Litigating the Lipstick Index

BIOST557

Team Members
Ananya Bajaj, Sue Boyd

Abstract

In this study, we investigate the level at which 30 different retail trade categories respond to or withstand economic fluctuations, and the factors that influence sales in those retail categories. The three questions addressed are: (1) do some categories of spending, as a percent of total spending, statistically increase or decrease during a recession, (2) are some categories of spending positively or negatively correlated with increases in Gross Domestic Product (GDP), and (3) what factors best predict the overall level of sales in selected categories? We utilize a two-sample mean test and simple linear regression to identify the effect of recessions periods and the percent change in GDP has on retail sales, respectively, for Questions 1 and 2. We conclude that some categories do behave differently during recessionary periods or in response to short term changes in GDP. Notably, some categories of spending *increase* (as a percent of total spending budget) during economic downturns. This is broadly consistent with the theory behind the "lipstick index" which suggests that consumers overweight spending in categories they may consider to be necessities or "small luxuries" during these periods. In Question 3, we construct the best predictor of future sales using a multivariate approach. We conclude that the best predictors of future sales vary by category. In all categories, including recession and/or changes in GDP enhances the predictive power of the model.

Table of Contents

1 Introduction

1.1 Background

The lipstick index, as defined by Estée Lauder's Leonard Lauder, discusses the possibility of increased sales of small and affordable luxuries during economic downturns. While the term "lipstick index" was coined to demonstrate the increased purchase of lipsticks instead of more expensive apparels during the early 2000s recession, it later became widely known as an economic indicator. Currently, it covers a wide array of markets mostly within the beauty and personal care industry such as fragrance, skincare, jewelry, clothing and more. Building on the idea that consumers change their spending habits in response to economic events, we consider whether consumers reallocate their budget towards (or away from) certain categories of spending during economic downturns. We also consider what variables best predict spending in future quarters. We investigate spending across 30 categories of retail spending.

1.2 Project Aims and Questions

The objectives of this study are to investigate the impact of recessions on different retail trades, and sales trends with fluctuations in economic performance with the hopes of identifying recession-proof and recession-prone retail trades. The questions addressed in this study are: (1) do some categories of spending, as a percent of total spending, statistically increase or decrease during a recession, (2) are some categories of spending, as a percent of total spending, positively or negatively correlated with increases in Gross Domestic Product (GDP), and (3) what factors best predict the overall level of sales in selected categories?

2 Methods

2.1 Data Description

2.1.1 Data Sources and Collection

That data for this study was obtained from the United States Census Bureau and the Research Department at the Federal Reserve Bank of St. Louis (FRED). The United States Census Bureau, who provides the Retail and Food Services Sales Excel (1992 - Present) dataset, collects data using the Advance Monthly and Monthly Retail Trade Surveys (MARTS and MRTS), the Annual Retail Trade Survey (ARTS), and the Quarterly E-Commerce Report to produce national estimates of sales for retail businesses that reside in the United States. FRED provides both the Dates of U.S. recessions as inferred by GDP-based recession indicator and the Gross Domestic Product (GDP) datasets. The recession indicator dataset is an interpretation of US Business Cycle Expansions and Contractions data published by the National Bureau of Economic Research (NBER). The GDP dataset was collected from reports published by the US Bureau of Economic Analysis. All datasets were obtained and downloaded as .csv or .xls files.

2.1.2 Study Design

We conducted an observational study of retail sales response to changes in GDP and recessions. We focused on observing retail sales over time in 30 of the 32 sales categories included in the Retail and

Food Services sales Dataset. We excluded the Jewelry and Men's Clothing categories as there are seven months of missing data in both categories. This study was restricted to retail business and economic changes within the US. Therefore, the population analyzed included retail businesses that reside in the US. These retail businesses were then grouped into categories.

2.1.3 Data Dictionary

The full data dictionary can be found in Section 6.1, *Additional Tables*.

2.1.4 Discussion

All datasets obtained from the FRED along with the data obtained from the United States Census Bureau were seasonally adjusted, excluding the Dates of U.S. recessions as inferred by GDP-based recession indicator dataset. Seasonal adjustment is a statistical technique used to better reflect changes in employment from month to month by identifying, measuring, and adjusting for influences in seasonal economic patterns (*What Is Seasonal Adjustment?*, 2001). This adjustment removes the seasonal influence on statistical trends. While these datasets account for seasonal fluctuations, they do not account for inflation or population growth over time. Failure to account for inflation or population growth can result in misleading conclusions because these variables make it so that sales or GDP data may be growing for reasons unrelated to our variables of interest, as seen in Fig. 1. To help ensure inflation and population growth do not skew our results, we analyzed the percent of total sales for each retail category and the percent change in GDP for Questions 1 and 2. For Question 3, we used a multivariate approach.

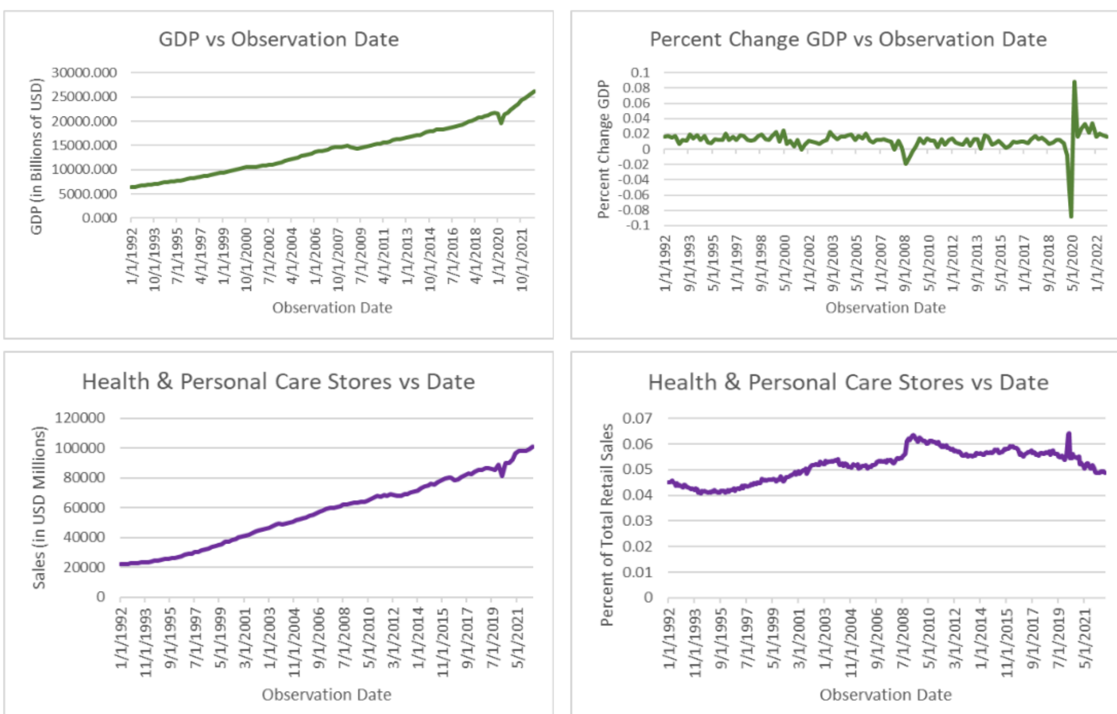


Figure 1: Original Health & Personal Care Stores retail sales and GDP data as compared to the calculated fields that reflect economic changes over time.

