How Do Corporate Health Claims and Food Safety Laws Affect Obesity Rates in the U.S.?

Executive Summary:

With obesity rates continuing to rise, the scrutiny of corporate practices in the food service industry has never been more important. In this paper, we explore company and state sentiment surrounding public health in conjunction with stock price and American behavior with the aim of predicting and better understanding obesity rates. Companies and lawmakers can make on-paper efforts regarding health, but do the results justify their efforts?

Utilizing sentiment analysis, we assess the emphasis that companies place on food quality and track how their focus on this issue evolves over time. We also explore the evolution of federal and state legislation related to food quality, observing how regulatory measures have changed over time and their potential influence on corporate behavior and public health outcomes. We aim to capture the complex interplay between corporate practices, public perception, health outcomes, and financial markets, underscoring the importance of aligning corporate claims with genuine public health improvements.

Given this background, our main questions were as follows:

- How do state-level variations in food safety laws and countrywide corporate health claims associate with differences in obesity rates across the U.S.?
- Which factors are informative and effective at influencing obesity rates at the state and federal levels?

From country-level time series analysis using LASSO regression and GLS, we find that the health direction of major food service companies, given by sentiment analysis of company annual reports, in conjunction with stock price is informative and important to predict the degree of obesity in America and that obesity rates are accurately predicted by food service company features along with the degree of activity of Americans (R-squared = 0.94; 0.96). These results reaffirm that the association between food service companies and obesity rates can be captured linearly to a degree, and suggest that companies' health direction is an important component of

that association. However, the company reports' sentiment aligning with health goals do not ubiquitously relate to a rise or fall in obesity rates.

State-level analysis reveals no significant relationship between the percentage of laws related to food and health and the average number of obese state residents. The weak correlation and non-significant regression results suggest that other factors might be more influential in determining obesity rates. Further research is needed to explore additional variables and more complex models to gain a better understanding of the factors influencing obesity.

Technical Exposition:

Data Exploration and Preprocessing, Country-Level

To gain a prior understanding of the relationship between obesity and corporate direction in the food service industry, we complemented our data preprocessing with outside readings: Edwards et al. find that the food service industry can bear some responsibility in the rise in obesity; Cobb et al. take a systematic review of 71 studies relating obesity to food environment, and while associations are predominantly null, we found that there was grounds for fast food availability positively corresponding to obesity rates. Thus, we hypothesize that health-positive sentiment from company annual reports in conjunction with strong stock performance could lead to a possible decline in obesity rates.

We first collected data on fast-food companies' commitment to making food healthier year by year. We focused on the five largest market-cap single restaurant companies: McDonald's, Chipotle, Texas Roadhouse, Wendy's, and Domino's.

To determine their devotion to healthier food, we performed sentiment analysis on companies' annual SEC report forms (10-K). We extracted section 1, which is the "Business" section, which typically contains information about current goals and the company's state.

Initially, we used OpenAI's embedding model for our own NLP model to determine commitment to healthier food, creating vector embeddings of the text found in this section and comparing it to vector embeddings of definitive statements like "One goal of ours is to make our food healthier." However, we found that performance was typically better if we simply used OpenAI's GPT-4o-mini model and prompted it to act as an NLP expert. We would input the excerpt from the given 10-K, and ask it to give a score from 1-to-100 on how much the company cares about healthier food.

When preprocessing *Nutrition Physical Activity and Obesity Data*, performing VIF (Variance Inflation Factor) analysis reveals a great degree of multicollinearity between different questions in the data (when considering all potential variables, VIF factor for our initial features was as high as 10*e+4). This can be understood by the similar context in questions such as 'percentage of adults who have obesity' and 'percentage of adults who have overweight classification. As such, we selected three features to use in later multivariate regressions, which we judged to be suitable for capturing the health-related behavior of Americans, and which were available over the entirety of the interested time frame. Further we performed Breuch-Pagan tests between the proportion of respondents agreeing to our selected questions and corresponding interested statistics such as state and sample size. Our results suggest that the selected data has a significant degree of heteroscedasticity, which we consider in model selection (for all tests, $p << 0.05$). When transforming the data into a time series of respondents' average responses to questions at a given time, we take the weighted average of the percentage of respondents affirming a question by the sample size, over either an individual state or the entire country (depending on our scale of analysis, described later).

