abstract

In recent years, the U.S. economy has experienced growth, paralleled by an increase in a disease of affluence, obesity. A significant contributor behind these trends may be the processed meat industry, a cornerstone of the economy and a staple in the American fast-food culture. Until now, the comprehensive impact of meat production on the financial market and the society has remained underexplored. This report aims to elucidate the relationship between meat production and its broader financial and societal effects.

Our study begins with an analysis of the correlation between meat production and financial markets, revealing a negative correlation with stocks in sectors related to processed food. We further examined the social ramifications of meat production, highlighting its role in promoting unhealthy eating habits and increasing obesity rates, while also contributing to a lower unemployment rate by providing employment opportunities.

In summary, although meat production bolsters the macroeconomy, it necessitates governmental intervention to address the accompanying health challenges.

Contents

1 Non-Technical Executive Summary

1.1 background

The US economy has been flourishing since the 2008 financial crisis. Without a shadow of a doubt, prosperity is always one of the goals that every government fights for. Nevertheless, some may have noticed affluence is also associated with some illnesses. This is also known as diseases of affluence. Office workers may order fast food so as to strive for more time to work. Some may just want to be lazy.

Originating in the US, fast food has become a popular choice for many citizens. Just by walking to a nearby fast food restaurant, customers can quickly enjoy delicious burgers, French fries, and Coke. However, when one overlooks health issues, processed meat can be very tempting. Due to this culture, related corporations, such as fast food chains and processed meat firms, are experiencing strong financial health.

Unfortunately, by not paying enough attention to their dietary habits and healthy lifestyles, obesity has become increasingly prevalent, surging to 33% in recent years. Consuming more meat and fewer vegetables can lead to obesity, which is associated with numerous illnesses, including diabetes. Therefore, it is of utmost importance to combat obesity. Otherwise, the cost to the medical system and even to society will be overwhelming.

To further investigate the relationship between and explore the effect of increasing processed meat production, in this report, we conduct a research to find out how they affect our society.

1.2 Main Questions

In this study, we address two central questions:

1. Could meat production be an indicator for the financial health, as reflected by their stock market performance?
2. What's the social influence of meat production?

1.3 Key Findings

1. To explore the interrelation between meat production and stock market, we introduce the Poultry Production Index (PPI) as an innovative metric that reflects meat production levels and notably correlates with obesity rates. The investigation reveals significant negative correlations between PPI and the annual stock returns of companies within the processed food domain. Notably, firms such as Starbucks, Pilgrim's Pride Corp, and Conagra Brands are highlighted as prominent examples of this trend. Moreover, our Granger causality analysis substantiates the predictive influence of preceding PPI values on stock returns, predominantly within the manufacturing sector. This causative connection suggests that escalations in meat production, as indicated by elevated PPI values, might paradoxically exert detrimental effects on the financial landscape of the processed food industry.
2. The meat production has a positive effect on employment. This may due to it provides more jobs opportunities. However, it also causes a higher obesity rate. This may due to it changes people's eating habits. The rate of people not eating enough vegetables increase along with the meat production.

2 Technical Exposition

2.1 Overall Workflow

The overall analysis workflow is illustrated in Figure 1, encompassing three categories of data: (1) **Meat Industry**, which includes data related to two stages of the meat production process; (2) **Financial Market**, which comprises information on stocks or ETFs from four sectors; and (3) **Social Influence**, which details the challenges encountered by diverse community groups within society. Based on these data, our analysis proceeded in two phases. Firstly, we explored the interconnections between meat production data and financial market trends, with a particular emphasis on the influence of manufacturing and technology stocks on meat production, as well as the reciprocal impact of meat production on trade & services stocks and various indexes. Secondly, we specifically analyzed how fluctuations in meat production and financial markets affect different communities, especially for their health condition and employment condition.