As all datasets used are time series data, the assumption that the responses (retail sales) are independent cannot be made as they do not hold. The issue of autocorrelation is visualized in Fig. 2. These graphs depict the autocorrelation function (ACF) values of the following retail categories: Automotive parts, acc., and tire stores, and Electronic shopping and mail order houses for seasonally and not seasonally adjusted sales, along with seasonally adjusted percent of total sales. Autocorrelation measures the relationship between multiple values of a variable over set time periods. As the ACF values are close to 1 in each category regardless if they are seasonally adjusted, this indicates the variables are highly dependent. We assume independence for all tests in this study, therefore, the clear violation of this assumption will most likely have an effect on the p-values and confidence intervals reported. Adding techniques to our models to address the autocorrelation is an important area for future research.

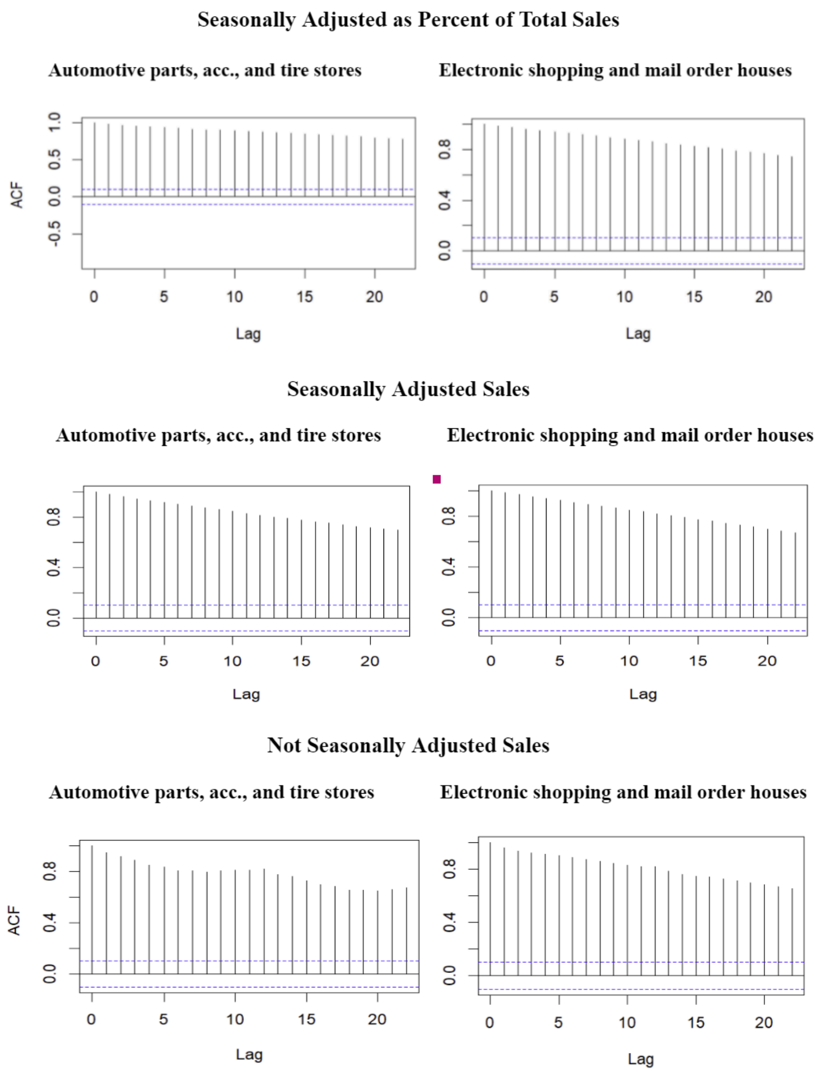


Figure 2: Comparison of seasonally and not seasonally adjusted sales, along with seasonally adjusted percent of total sales autocorrelation for the Automotive parts, acc., and tire store, and Electronic shopping and mail order houses retail categories

2.2 Statistical Methods

2.2.1 Question 1: Do some categories of spending as a percent of total spending statistically increases or decreases during a recession?

We started by combining the two datasets used in this section, Dates of US Recessions (Monthly) and Retail and Food Services Sales by matching the dates of the U.S. recessions dataset with the dates of the Retail and Food Services Sales dataset. We also cleaned the Retail and Food Sales dataset by updating values indicated as "(S)" to "NA" and transforming the data to include the percent of total spending for each category. As mentioned in Section 2.1.2, the Jewelry and Men's clothing retail categories were excluded due to many missing values. Line and box plots were produced to visualize the percent of spending for each category over time, highlighting recession and non-recession periods.

For this question, we defined the null hypothesis to be the mean spending, as a percent of total spending, in a particular category during a recession is equal to the mean percent of total spending in the same category outside of a recession. The alternate hypothesis was that the mean spending in a particular category, as a percent of total spending, during a recession was different from the mean percent of total spending in the same category outside of a recession.

Welch's two-sample t-test was used to compare the mean percent spending in a particular category during and outside of a recession. The mean spending as a percent of total spending in each category during and outside of a recession was calculated, along with the standard error of the difference. After running the two-sample t-test in R, the test statistic and p-value was obtained. We did not adjust for multiple testing and instead used a significance level of 0.05 and p-values to determine if the null hypothesis could be rejected. A 95% confidence interval was also calculated for the difference in mean percentages to estimate the magnitude of the effect. The following are the underlying assumptions:

Independence:	Our model assumes that each observed monthly sales (as a percent of total) in a particular category is independent from other observed monthly sales (as a percent of total). As noted above, this assumption may not hold.
Variance:	In using the Welch's two-sample t-test, we allow for unequal variance to ensure robustness of our tests.
Normality/Sample Size	The sample sizes are sufficiently large (n=370), and the central limit theorem can be used to justify the normality of sample means.

2.2.2 Question 2: Are some categories of spending positively or negatively correlated with increases in GDP?

We began by cleaning the data as noted in the previous section. We then calculated Quarterly, Seasonally Adjusted, Retail and Food Services Sales Data in each sales category, by summing the three months of sales data relevant to that quarter. The Quarterly, Seasonally Adjusted, Retail and Food Services Sales Data was combined with the Quarterly Adjusted GDP Data by matching the dates of the two datasets. The same 30 retail categories, as previously mentioned, were used in this question. The percent change in GDP variable was calculated as described in Table 1.

For each category of spending, we tested the null hypothesis that the difference in estimated mean percent of total sales per 1 unit difference in percent change GDP (for a given retail trade) is zero against the alternative hypothesis that the difference in estimated mean percent of total sales per 1 unit difference in percent change GDP (for a given retail trade) is not zero. We ran 30 Partial F-tests, utilizing simple linear regression, with the percent of total retail sales as the response and the percent change in GDP as the predictor. We used a statistical significance level of 0.05 and did not adjust for multiple testing. The estimated β coefficients were calculated along with the p-values to determine whether to reject or fail to reject the null hypothesis. 95% confidence intervals were calculated along with R^2 values to estimate the range of the β coefficients and effectiveness of the model, respectively. The following are the underlying assumptions:

Independence:	Our model assumed that sample observations were independent. As noted above, this assumption may be violated, thereby undermining our p-values and confidence intervals.
Linearity:	Our model assumed that expected sales as a percent of total sales are linearly related to percent change in GDP.
Variance	Our model in Question 2 assumed constant variance or homoscedasticity.
Normality/Sample Size	Our sample size (n=123 quarters) was sufficiently large that we did not need to rely on an underlying assumption of normality of the residuals in order to predict the β coefficients; rather we appealed to the Central Limit Theorem. In this question, predicting future quarters was not a goal. Prediction is explored in Question 3.

2.2.3 Question 3: What factors best predict the overall level of sales in selected categories?

In Question 3, we focused on constructing the best predictor of future sales from the data available to us. To do this, we switched back from looking at sales as a percent of total sales, and instead focused on the total level of sales. Data was cleaned and prepared as for Question 2 above. In addition, we calculated a variable for "Time" and "Covid", as described in the data dictionary. Finally, sales data was converted to USD in Billion and GDP data was converted to USD in Trillions to make the results more interpretable.