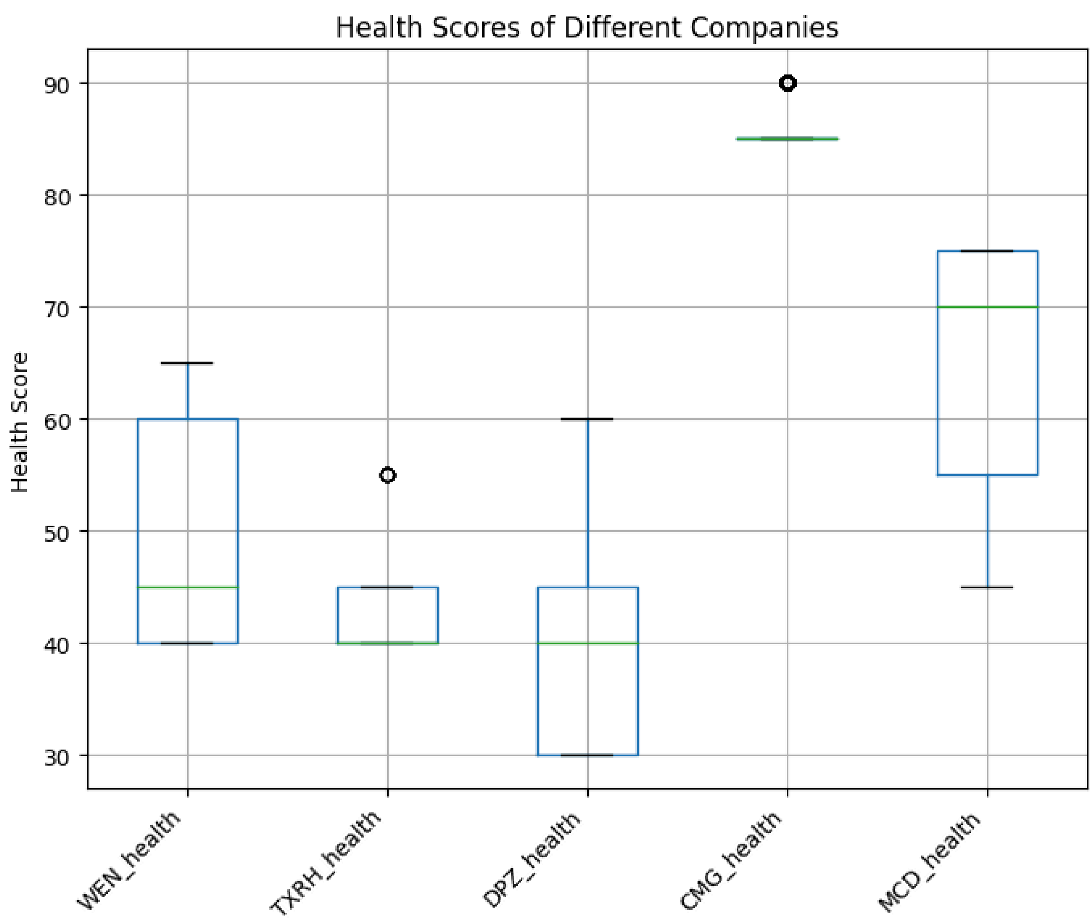

Figure: Health scores of different companies

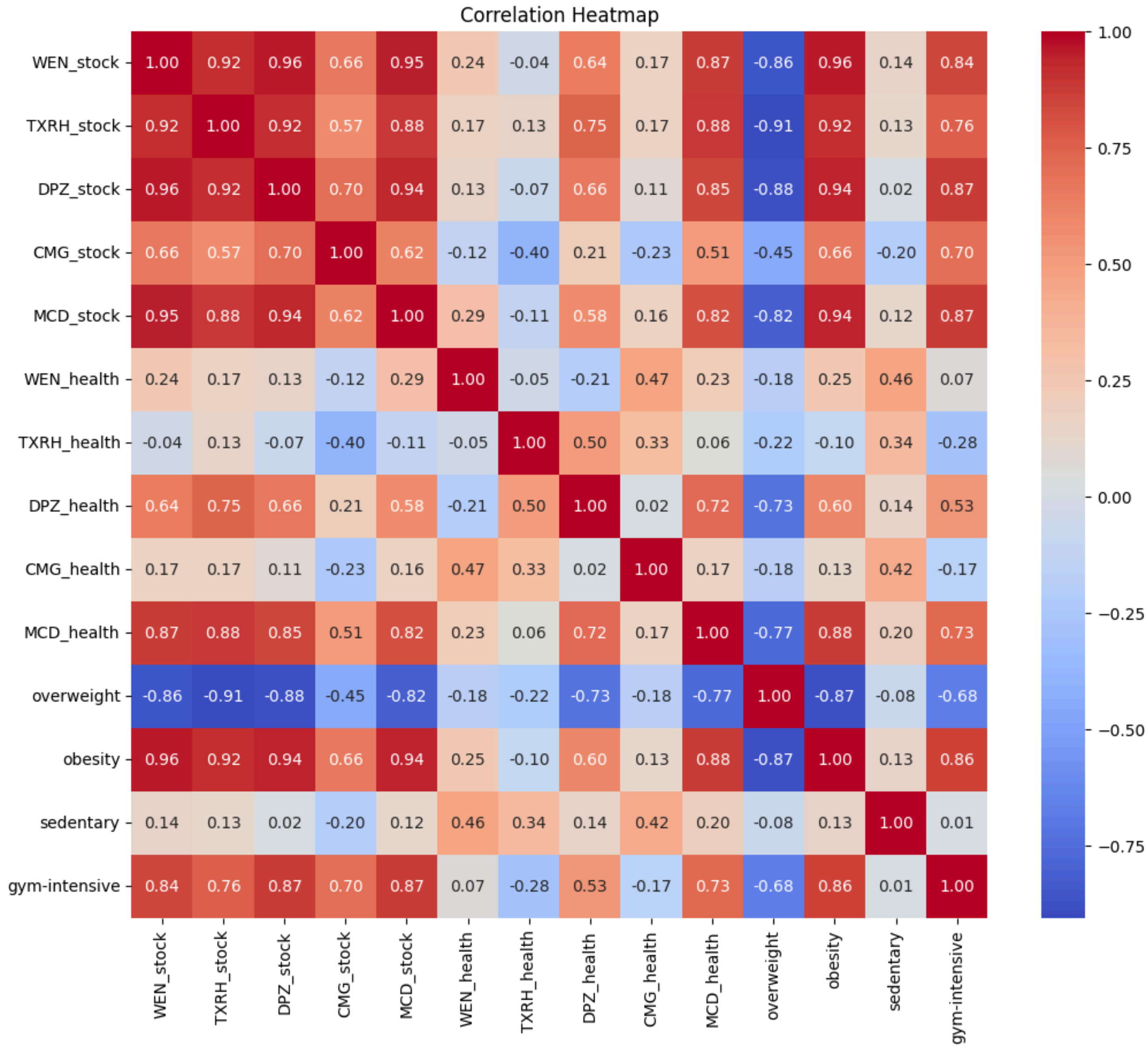

Figure: Correlation heatmap between features considered in countrywide analysis

Data Exploration and Preprocessing, State-level

Next, we needed statewide data that gave us info on sentiment at the government level toward healthier foods. We decided to use the proportion of food safety and health-related laws in relation to the total amount of laws passed. From the NCSL website, we scraped a list of all health laws passed, then extracted the total number of health and food laws passed per state per year using sentiment