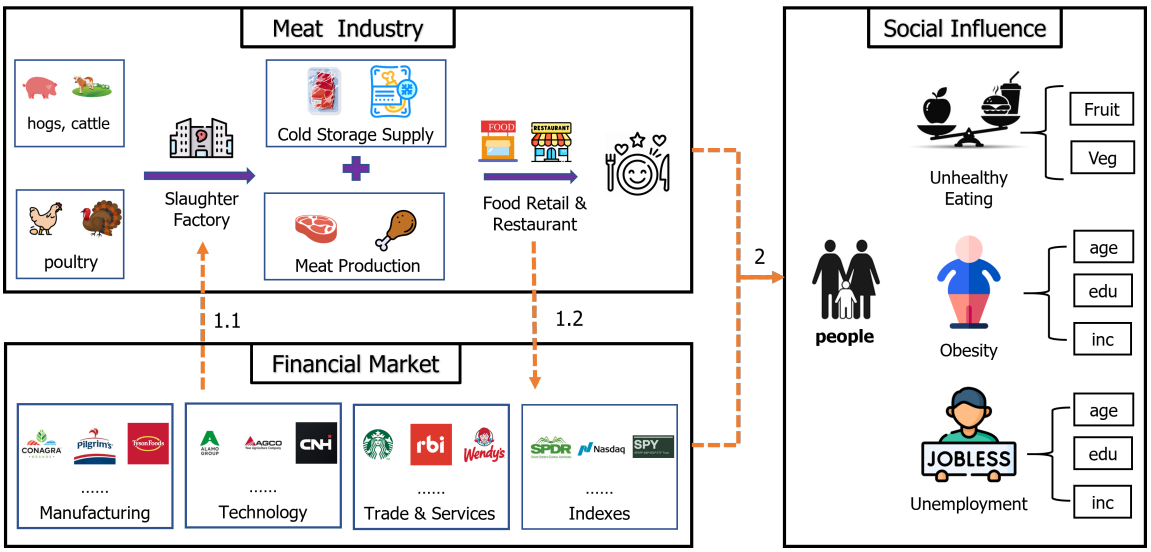


Figure 1: Overall Analysis Workflow

Note: - - → denotes the impact direction, ⟶ denotes the pipeline of meat production. 1.1 denotes the impact from Manufacturing and Technology to Meat Slaughter; 1.2 denotes the impact of meat production to Trade & Services stocks and Indexes; 2 denotes the impact from meat industry and financial market to individual health and unemployment condition.

2.2 Data Preprocessing and Feature Engineering

2.2.1 Data Preprocessing

Before our analysis, we first conducted a cleaning and transformation of the data. Below, we will detail the data source we used to construct the three categories of data and explain the transformation we did.

2.2.2 Data Transformation

- **Meat Industry**.

  We derived our *Meat Industry* data from United States Department of Agriculture (USDA), which are specifically `Meat_Stats_Slaughter_Counts.csv`, `Meat_Stats_Slaughter_Counts.csv`, `Meat_Stats_Slaughter_Weights.csv`, `Meat_Stats_Cold_Storage.csv`. We first combined data from `Slaughter_Counts.csv` and `Slaughter_Weights.csv` to get the total weight of slaughter. Specifically, we calculated the total slaughter weight of an animal by multiplying their slaughter count and average slaughter weight. Then for each csv data, we summed up the total weights of different animals from same "Commercial_Or_Federally_Inspected" and "Type_Of_Meat" types and joined them together. After aligning the Unit, we get the *Meat Industry* data in the following format:

column	unit
year	-
month	-
Slaughter Commercial Red Meat	pounds
Slaughter Federally Inspected Red Meat	pounds
Slaughter Federally Inspected Poultry	pounds
Cold Storage Red Meat	pounds
Cold Storage Poultry	pounds
Production Commercial Red Meat	pounds
Production Federally Inspected Red Meat	pounds
Production Federally Inspected Poultry	pounds

Table 1: data format of *Meat Industry* data.

- **Financial Market**.

  We derived *Financial Market* data from Alpha Vantage and Yahoo! Finance, which includes `all_stock_and_etfs.csv` and `stock_descriptions.csv`. In the analysis, we mainly focus on the price of each stock and we choose the close price as the indicator. Furthermore, we calculate the log return rate of each day with the following formula:

$$\text{log\_daily\_return}_i = \log\left(\frac{\text{close\_price}_{i+1}}{\text{close\_price}_i}\right)$$

  And we also extend this to monthly log return and annual return by:

$$\text{log\_monthly\_return}_m = \sum_{i=\text{first\_day\_of\_m}}^{\text{last\_day\_of\_m}} \text{log\_daily\_return}_i \tag{1}$$

$$\text{log\_annual\_return}_y = \sum_{i=\text{first\_day\_of\_y}}^{\text{last\_day\_of\_y}} \text{log\_daily\_return}_i \tag{2}$$

  Finally, we derived the *Financial Market* in the following format:

column	Meaning
Date	Date of the line
Year	Year of the Date
Month	Month of the Date
ADM_Close_Price	ADM's close price of the Date
ADM_Log_Daily_Return	ADM's log return for the Date
ADM_Log_Monthly_Return	ADM's monthly return for the Month
ADM_Log_annual_Return	ADM's annual return for the Year
Other Stocks...	Other Stocks...

