Marlaina Wakim

Written Report

Do people spend less when inflation is higher? Inflation affects the economy and consumers in a variety of ways, including consumer confidence, demand for luxury goods, and purchasing power. Will those factors influence how much consumers are willing to spend around the holidays? My research question is "How does November inflation affect the U.S. Holiday Sales?"

For this study, I used Fred to find the CPI for each year from 2003 to 2024, specifically for November. I also used Capital One to find total holiday sales in billions from November 1st through December 31st for the same 21 years. Both of these sources seem reliable. FRED gets their data from the U.S. Bureau of Labor and Statistics and Capital One gets theirs from many credible sources such as Statista and The National Retail Federation. To find the year-over-year consumer price index (CPI) inflation, I subtracted the CPI from the previous year from the CPI from that year, divided the result by the CPI from the previous year, and multiplied by 100. My independent variable is inflation, and my dependent variable is Holiday Sales. My summary statistics are as follows: my counts are 21 for both Sales in Billions and CPI Year over Year. My mean for sales in billions is 653.8, and my mean for CPI YOY is 2.6. The standard deviation for sales in billions is 162.3, and 1.77 for CPI YOY. The minimum and maximum sales figures, in billions, are 467.2 and 984.3, respectively. The minimum and maximum for CPI YOY are 0.436 and 7.13, respectively.

I used a Generalized Linear Model to examine how year-over-year inflation extracted from the November Consumer Price Index related to holiday sales in the U.S. My model is $Sales=\beta0+\beta1(Inflation)+ \varepsilon$. Using my data, this becomes $Sales= 553.65 + 38.46(inflation) + \varepsilon$.

The y-intercept, also known as β0, which in this case is 553.65 billion, represents holiday sales in the U.S. when inflation is zero. β1 captures how much sales change for every 1% increase in inflation. Although I am using a linear regression model, there are limitations to my research because my data is a time series. Since our final project requires an Ordinary Least Squares model, I will proceed with my analysis while acknowledging this as a limitation of my model, then proceed to validate my research using methods taught to us in class.

Given my model, Sales=553.65 +38.46(CPI YOY)+ε, we can see that the coefficient on inflation is positive, meaning when inflation goes up, so do holiday sales. My p-value is 0.059, which is not statistically significant. Based on this p-value, there is insufficient statistical evidence to conclude that inflation is linked to holiday sales. My residual plot displays a roughly random scatter around zero, with no apparent pattern. This residual plot alone should suggest that the linearity and constant-variance assumptions are met. However, after further examination of my data using a Lag Plot, we can see a strong positive autocorrelation. Due to this violation of the Generalized Linear Model, I decided to use the "Double First Differences" approach to reduce autocorrelation in the error terms. With this approach, my new model is ΔSales=-26.13+6.06(ΔInflation)+ε, and my new p-value is 0.019, which is statistically significant because it is below our significance threshold of 0.05. This suggests that a 1% increase in the year-over-year change in inflation is associated with a 6.1 billion dollar increase in the year-over-year change in holiday sales. Although this relationship initially seems counterintuitive, it makes sense in real-world situations. If inflation is higher this year than last, you may try to save more by utilizing Black Friday and Cyber Monday deals than if you were not worried about prices. If inflation is higher than the previous calendar year, retailers might raise their prices to negate inflation; this might also lead to higher sales than the previous year.

In order to validate my research further, it is important to consider what other factors could be affecting both my independent and dependent variables. A factor that might contribute to the perceived relationship between my independent and dependent variable is the normal growth that holiday sales have each year, unrelated to inflation. The U.S. population is growing and so is online shopping among other things, so it only makes sense that holiday sales would also grow. To account for this, I made a control variable called 'Year Trend', this would turn my model into $\Delta Sales=\beta 0+\beta 1\Delta Inflation+\beta 2YearTrend+\epsilon$. I used python to analyze if there was still a correlation between the change in inflation compared to the change in holiday sales and it turns out there is. My new p-value is 0.007 which is statistically significant. The final controlled regression equation is $\Delta Sales=-46.73+5.75(\Delta Inflation)+2.17(Year\ Trend)+\epsilon$ This means that if inflation increases one percentage point more than the year prior, holiday sales increase by roughly 5.7 billion dollars, after controlling for non-inflation related year-to-year growth. After removing time trends and autocorrelation using double first differences, I found that the correlation between changes in inflation and changes in holiday sales is statistically significant.

References

Where to find the CPI for November of each year: https://fred.stlouisfed.org/series/CPIAUCSL?

Where to find holiday sales (November 1st - December 31st) starting from 2003:

https://capitaloneshopping.com/research/black-friday-statistics/

I used Python to code and collect valuable insights on my data.

I changed the column names of my raw data towards the end of the project, which affected my code. To quickly fix all my code, I used AI.

I used Grammarly to review my grammar for this written report.