analysis, and the total number of food laws passed per state per year. This gave a measurement of how much health and food considerations were a priority in the eyes of lawmakers and politicians. Using the heatmap format across the United States, we can observe trends in commitment to healthier food.

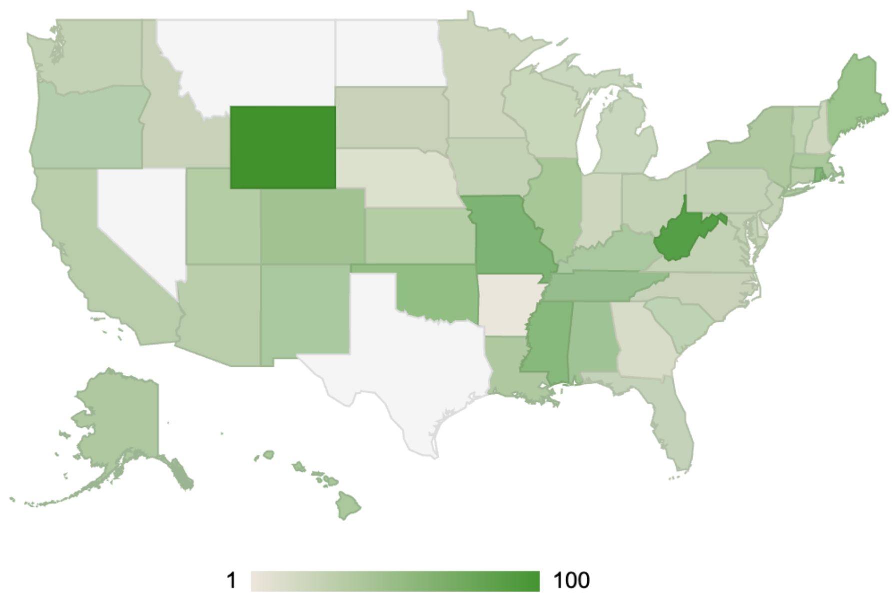

Figure: U.S. Heatmap by Percent of Health Laws Related to Food, 2018

In this analysis, we used 2018 data to investigate the correlation between obesity rates and food safety laws across various U.S. states. After loading the data, we performed interpolation on numerical columns to handle missing values.