First, for each category of spending, we fit a "Baseline Model" from the data using Linear Regression with GDP as the predictor variable and total sales in that quarter as the dependent variable. We chose GDP as the predictor for the Baseline Model because data exploration revealed that GDP was highly correlated with sales across all categories. We then fit Expanded Models, adding in one or more additional predictor variables. We considered the following additional predictor variables as our exploratory data analysis suggested they may have predictive value (total of 15 combinations):

- COVID = Indicator Variable (1 for Q2_2020 and 0 otherwise)
- Time = Number of Quarters Elapsed Since 1992_Q1
- Percent_Change_GDP = Quarter-over-Quarter Change in Percent GDP
- Rec = Quarterly Recession Indicator

We then selected as the "Best" model the model with the highest adjusted R^2. The selected model was permitted to (and did) vary across different spending categories. The Baseline Model was compared to the Best model by testing the null hypothesis that the β coefficient for all predictors, *except* the coefficient for GDP, was equal to zero. The null hypothesis was rejected at the 5% significance level if the p-value of the F statistic with k, n-k-2 degrees of freedom, was less than 0.05 (where n is the sample size, and k is the number of *additional* predators in the multivariate model, beyond GDP). We also checked whether our results were robust to relaxation of the equal variance assumption by also performing a Wald test, with unequal variance, using a significance test of 0.05.

Finally, for selected categories of interest, we used visual diagnostics to further assess the model, underlying assumptions, and potential limitations:

Independence:	As in Questions 1 and 2, we assumed independence of sample observations - an assumption that may not hold. For selected categories of interest, we graphed the autocorrelation of model residuals to assess the level of autocorrelation that remains in the data even accounting for the predictive values in the model. Below, we use these observations to discuss potential impact/limitations on our findings and issues for future research.
Linearity:	Our model assumed that expected sales were linearly related to the predictor variables.
Variance:	Standard linear regression techniques assume equal variance of the residuals. But we also tested our hypotheses relaxing the equal variance assumption.
Normality/Sample Size	Our sample size (n=123 quarters) is sufficiently large that we did not need to rely on an underlying assumption of normality of the residuals in order to predict the β coefficients; rather we appealed to the Central Limit Theorem. For predicting future spending, however, the assumption of normality of the residuals would be required. For selected categories of interest, we produced and assessed QQ plots to evaluate the reasonableness of this assumption.

3 Results

3.1 Question 1

Based on the results in Table 6, we can see that there is a statistically significant difference in mean percent of total sales during recession periods compared to non-recession periods, as indicated by the p-value column, for 14 retail categories. For instance, automobile and motor vehicle dealers had a mean percent of total retail sales significantly lower than its mean in non-recessionary periods. Conversely, for categories like Gasoline and Health and personal care, there is a higher mean in the recession periods compared to non-recession periods. A mean difference value less than 0 indicates that that particular category performed better during a recession. Again, p-values should be interpreted with caution as they may not be accurate due to lack of independence.

Table 6: Two-sided T-test output using percent of total retail sales and a recession indicator ("Yes" indicating recession period, "No" indicating not a recession period) by retail category.

Retail Category	Mean No	Mean Yes	Mean Diff	95% CI	P-Value
All other gen. merchandise stores	0.0119	0.0111	0.0008	(4.529e-06, 1.509e-03)	0.0487
Automobile and motor vehicle dealers	0.1949	0.1743	0.0206	(0.0068, 0.0343)	0.00487
Automotive parts, acc., and tire stores	0.0181	0.0178	0.0003	(-0.0001, 0.0008)	0.1429
Beer, wine and liquor stores	0.0091	0.0093	-0.0002	(-0.0007, 0.0001)	0.1763
Build. mat., garden equip., supp dealers	0.0665	0.0691	-0.0026	(-0.0038, -0.0015)	1.778e-05
Building mat. and supplies dealers	0.0577	0.0601	-0.0024	(-0.0034, -0.0014)	1.511e-05
Clothing and clothing access. stores	0.0491	0.0473	0.0018	(-0.0018, 0.0055)	0.3109
Clothing stores	0.0351	0.0343	0.0008	(-0.0019, 0.0035)	0.5622
Department stores	0.0515	0.0499	0.0016	(-0.0041, 0.0074)	0.5659
Electronic shop. and mail order houses	0.0589	0.0552	0.0037	(-0.0084, 0.0159)	0.5343
Electronics and appliance stores	0.0227	0.0241	-0.0014	(-0.0030, 0.0002)	0.0845
Food and beverage stores	0.1369	0.1373	-0.0004	(-0.0052, 0.0043)	0.8476
Food services and drinking places	0.1049	0.1018	0.0031	(-0.0006, 0.0069)	0.0999
Fuel dealers	0.0073	0.0086	-0.0013	(-0.0019, -0.0007)	0.0002
Furniture and home furnishings stores	0.0232	0.0229	0.0003	(-0.0011, 0.0019)	0.5767
Furn., home furn, elec., and app. stores	0.0459	0.0469	-0.001	(-0.0039, 0.0019)	0.4907
GAFO(1)	0.2467	0.2575	-0.0108	(-0.0192, -0.0024)	0.0132
Gasoline stations	0.0858	0.0941	-0.0083	(-0.0163, -0.0003)	0.0436
General merchandise stores	0.1248	0.1356	-0.0108	(-0.0139, -0.0075)	1.296e-07
Grocery stores	0.1234	0.1236	-0.0002	(-0.0046, 0.0042)	0.931
Health and personal care stores	0.0515	0.0562	-0.0047	(-0.0069, -0.0024)	0.0002
Miscellaneous stores retailers	0.0263	0.0264	-0.0001	(-0.0013, 0.0013)	0.954
Motor vehicle and parts dealers	0.2130	0.1921	0.0209	(0.0071, 0.0346)	0.0043
Nonstore retailers	0.0771	0.0746	0.0025	(-0.0092, 0.0142)	0.6612
Other general merchandise stores	0.0732	0.0857	-0.0125	(-0.0201, -0.0047)	0.0025
Pharmacies and drug stores	0.0435	0.0483	-0.0048	(-0.0069, -0.0027)	4.953e-05
Shoe stores	0.0068	0.0061	0.0007	(0.0002, 0.0012)	0.0037
Sporting goods, hobby, musical instrument, and book stores	0.0181	0.0186	-0.0005	(-0.0016, 0.0005)	0.3046
Warehouse clubs and superstores	0.0614	0.0746	-0.0132	(-0.0203, -0.0059)	0.0008
Women's clothing stores	0.0091	0.0085	0.0006	(-0.0002, 0.0013)	0.1277

*Retail categories highlighted in green indicate the null hypothesis was rejected when sales during recessions had a higher mean. Retail categories highlighted in purple indicate the null hypothesis was rejected when sales during a non-recession period had a higher mean.

Looking specifically at the Health and Personal Care Stores retail category, we see an increase in sales for all recessionary periods which is also supported by the box plot that demonstrates an increase in mean percent of total spending in the recessionary period, as seen in Fig 3. We also see that for the Shoe Stores retail category, there was a dip in the 2001 and 2020 recession periods, whereas it saw little change during the 2008 recession. However, the overall mean percent of total spending dipped for shoe stores during recessionary periods.