Table 2: data format of *Financial Market* data.

- **Social Influence**.

  We derived *Social Influence* Data from CDC Behavioral Risk Factor Surveillance System, including `Nutrition_Physical_Activity_and_Obesity_Data.csv` and `acs_5yr_est_selected_economic_characteristics_2010-2022.csv`. For Unhealthy Eating, we analyzed the first csv and focused on Q018. Q019, Q020 and Q021, for Obesity, we mainly focus on Q036 and Q038. Then We record the value of each year, each Stratification in the following format:

column	Meaning
Year	Year of the data
SCID_i-SID_i-FruitConsuming	percent of communities not eating enough fruit
SCID_i-SID_i-VegConsuming	percent of communities not eating enough vegetable
SCID_i-SID_i-Obesity	percent of communities with obesity question
Other Communities...	Other Communities...

Table 3: data format of *Social Influence*-unhealthy data.

For Unemployment data, we analyzed the second csv and focused the data related to UNEMPLOYMENT STATUS. We extract the Unemployed percentage from the Civilian labor force and use it as the annual unemployment rate data. We extract the data from each state and also download the US National data from the official website. Finally we record the data in the following format:

column	Meaning
Year	Year of the data
National	National unemployment rate
$State_i$	Unemployment rate of a specific state
Other States...	Other States...

Table 4: data format of *Social Influence*-unemployment data.

2.2.3 Data Cleaning

During the process of data transformation, we also conducted a data cleaning to make sure the quality of it. We mainly did the following job:

1. **Filter missing values**.

   We removed data entries containing missing values to ensure the calculation process would not produce NaN errors.

2. **Filter extreme values**.

   We eliminated outliers to ensure the analysis results would not be affected by extreme cases. Specifically, if a data point falls outside the 3 sigma range of its data from the recent year, we consider it an outlier and remove it.

3. **Unit Alignment**

   During processing of *Meat Industry* data, some data were in the Unit of "Million Pounds", and some data were in the Unit of "Average Pound". For conveniently usage, we transform all of the weights related data's Unit into Pound for alignment.

4. **Numeric Value Parsing**

   Some of the numerical value was stored in the format of "xxx,xxx.xx". To convert it into numerical, we first remove all the commas and then translate it into numerical value.

2.3 Feature Engineering

2.3.1 Measure of Stock Market Performance in Sectors Related to Processed Food

We opt for annual log returns due to their advantages in financial analysis:

- **Mitigating Price Scale Effects**: Log returns effectively minimize the impact of securities' absolute price levels, prioritizing the rate of return. This approach is invaluable for comparing securities across different price scales, as it ensures analyses reflect relative percentage changes rather than incomparable absolute dollar values.
- **Handling Asymmetry in Gains and Losses**: Log returns address the asymmetry between gains and losses, recognizing that a specific percentage loss necessitates a more significant percentage gain for recovery. This nuanced relationship is more accurately represented by log returns than by linear returns.

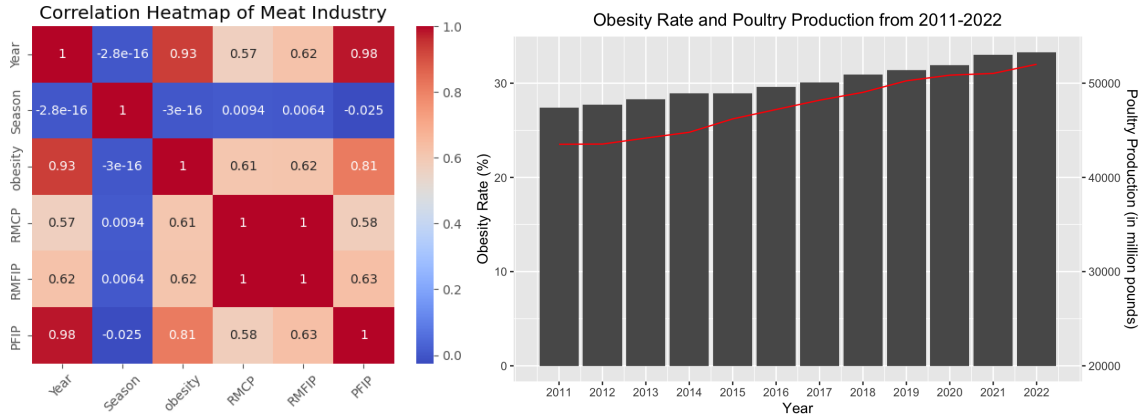
The formula for calculating the annual log return is expressed as:

$$\text{annual\_return\_rate} = \log\left(\frac{\text{last\_day\_close}}{\text{first\_day\_close}}\right)$$

2.3.2 Measure of Meat Production

To establish a measure of meat production, several factors need consideration, including its capacity to mirror aggregate meat supply and demand trends and its association with obesity.

We begin by analyzing the correlation between meat production and obesity rates. Surprisingly, our findings reveal a stronger correlation between obesity rates and poultry production compared to red meat. This may be attributed to the widespread perception of poultry as a "healthier" alternative to red meat, leading to heightened consumption due to health-conscious dietary trends. However, studies indicate that there is no significant health discrepancy between the two. In fact, research from the American Journal of Clinical Nutrition suggests that white meat can raise LDL (bad) cholesterol levels similarly to red meat.



(a) Heatmap of Food Production (b) Obesity Rate and Poultry Production from 2011-2022

Note: RMCP: Red Meat Commercial Production, RMFIP: Red Meat Federally Inspected Production, PFIP: Poultry Federally Inspected Production. Poultry does not have commercial production from our given data set.

Figure 2: Comparison of Food Production and Obesity Rate

Collinearity is another critical factor for consideration. The strong correlation observed between red meat and poultry production could introduce collinearity issues if both variables are included in our analysis. To circumvent this, we choose to separate red meat production and poultry production. After careful considerations, we introduce the Red Meat Production Index (RMI) and Poultry Production Index (PPI), and decide to use PPI for our main analysis.

Subsequently, we analyzed the relationship between the quantities of meat slaughtered meat, the weight of slaughtered meat, and the volume of meat in cold storage in each year. These metrics serve as indicators of the annual meat supply that satisfies consumer demand.

$$M_t = \left(\sum_{i=0}^{n} \text{slaughter\_weight}_{i,t} \times \text{slaughter\_count}\_i, t \right) + \text{cold\_storage\_weight}_{t-1} - \text{cold\_storage\_weight}_t$$

where:

- M_t represents the total poultry meat supply in year t.

- i is the index for each animal, with n being the total number of poultry animals.
- Slaughter Weight$_{i,t}$ denotes the average weight of poultry animal i slaughtered in year t.
- Slaughter Count$_{i,t}$ indicates the number of animal i poultry slaughtered in year t.
- Cold Storage Weight$_t$ represents the weight of poultry stored in cold storage at the end of year t.

We define the Poultry Production Index PPI as the logarithmic change in the supply of poultry meat from year $t-1$ to year t. This formulation aligns the PPI with the methodology employed for calculating stock returns:

$$PPI_t = \log\left(\frac{M_t}{M_{t-1}}\right)$$

Likewise, we develop a Red Meat Production Index (RMI) using the same approach with red meat data.

2.3.3 Scale and Range

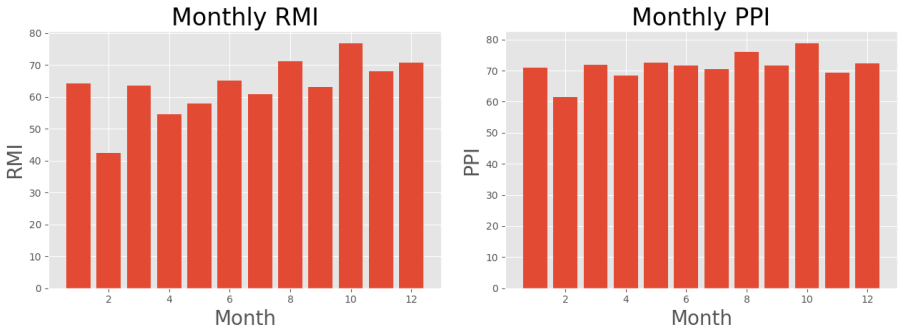
The selection of the 1999-2024 time frame for our analysis is driven by two primary considerations:

- **Temporal Relevance and Trends:** The period from 1999 to 2024 is particularly critical for our discussion due to the significant increase in both obesity rates and meat production during these years. It is a time period where lifestyle changes and dietary patterns have evolved substantially, making it a rich dataset for analyzing trends and drawing more current and applicable insights. Focusing on this timeframe allows for the examination of more recent phenomena, ensuring that the analysis reflects contemporary market dynamics and health trends.
- **Data Seasonality and Annual Aggregation:** By utilizing annual data, the analysis effectively addresses and mitigates the impact of seasonal trends in meat production, which may not directly correlate with stock market movements. Seasonal variations in agriculture and food production can introduce confounding variables, obscuring the underlying trends and relationships between meat production and market performance. Annual aggregation smooths out these seasonal fluctuations, providing a clearer view of long-term trends and their implications for the stock market.

  As shown in fig 3, the production across different months can vary up to 75%. So we choose year granularity to eliminate this effect in analysis.

2.4 Correlation and Causal Relations Between PPI and Annual Stock Returns

Employing the newly introduced Poultry Production Index (PPI), we can now assess the correlation and causation between meat supply and stock market dynamics in the processed food sector to investigate the financial implications of meat production.



(a) RMI across different months (b) PPI across different months

Figure 3: Meat Production across Different Months

2.4.1 Correlation

We proceed to examine the performance of each stock, identifying the five stocks exhibiting the strongest correlation with the PPI, which are shown in Table 5:

Ticker Symbol	**Correlation**
SBUX	-0.697
PPC	-0.656
CAG	-0.37
SAP	-0.330
DE	-0.329

Table 5: Correlation between PPI and Annual Log Return by Stock

The negative and strong correlation suggests that meat production is inversely related to the financial performance of the processed food-related stock market. To quantify the correlations, we perform ordinary linear regression with PPI, obesity rate, year, and previous year's log return as covariates on the dependent variable of annual log return:

$$\text{annual log return}_t = \alpha + \beta_1 \text{PPI}_t + \beta_2 OBI_t + \beta_3 \text{annual log return}_{t-1} + \epsilon_t$$

where t represents the year.

We summarize the result in Table 6:

While our model demonstrates a low R-squared value, as anticipated when predicting stock market outcomes, we observe some significant relationship between our covariates and annual log returns. The analysis shows a predominant negative correlation of stocks with the Poultry Production Index (PPI) and obesity rate, contrasting with positive correlations with annual and prior year log returns. This pattern suggests a minimal adverse effect of reducing poultry production on financial markets. A feasible explanation is that poultry is closer to an inferior good compared to other healthier alternatives. Thus, in a robust economy with elevated stock log returns, demand and consequently production of poultry tend to decrease.

Variables	All Stocks	Trade and Services	Manufacturing	Technology
PPI	-1.9645*	-1.4793	-3.0951*	-2.1267
	(0.855)	(1.630)	(1.430)	(1.587)
obesity rate	-1.1590*	0.0521	-1.8851*	-1.5012
	(0.504)	(0.955)	(0.868)	(0.914)
Year	0.0139*	0.0006	0.0224*	0.0211
	(0.006)	(0.012)	(0.011)	(0.012)
log return lag	0.0842	0.1768	0.0359	0.0723
	(0.046)	(0.099)	(0.080)	(0.072)
Constant	-27.5121*	-1.1513	-44.5014*	-41.9721
	(12.928)	(24.645)	(22.109)	(23.362)
F-Value	5.006	1.057	3.333	2.398
p-Value	0.000587	0.380	0.0118	0.0560
R-squared	0.042	0.030	0.075	0.097

***Reject H0 at 1% level of significance, **Reject H0 at 5% level of significance, *Reject H0 at 10% level of significance.