We focused on adults aged 18+ with obesity in the year 2018. To handle multiple data values per state due to gender and race stratification, we calculated a weighted average by multiplying the data value by sample size, grouping by state, summing the weighted values and sample sizes, and then computing the weighted average for each state.

Simultaneously, we processed a dataset on health and food safety laws by state and year. We transposed the data, set appropriate column names, filtered for 2018, reset the index, renamed columns, and merged it with the obesity data on state abbreviation. Finally, we filtered the merged DataFrame to include only valid state abbreviations.

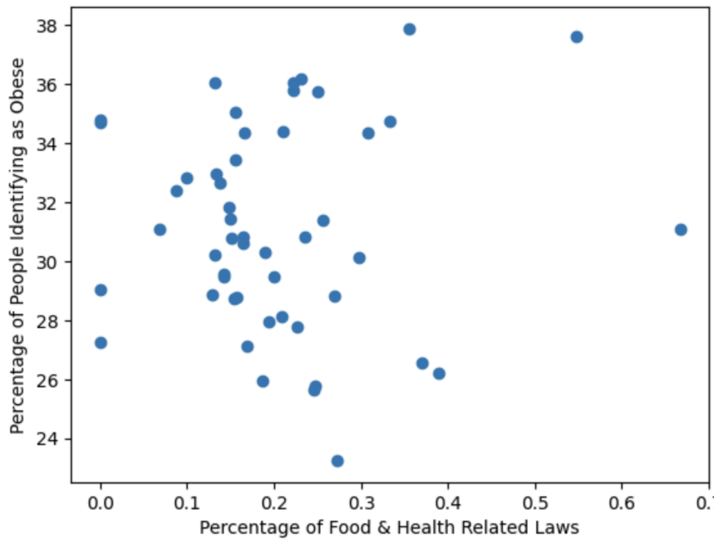

Figure: Scatterplot of Food and Health-Related Laws vs. Percentage of People Identifying as Obese

There is no clear relationship between the percentage of food & health related laws and the percentage of people identifying as obese.

Since our data proved to be heteroskedastic, we used a log-log transformation. To prepare the data, we first replaced all of the 0 values with 1 to ensure compatibility with the logarithm function. We then applied the natural logarithm to both the percentage of laws passed and the percentage of identified obese people. Additionally, we winsorized the top and bottom 1% of the log-transformed data to reduce the impact of outliers. The resulting data is as follows:

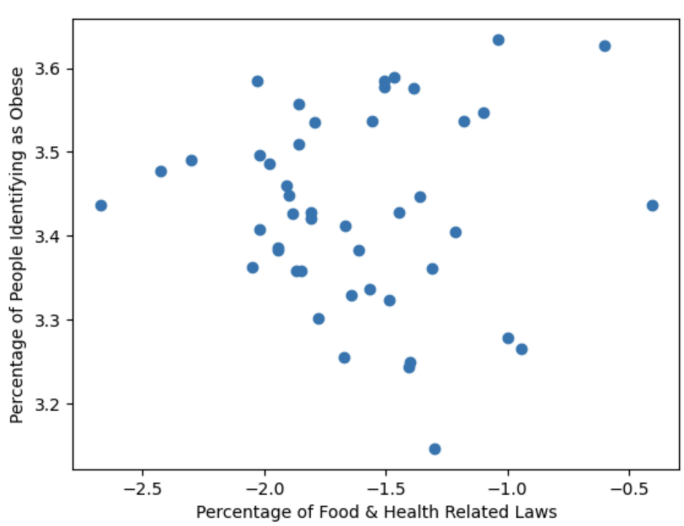

Figure: Post-Transformation Scatterplot

Modeling and Results, Country-level

We performed predictive modeling country-wide. Our country-wide data was given by features from the *Nutrition*, company stock prices at close on market days, and health scores of company reports given by sentiment analysis. We performed analysis from 2014 to 2022 with values taken daily; if a value has less than daily granularity or is unavailable on a given day (such as stock prices during weekends), we imputed the most recent given value. We hoped to predict the degree of obesity from other features (not including the degree of overweight condition) by two methods: LASSO regression, chosen in order to limit the number of features necessary for prediction and to have insight into feature importances; and a generalized least squares model (GLS), chosen because GLS can account for heteroscedasticity.