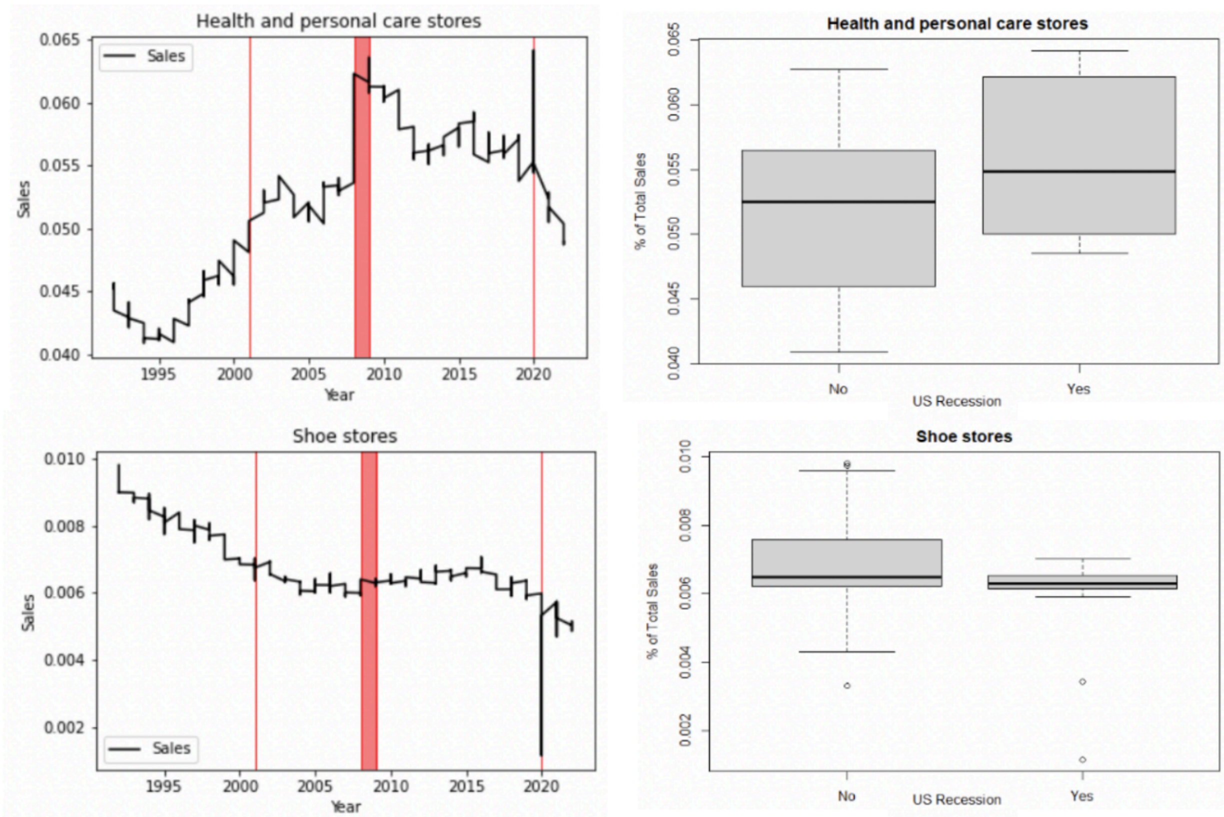


Figure 3: Line plots indicated with recession periods indicated in a red vertical line, and corresponding box plots indicating mean value of the percent of total spending in recession and non-recession periods for categories.

3.2 Question 2

After running Partial T-test using simple linear regression, we obtained the estimated β coefficients, 95% confidence intervals, p-values and R^2 as shown in Table 7. Before interpreting these results, it's important to note the extremely low R^2 values. The largest R^2 value is 0.122 (a statistic for the General merchandise stores model). These values indicate that all models do not explain much of the variance in the response and Percent Change GDP might not be effective as a predictor alone. Only eight of the 30 models reject the null hypothesis as compared to Question 1, where 14 rejects the null hypothesis. Of those eight, the Automotive and motor vehicle, and Motor vehicles and parts dealers retail categories were the only categories that had a slight positive correlation with increased percent change in GDP (highlighted in purple), indicating consumers allocate more of their budget to these categories when the economy experiences growth. This aligns what was observed in the previous question where these same retail categories saw a higher mean during non-recession periods. The remaining six retail categories that saw a slight negative correlation with percent change in GDP (highlighted in green) indicate that consumers allocate more of their budget to these retail categories during economic downturns. It is important to note that the p-values and confidence intervals recorded in this table may not be accurate due to the violation of independence.

Table 7: Estimated β Coefficient, 95% Confidence Intervals, P-Values and R^2 output to simple linear regression models for each retail category

Retail Category	Estimated Beta Coef.	95% CI	P-value	Model R^2
All other gen. merchandise stores	-0.0110	(-0.0317, 0.0096)	0.2921	0.0092
Automobile and motor vehicle dealers	0.3470	(0.0498, 0.6442)	0.0225	0.0423
Automotive parts, acc., and tire stores	-0.0008	(-0.0274, 0.0256)	0.9485	3.457e-05
Beer, wine and liquor stores	-0.0101	(-0.0184, -0.0019)	0.0162	0.0468
Build. mat., garden equip., supp dealers	0.0093	(-0.0689, 0.0876)	0.8139	0.0005
Building mat. and supplies dealers	0.0099	(-0.0636, 0.0835)	0.7894	0.0006
Clothing and clothing access. stores	0.0658	(-0.0050, 0.1366)	0.0682	0.0272
Clothing stores	0.0320	(-0.0208, 0.0849)	0.2327	0.0118
Department stores	0.1123	(-0.1959, 0.4206)	0.4719	0.0043
Electronic shop. and mail order houses	-0.0327	(-0.5099, 0.4444)	0.8922	0.0002
Electronics and appliance stores	0.0109	(-0.0599, 0.0818)	0.7603	0.0008
Food and beverage stores	-0.0953	(-0.3084, 0.1177)	0.3776	0.0064
Food services and drinking places	0.0241	(-0.1176, 0.1657)	0.7374	0.0009
Fuel dealers	-0.0103	(-0.0284, 0.0078)	0.2646	0.0103
Furniture and home furnishings stores	0.0443	(-0.0029, 0.0915)	0.0657	0.0277
Furn., home furn, elec., and app. stores	0.0552	(-0.0586, 0.1691)	0.3389	0.0076
GAFO(1)	-0.0709	(-0.3731, 0.2312)	0.6429	0.0018
Gasoline stations	-0.1059	(-0.3098, 0.0978)	0.3054	0.0087
General merchandise stores	-0.2176	(-0.3225, -0.1125)	0.0001	0.122
Grocery stores	-0.0836	(-0.2869, 0.1196)	0.4167	0.0055
Health and personal care stores	-0.1054	(-0.1839, -0.0270)	0.0088	0.0553
Miscellaneous stores retailers	0.0352	(-0.0211, 0.0917)	0.2184	0.0125
Motor vehicle and parts dealers	0.3462	(0.0347, 0.6575)	0.0296	0.0385
Nonstore retailers	-0.0236	(-0.4725, 0.4252)	0.9172	8.972e-05
Other general merchandise stores	-0.3299	(-0.6410, -0.0187)	0.0379	0.0351
Pharmacies and drug stores	-0.1058	(-0.1714, -0.0401)	0.0018	0.0776
Shoe stores	0.0105	(-0.0034, 0.0245)	0.1389	0.0180
Sporting goods, hobby, musical instrument, and book stores	0.0121	(-0.0246, 0.0488)	0.5151	0.0035
Warehouse clubs and superstores	-0.3189	(-0.6279, -0.0098)	0.0433	0.0333
Women's clothing stores	0.0163	(-0.0168, 0.0494)	0.3313	0.0078

*Retail categories highlighted in purple indicate the null hypothesis was rejected with a positive correlation. Retail categories highlighted in green indicate the null hypothesis was rejected with a negative correlation.