Table 6: Correlation and Causation Between PPI and log Return by Sectors

Notably, the correlation appears strongest in the manufacturing sector. Among the stocks in the manufacturing sector, Pilgrim's Pride Corp (PPC) stands out with a statistically significant correlation between PPI and annual log returns. Pilgrim's Pride Corp exemplifies the poultry manufacturing industry, specializing in the global production, processing, marketing, and distribution of fresh, frozen, and value-added chicken and pork products. One explanation for this correlation is the intense competition, increased production efficiency within the poultry industry, and higher cost of inputs, which may result in pricing competition and reduced margins for companies like Pilgrim's Pride Corp. Moreover, the sector might face overcapacity challenges, where production surpasses demand for poultry, possibly leading to diminished stock returns despite increased production levels. However, it's crucial to consider that these relationships are part of a broader, multifaceted market ecosystem, which might be moderated or accentuated by other factors such as different companies' varied resilience to changes in meat production

2.4.2 Granger-Causality

After exploring the correlation between meat production and the financial market for processed food, we also want to see if a causal relationship exist between these two variables. To achieve this, we employed the Granger causality test to determine if changes in one could be used to predict changes in the other. Take note that this test does not establish a true cause-and-effect relationship, but it could provide insights into our analysis. Table 7 presents some noteworthy results from this analysis:

The Granger causality tests results reveal no causality from log returns to the Producer Price Index (PPI) due to high p-values. Conversely, low p-values at lags 1 and 2 indicate Granger causality from PPI to log returns, suggesting a directional causal effect from PPI on log returns.

After testing across various stock sectors, we find that this causal relationship is uniquely identified in the manufacturing sector. Recall our prior analysis identified a negative correlation between stock returns in the manufacturing sector and PPI. This suggests that past PPI values could detri-

Lag length	1	2	3
All Stocks			
Log Return does not Granger Cause PPI	F=0.3454	F=0.3211	F=0.4820
	p=0.5570	p=0.7255	p=0.6949
PPI does not Granger Cause Log Return	$F = 7.9401^{**}$	$F = 4.2479^{*}$	F=2.5760
	p=0.0050	p=0.0149	p=0.0533
Manufacturing Sector			
PPI does not Granger Cause Log Return	$F = 5.3640^{*}$	F=2.3832	F=1.6471
	p=0.0218	p=0.0955	p=0.1807

***Reject H0 at 1% level of significance, **Reject H0 at 5% level of significance, *Reject H0 at 10% level of significance.

Table 7: Granger Causality Tests

mentally influence stock returns, which evidence supports our hypothesis that increased meat production, as suggested by PPI, heightens price competition, adversely affecting the manufacturing sector.

2.5 Social Influence of Meat Production

After analyzing the correlation between meat production and financial market, we obtained insights into the impact of meat production on the macro-economy. Furthermore, we also analyzed its impact on individuals. Specifically, we analyzed the correlation between it and two important aspects of societal conditions, obesity and employment situations. The results show that Red Meat production has a more positive social influence and while the Poultry production has a negative one. The detailed results are shown in the following sections.

2.5.1 Impact on Unhealthy Eating

To evaluate meat production's correlation with people's obesity, we first analyze its impact on people's eating habit. Specifically, we use the percentage of people not eating fruit or vegetable everyday as an indicator of unhealthy eating.

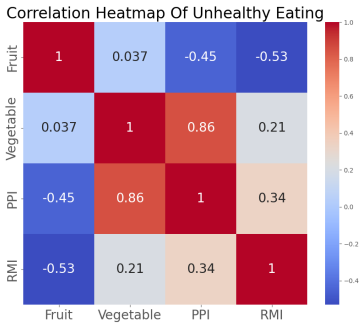
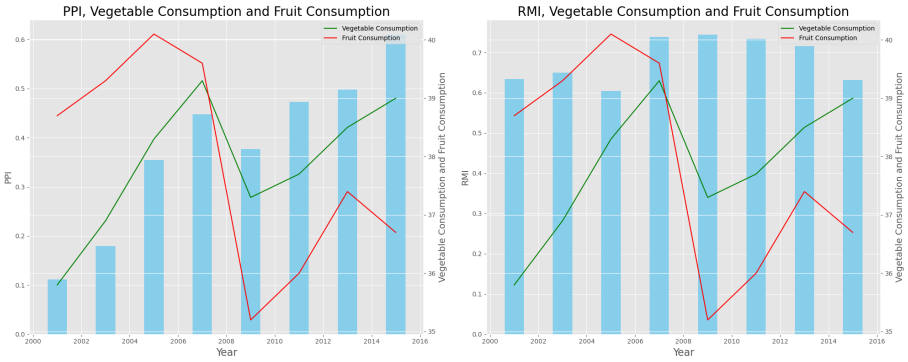


Figure 4: Correlation Between Unhealthy Eating Condition and Meat Production.