Our LASSO regression resulted in an R-squared of 0.94 and an MSE of 0.12, suggesting obesity is strongly predicted by our selected features. The nonzero coefficients were from highest to lowest magnitude

WEN_stock: 0.3974
MCD_stock: 0.2881
TXRH_stock: 0.1705
MCD_health: 0.1399
Gym-intensive: 0.1346
DPZ_stock: 0.1293

<u>Table: the absolute value of nonzero coefficients of LASSO regression</u>

The clear downside of this model is that LASSO regression assumes homoscedasticity, which is arguably untrue here. Our GLS resulted in an R-squared of 0.96 and an MSE of 0.08, confirming obesity is predicted by our selected features. The calculated effect sizes from our GLS were

XRH_stock: -0.0691
DPZ_stock: -0.2303
CMG_stock: 0.2313
MCD_stock: 0.2073
WEN_health: 0.1893
TXRH_health: -0.0017
DPZ_health: 0.3042
CMG_health: -0.0593
MCD_health: 0.2123
Gym-intensive: 0.1030

<u>Table: the effect sizes of features used in our GLS model</u>

which noticeably agrees with McDonald's stock price and company report health score, as well as Domino's Pizza stock price, being informative to the degree of obesity.

The results of our country-level modeling indicate that company reports in conjunction with their stock performance, are a powerful predictor of obesity rates in the United States, aligning with our hypothesis. Company sentiments aligning with health goals do not ubiquitously relate to a rise or fall in obesity rates. What motivates certain companies' stock and health reports to be more informative about obesity rates is a point of future research.

Modeling and Results, State-level

After conducting a country-level analysis, we conduct a state-level analysis. Based on our findings in the country-level analysis, we investigate the following question at the state level: *is there a statistically significant relationship between the percentage of health-related laws passed and the obesity rate for that state?*

For the statistical analysis, we performed an Ordinary Least Squares (OLS) regression to explore the relationship between the variables. An OLS regression provides straightforward coefficient interpretation and is computationally efficient. Additionally, OLS forms the foundation for more advanced statistical methods, making it a good starting point for further analysis. The model was defined as:

$$\log(\text{WeightedAverage}) = \beta_0 + \beta_1 \log(\text{Percentage})$$

The regression results were as follows:

- *Intercept (β_0)*: 3.4302 ($p < 0.001$)
- *Slope (β_1)*: -0.0006 ($p = 0.843$)
- *R-squared*: 0.001 (indicating that the model explains 0.1% of the variance in the log-transformed weighted average obesity rate)

The regression results suggest that the slope (-0.0006) is not statistically significant, with a p-value of 0.843. This indicates that the percentage of food and health-related laws does not have a meaningful effect on the weighted average obesity rates across states. The R-squared value is also very close to zero, implying that the model does not explain much of the variability in obesity rates.

The model has several shortcomings, including its simplicity, which may not capture more complex relationships, and the potential omission of other influential factors. Winsorization, while reducing the effect of outliers, may also exclude valuable information. The overall performance of the model was weak, with an R-squared value of 0.001, indicating that it explains only 0.1% of the variance in obesity rates.

This analysis reveals no significant relationship between the percentage of laws related to food and health and the average number of obese state residents. The weak correlation and non-significant regression results suggest that other factors might be more influential in determining obesity rates. Further research is needed to explore additional variables and more complex models to gain a better understanding of the factors influencing obesity.

Citations:

Cobb, Laura K., et al. “The Relationship of the Local Food Environment with Obesity: A Systematic Review of Methods, Study Quality, and Results.” Obesity, vol. 23, no. 7, 12 June 2015, pp. 1331–1344, https://doi.org/10.1002/oby.21118.

Edwards, John S. A., et al. “Overweight, Obesity and the Food Service Industry.” Food Service Technology, vol. 5, no. 2-4, June 2005, pp. 85–94,

“Database Environmental Health State Bill Tracking Database Archive: 2009 - 2022.” National Conference of State Legislatures, www.ncsl.org/environment-and-natural-resources/environmental-health-state-bill-tracking-database-archive. Accessed 04 Aug. 2024.

“Edgar.” *SEC.Gov | EDGAR Full Text Search*, www.sec.gov/edgar/search/. Accessed 04 Aug. 2024.