To better understand the statistical interpretation of the results in the table above, Fig. 4 visualizes the Perchange Change in GDP plotted against the Percent of Total Retail Sales for the Automobile and other motor vehicles retail category for each quarter. Although our results indicate there exists a slight positive correlation, this graph shows there is a lot of variation in the data as also seen in the R^2 value. The same can be seen in Fig. 5, but this time, we are examining a negative correlation between Percent Change in GDP and Percent of Total Retail Sales for Other general merchandise stores. The addition of other predictors could potentially account for more variation in the data and strengthen the models, and is what is investigated in Question 3.

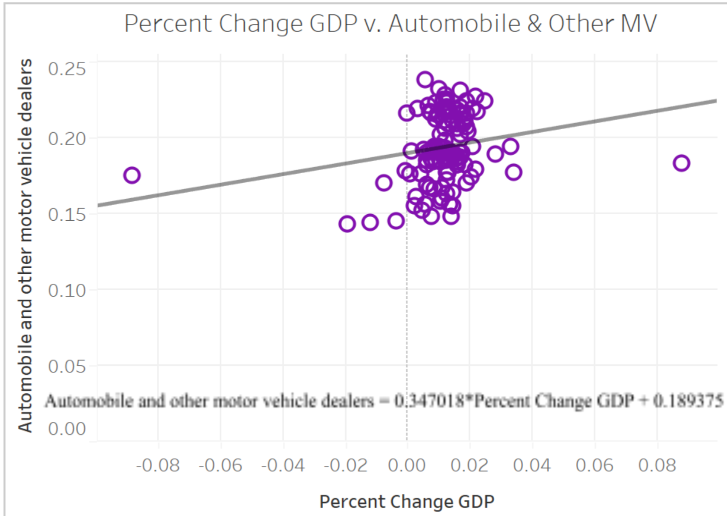


Figure 4: Scatter plot of Percent Change GDP plotted against the Percent of Total Sales in the Automobile and Other Motor Vehicles by quarter.

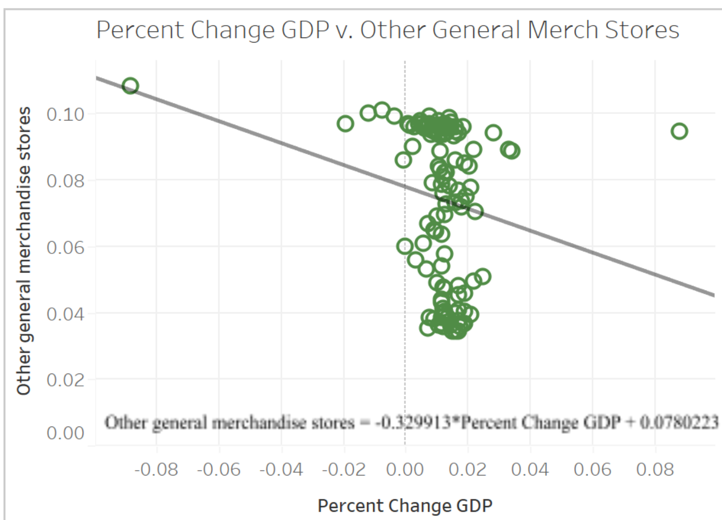
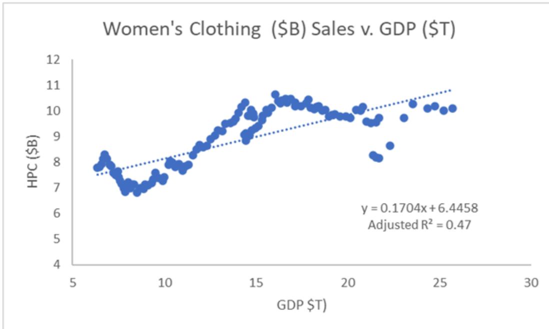


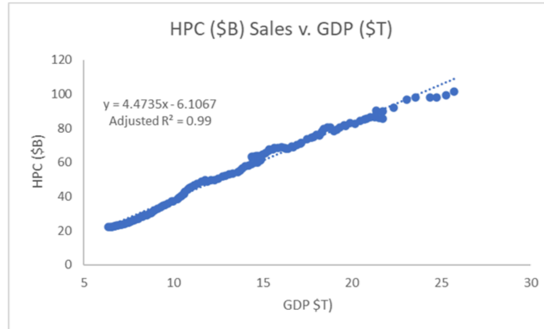
Figure 5: Scatter plot of Percent Change GDP plotted against the Percent of Total Sales in the Other General Merch Stores by quarter.

3.3 Question 3

3.3.1 Baseline Model

For each category of spending, GDP alone is a strong predictor of sales. Across all 30 categories, the β coefficient for the GDP variable was significant at the 0.05 level using a t-statistic. Adjusted R^2 varied from a low of 0.47, in Women's clothing, to a high of 0.99, in Health and personal care stores (HPC). Fig. 6 demonstrates the effectiveness of GDP as a predictor for the Women's clothing and Health and personal care stores retail categories.





Regression Estimates- Women's Clothing

	Est	95%CI	P-val
Intercept	6.44	(5.96, 6.93)	<2e-16
GDP	0.17	(.14, .20)	<2e-16

Regression Estimates- HPC

	Est	95%CI	P-val
Intercept	-6.10	(-7.44, -4.78)	<2e-15
GDP	4.47	(4.38, 4.56)	<2e-16

Figure 6: Baseline Model Regression Results, Women's Clothing and Health & Personal Care

3.3.2 Additional Predictors

After creating a Baseline Model for each category, we considered additional potential predictors. Fig. 7 graphs both GDP and Women's Clothing Sales Quarterly, over time. Women's Clothing Sales trends upward over time, but the trend is less dramatic than the upward trend in GDP. This suggests potential long-term trends in the Women's Clothing category, separate and apart from the impact of GDP changes over time. The graph also reveals a dramatic dip in sales during Q2_2020, corresponding with the period in which many states and localities had issued stay-at-home orders for COVID. These two variables, Time and COVID, appeared relevant in many categories based on preliminary exploratory analysis. Results from Questions 1 and 2 also suggested that percent change in GDP and recession indicator variables might be relevant predictors.

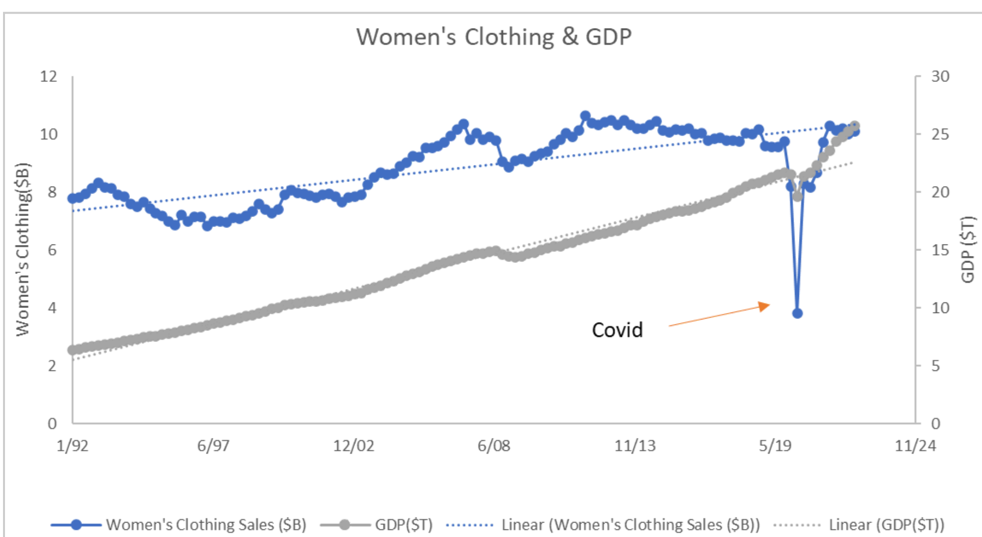


Figure 7: Chart of Women's Clothing Sales & GDP Over Time

3.3.3 Best Model for Each Category

For each category, we fit multiple Expanded Models using various combinations of the additional predictors. We selected the "Best" model to be the one with the highest Adjusted R^2.