The correlation results between meat production and these two indicators is shown in fig 4 and the change through years is shown in fig 5.



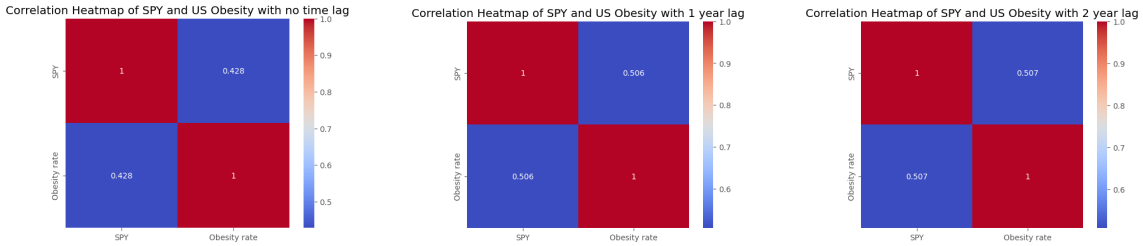
(a) PPI and Unhealhty Eating (b) RMI and Unhealthy Eating

Figure 5: Comparison of Meat Production and Unhealthy Eating Condition

From the results, we can see that the Poultry production has a strong impact on people's unhealthy eating habit. As the Poultry increases, the percentage of people eating fruits and vegetables increase. This may lead to a worse health condition of the society.

2.5.2 Impact on Obesity

Obesity rates are always the lowest for the youngest age group. Then for any other age group across all year, the rates never fall below 25%. The figures are very worrying, especially for the middle age group. In some years, the obesity rates drop for age 65+ group while it just maintains at a high level of 30%+ for the middle-aged and elderly. This is obviously the result of the change in lifestyle after they graduate.



(a) SPY and Obesity (b) SPY and Obesity with 1 lag (c) SPY and Obesity with 2 lag

Figure 6: SPY and Obesity data

2.5.3 Impact on Unemployment

To discern the interconnectedness between employment trends and the poultry industry, we delve into an exploration of their mutual relationships. We focus on two significant employment indica-

tors—overall unemployment rates and specifically, female unemployment rates—as they relate to various poultry industry metrics. These include production levels, cold storage capacity, slaughter count, and the weight of poultry slaughtered. A comprehensive heatmap of correlations, as depicted in Figure 7, lays out the intricate web of interactions among these variables.

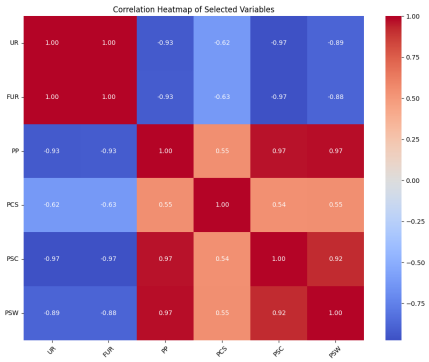


Figure 7: Correlation Between Unemployment and Meat Production.

The results emanating from our correlation heatmap and time series analysis are quite revealing. They suggest a substantial inverse relationship between the unemployment metrics and poultry industry parameters.

As unemployment rates decrease, a significant rise is observed in the measures of poultry production, suggesting that a robust poultry industry might be indicative of a healthier job market. This interplay between industry health and employment may also reflect broader economic conditions and societal trends.

3 Conclusion

In this study, we investigated the relationship between meat production, its economic impacts, and social consequences. Initially, we introduced two indices, RMI and PPI, as metrics for quantifying meat production levels. Subsequently, we examined the relationship between the yearly output of meat and the annual logarithmic returns of various stock categories. It was observed that all stocks, especially from the Manufacturing and Technology sector, exhibited a negative relationship with meat production volumes.

Furthermore, we investigated the societal impact of meat production, with a focus on public health and unemployment rates. Our findings suggest that while meat production contributes positively to employment rates, it also leads to an increased prevalence of obesity by influencing dietary habits adversely.

To conclude, although meat production contributes positively to the macroeconomic landscape, it prompts the need for policy interventions to mitigate the resultant health issues.

4 References

Lardieri, A. (2019, June 4). Study: White meat increases cholesterol levels similar to Red Meat. U.S. news. https://www.usnews.com/news/health-news/articles/2019-06-04/study-white-meat-increases-cholesterol-levels-similar-to-red-meat