Table 8: Adjusted R^2 By Model in Selected Categories

Model	Auto. & other MV	Auto. parts & tires	HPC	W. cloth
GDP	0.8557	0.9746	0.9878	0.4665
GDP, REC	0.8653	0.9751	0.9881	0.4829
GDP, PerChange	0.8715	0.9751	0.9881	0.4895
GDP, PerChange, REC	0.8725	0.9751	0.9881	0.4906
GDP, Time	0.8748	0.9749	0.9951	0.4675
GDP, Time, REC	0.881	0.9752	0.9952	0.4876
GDP, Time, PerChange	0.8808	0.975	0.9953	0.5065
GDP, Time, PerChange, REC	0.8825	0.9751	0.9955	0.5068
GDP, Covid	0.8553	0.9745	0.9877	0.655
GDP, Covid, REC	0.8642	0.9751	0.988	0.6524
GDP, Covid, PerChange	0.8773	**0.9758**	0.9885	0.6847
GDP, Covid, PerChange, REC	0.8779	0.9758	0.9885	0.6928
GDP, Covid, Time	0.8738	0.9748	0.9957	0.6874
GDP, Covid, Time, REC	0.881	0.9753	**0.9959**	0.6859
GDP, Covid, Time, PerChange	0.8866	0.9757	0.9957	0.7025
GDP, Covid, Time, PerChange, REC	**0.8878**	0.9757	0.9958	**0.7094**

The Best model varied by category. For example, in Table 8, the set of predictors that provide the highest Adjusted R^2 in the Automobile, part & tires category was the model that included only GDP, Covid indicator, and Percent Change GDP as predictors. For many, but not all categories, the highest Adjusted R^2 results from the model using all predictors, as shown in Fig. 8.

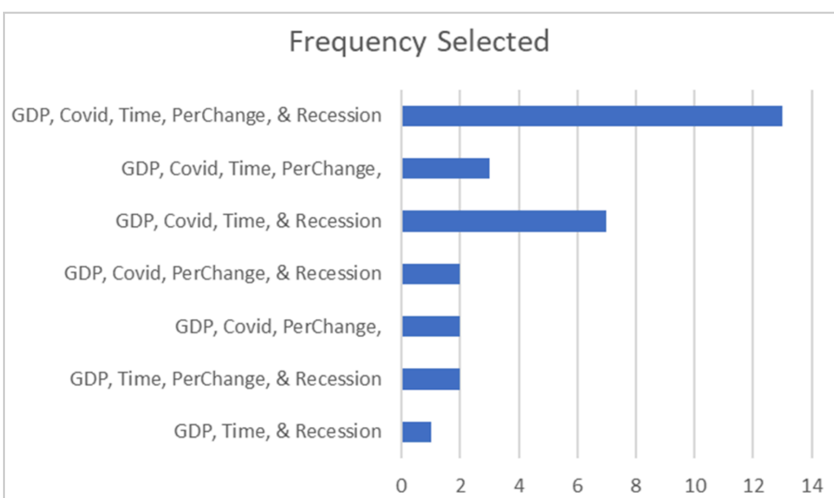


Figure 8: Best Models Selection Frequency

Notably, for all spending categories, the Best model included either, or both, of the variables Percent Change GDP and Recession Indicators as additional predictors. These results suggest that spending patterns fluctuate in response to *changes* in economic indicators, beyond a simple relationship based solely on the *level* of GDP. This is consistent with our findings in Questions 1 and 2 that consumers do re-allocate their budgets towards, or away from, certain categories of spending in response to economic indicators.

For all categories, the selected Best model had a higher Adjusted R^2 than the Baseline model, as shown in Fig. 9. Thus, the multivariate model better explains the variance in the data and therefore could be expected to better predict sales in future quarters (assuming model assumptions otherwise hold). In some categories, such as Women's clothing, the improvement in R^2 was dramatic. In other categories, such as Health and personal care stores, the Baseline model was already a strong fit and the increase in Adjusted R^2 was quite small.

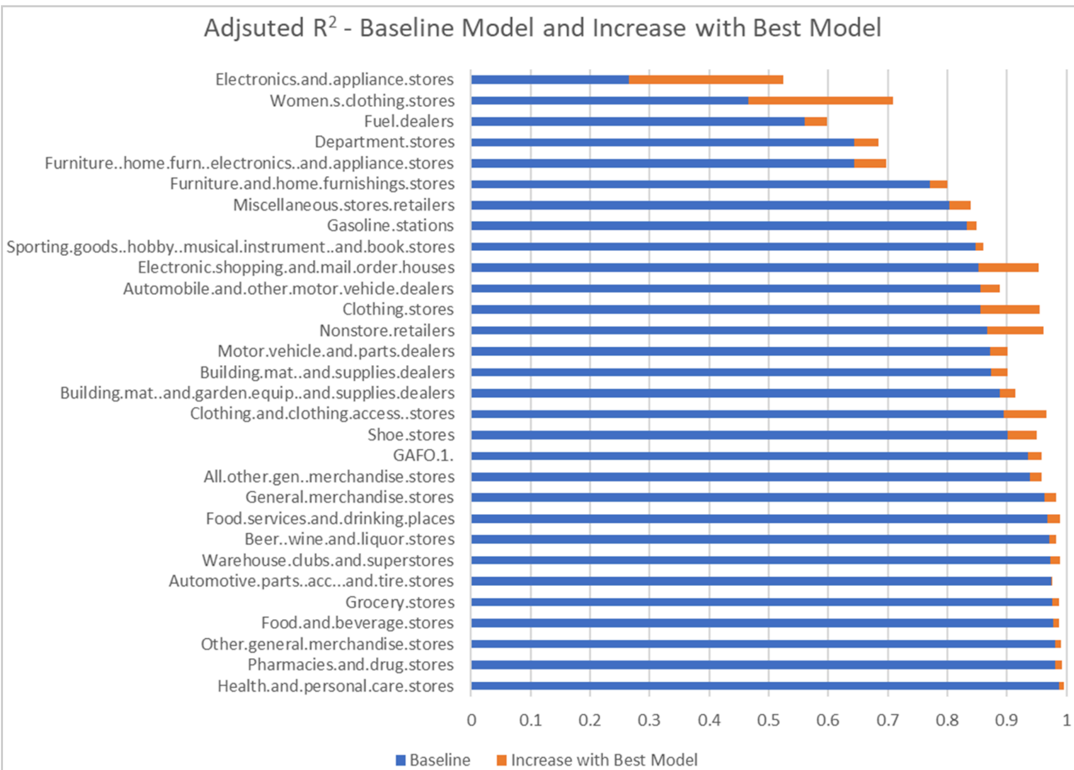


Figure 9: Adjusted R^2 Baseline and Best Model by Category

3.3.4 Hypothesis Testing

We compared the Baseline Model to the Best model in each category by testing the null hypothesis that the β for all predictors, *except* the coefficient for GDP, equaled zero. In **all** categories, we rejected the null hypothesis at a significance level of 0.05. We rejected the null hypothesis regardless of whether we used a partial F test, assuming constant variance, or a Wald test, with robust variance. These results suggest that the multivariate model better predicts future sales because it includes additional statistically relevant variables. Detailed results for selected categories provided in Fig. 10 and 11.

Women's Clothing – Baseline Model
$E[Sales|GDP] = \beta_0 + \beta_1 * GDP$

	Est	95%CI*	P-val*
Intercept	6.44	(5.96, 6.93)	<2e-16
GDP	0.17	(.14, .20)	<2e-16

$H_0 : \beta_2 = 0 = \beta_3 = \beta_4 = \beta_5 = 0$
P-value (Multiple F Test, constant Variance) = 1.5 e-15
P-value (Wald Test, Robust Var) = <2e-16

Reject Null Hypothesis At 0.05 Significance Level

Women's Clothing Best Model
$E[Sales|GDP, Covid, Time, PerChange, Rec] =$
$\beta_0 + \beta_1 * GDP + \beta_2 * Covid + \beta_3 * Time + \beta_4 * PerChange + \beta_5 * Rec$

	Est	95%CI*	P-val*
Intercept	7.97	(6.96, 8.98)	<2e-16
GDP	-.078	(-.27, .11)	.420
COVID	-8.49	(-10.33, -6.65)	2.49e-15
Time	.038	(0.01, 0.06)	.006
PerChange	-23.00	(-37.03, -8.97)	.002
Recession	-.48	(-.96, 0.0)	.053

*P-values and Confidence Intervals for individual slope coefficients assume constant variance and no Bonferroni adjustment

Figure 10: Women's Clothing Baseline v. Best Model.

HPC– Baseline Model
$E[Sales|GDP] = \beta_0 + \beta_1 * GDP$

	Est	95%CI	P-val
Intercept	-6.10	(-7.44, -4.78)	<2e-15
GDP	4.47	(4.38, 4.56)	<2e-16

$H_0 : \beta_2 = \beta_3 = \beta_4 = 0$
P-value (Multiple F Test, constant Variance) = <2e-16
P-value (Wald Test, Robust Var) = <2e-16

Reject Null Hypothesis At 0.05 Significance Level

HPC - Best Model
$E[Sales|GDP, Covid, Time, Rec] =$
$\beta_0 + \beta_1 * GDP + \beta_2 * Covid + \beta_3 * Time + \beta_4 * Rec$

	Est	95%CI	P-val
Intercept	9.13	(6.97, 11.30)	1.4e-13
GDP	1.51	(1.13, 1.91)	5.1e-12
COVID	-7.35	(-10.46, -4.23)	8.0e-6
Time	.42	(.37, .48)	2.63-29
Recession	1.17	(.21, 2.12)	.017

*P-values and Confidence Intervals for individual coefficients assume constant variance and no Bonferroni adjustment

Figure 11: HPC Baseline v. Best Model.

3.3.5 Visual Diagnostics

We performed visual diagnostics on the residuals in four categories, selected to be representative of the range of results: (1) Auto. & Other Motor Vehicles; (2) Automotive Parts & Accessories; (3) Health and Personal Care; and (4) Women's Clothing. Charts included in Section 6.2, Fig. 12(a) and 12(b). Analysis of QQ plots would suggest that the assumption of normality in the residuals is *somewhat* reasonable in the Best model across these four categories. But analysis of the autocorrelation of the residuals tells a different story. The Best (multivariate) model exhibited lower autocorrelation in the residuals in three of the four categories, but even the Best model residuals were highly autocorrelated. In the fourth category, Women's Clothing, both the Baseline and Best (multivariate) model had similar, and highly autocorrelated, residuals. As noted, autocorrelated residuals can bias p-values, confidence intervals, and predictive power of the models.

4. Discussion

Our analysis set out to investigate how retail spending responds to macroeconomic events, such as recession or changes in GDP. Results across all questions suggest that different spending categories react differently to these stimuli.

In Question 1, we concluded that the mean spending as a percent of total spending was statistically different between recessionary and non-recessionary periods for 14 of the 30 categories investigated. One limitation in our analysis relates to the limited number of recessionary periods in our data set. There were 28 recessionary months out of 370 total monthly observations, however, they were clustered in three groups (April 2001 - Nov 2001, Jan 2008 -June 2009, and Mar 2020 - April 2020).

To address the relatively small number of recessionary periods in our data set, in Question 2, we also analyzed how percent-spending correlated with changes in GDP, using quarterly data. Our research aim was similar to our aim in Question 1 – identify those categories of spending that increase or decrease (as a percent of total spending) in response to short-term changes in economic prospects. Using this approach, we concluded that eight categories of spending were correlated with changes in GDP to a statistically significant extent. With one exception (the Beer, Wine, and Liquor category), each of these eight categories were also identified in Question 1 as having statistical differences between recessionary and non-recessionary periods. This makes sense. We expect spending patterns to react similarly to short term changes in GDP as they do to formal recessions. In six categories, percent-spending was negatively correlated with changes in GDP, suggesting that consumers allocate *more* of their budget towards those categories during economic downturns. Again, with the exception of the Beer, Wine, and Liquor category, all of the categories we concluded were significantly negatively correlated with changes in GDP were also identified in Question 1 as categories that had an *increase* in mean percent spending in Question 1, as we would expect.

These results are broadly consistent with the theory behind the “lipstick index” in that they suggest that consumers consider certain spending categories to be necessities or “small luxuries” that can be afforded even during recessions or downturns, and thus shift spending toward those categories during such periods. In particular, percent-spending in the “Health and Personal Care” category, the closest proxy in our dataset for the type of good included in the lipstick index, was indeed negatively correlated with changes in GDP (Question 2) and increased during recessionary periods (Question 1) as the lipstick index would predict.

In Question 3, we constructed the best predictor of future sales after considering GDP, a COVID-19 indicator, a recession indicator, number of quarters elapsed, and the percent change in GDP as potential predictors. We concluded that each of these variables has predictive value in some categories. As with Questions 1 and 2, different categories of spending behaved differently in response to changes in economic indicators.

An overarching limitation in our results, across all questions, relates to the autocorrelation in the sales data. This autocorrelation undermines the assumption that our sample observations (or residuals, in the case of regression analysis) are independent. Thus, p-values and confidence intervals in this report should be interpreted with caution. An important area for future research would be to apply specialized methods to account for autocorrelation in the sales data. Possible techniques could include adding one or more lagged dependent variables (LDV) as predictors in our regression analysis or expressly modeling residuals as autocorrelated.

5 References

Bureau, U. S. C. (2019, April 15). Monthly retail trade - sales report. United States Census Bureau. Retrieved February 20, 2023, from https://www.census.gov/retail/sales.html

COVID-19 lockdowns. (2021, February 13). Wikipedia; Wikipedia. https://en.wikipedia.org/wiki/COVID-19_lockdowns

Gross Domestic Product: Implicit Price Deflator. (2023, February 23). Stlouisfed.org; Federal Reserve Bank of St. Louis. https://fred.stlouisfed.org/series/GDPDEF

Hamilton, J. (2023, January 27). *Dates of U.S. recessions as inferred by GDP-based recession indicator*. FRED, Federal Reserve Bank of St. Louis. Retrieved February 20, 2023, from https://fred.stlouisfed.org/series/JHDUSRGDPBR

Last Day of Stay at Home Order in the United States. (n.d.). Www.timeanddate.com. Retrieved March 10, 2023, from https://www.timeanddate.com/holidays/us/last-day-of-lockdown

U.S. Bureau of Economic Analysis. (2023, January 26). *Gross Domestic Product [GDP]*. FRED, Federal Reserve Bank of St. Louis. Retrieved February 20, 2023, from https://fred.stlouisfed.org/series/GDP#0

What is seasonal adjustment? (2001, October 16). U.S. Bureau of Labor Statistics; United States Department of Labor. https://www.bls.gov/cps/seasfaq.htm

6 Tables and Figures

6.1 Additional Tables

Table 1: Description of variables used in the Gross Domestic Product dataset

Variable Display Name	Description	Unit	Format
observation date	The date at which the GDP was measured. Recorded quarterly.	Date	YYYY-MM-DD
Seasonally Adjusted GDP	Gross Domestic Product. The measure of the market value of all the final goods and services produced in the US in a given time period. Seasonally adjusted annual rate.	Billions of US dollars	Float
Percent_Change_GDP **(Calculated)**	This was calculated as the difference (current quarter GDP - prior quarter GDP) divided by the GDP of the previous quarter.	Percent	Float

Table 2: Description of variables used in the Dates of U.S. recessions as inferred by GDP-based recession indicator (Quarterly) dataset

Variable Display Name	Description	Unit	Format
observation date	The first day of the Quarter in which a recession was inferred. Recorded quarterly. In situations where a portion of the quarter is included in the recession, the whole period is deemed to be included in the recession period.	Date	YYYY-MM-DD
JHDUSRGDPBR	A recession indicator based on GDP.	Indicator	Boolean

Table 3: Description of variables used in the Dates of U.S. recessions as inferred by GDP-based recession indicator (Monthly) dataset

Variable Display Name	Description	Unit	Format
observation date	The first day of the Quarter in which a recession was inferred. Recorded monthly. In situations where a portion of the quarter is included in the recession, the whole period is deemed to be included in the recession period.	Date	YYYY-MM-DD
USREC	A recession indicator based on GDP.	Indicator	Boolean

Table 4: Description of variables used in the COVID Indicator dataset

Variable Display Name	Description	Unit	Format
COVID **(Self Calculated)**	This indicates the quarter (Q2 _2020) in which the COVID-19 pandemic resulted in the most stay-at-home orders.	Indicator	Boolean

Information regarding stay at home orders derived from Wikipedia's COVID-19 Lockdowns article and timeanddate's record of stay-at-home observances (*COVID-19 Lockdowns*, 2021; *Last Day of Stay at Home Order in the United States*, n.d.).

Table 5: Description of variables used in the Retail and Food Services Sales Excel (1992 - Present)

Variable Display Name	Description	Unit	Format	Additional Values
NAICS Code	Codes that classify business establishments standardized by Federal statistical agencies.	**Sector:** 2-digit code **Subsector:** 3-digit code **Industry Group:** 4-digit code **NAICS Industry:** 5-digit code **National Industry:** 6-digit code	Integer	GAFO represents stores classified in the following NAICS codes: 442, 443, 448, 451, 452, and 4532.
Kind of Business	Subcategories of retail trades.	N/A	String	
Estimate of Seasonally Adjusted Monthly Retail and Food Sales	Estimate of sales by month based on data from the Monthly Retail Trade Survey, Annual Retail Trade Survey, and Service Annual Survey.	Millions of US dollars	Integer	(S): "Suppressed" or estimates that do not meet publication standards because of high sampling variability (variability coefficient > 30%), low response rate, or concerns about estimate's quality.
Estimate of Percent Seasonally Adjusted Monthly Retail and Food Sales **(Calculated)**	Takes the Estimate of Monthly Retail and Food Sales by month and divides it by the total retail sales for that year.	Percent	Float	
Seasonally Adjusted Quarterly Sales **(Calculated)**	Calculated using the monthly sales summed to quarters. Q1 is the sum of Jan, Feb, March, etc.	Millions of US dollars	Integer	
Time **(Calculated)**	Number of quarters since 1992. This variable counts the number of quarters that have passed since Q1 1992.	Count	Integer	

6.2 Additional Figures

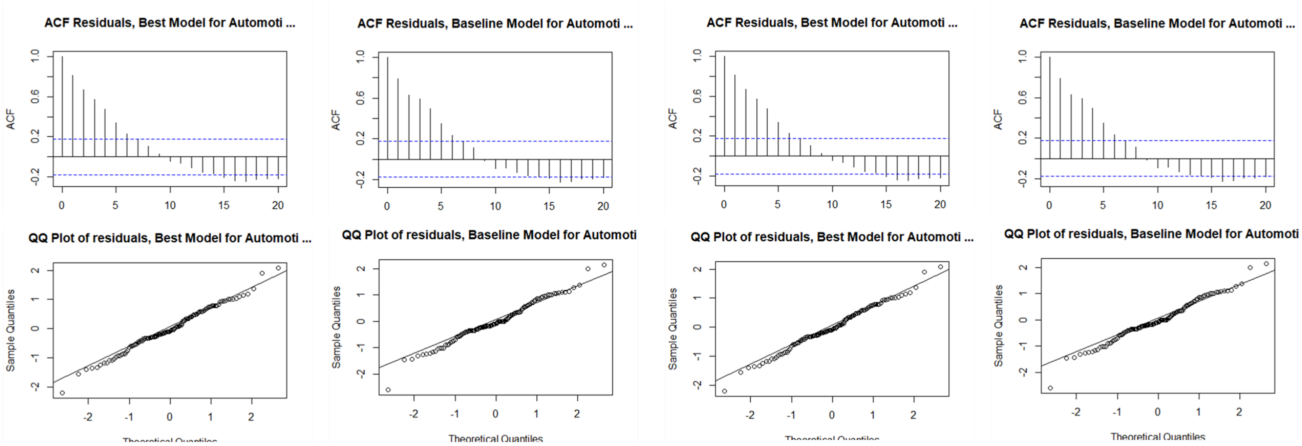


Figure 12(a): Visual Diagnostics of Residuals for Automobile & Other Motor Vehicles (Left) and Automobile, Tires & Accessories (Right)

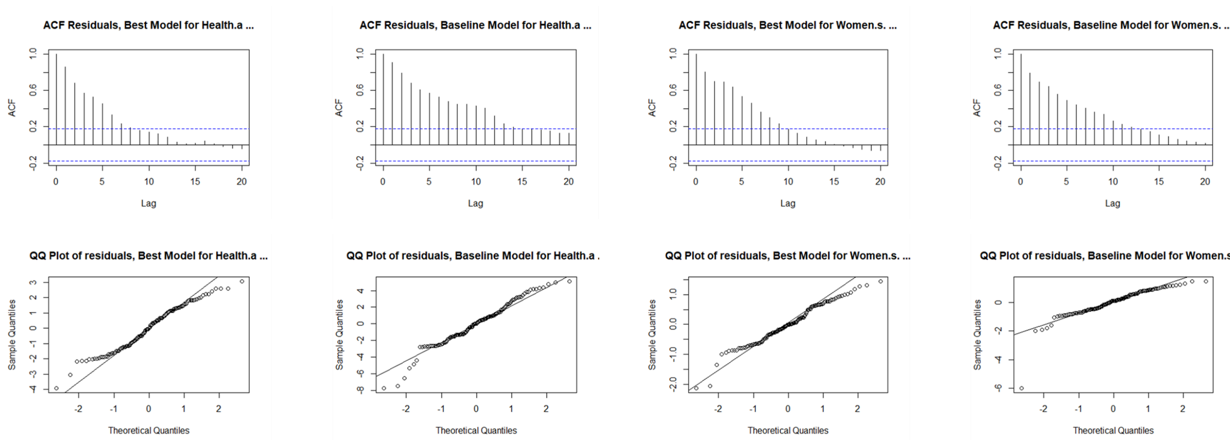


Figure 12(b): Visual Diagnostics of Residuals for Health and Personal Care (Left) and Women's Clothing Automobile, Tires & Accessories (Right)

7 Code

Here's the [Github Link]