Assignment 4A-1: Guac

Guacamole Sales and Pricing Analysis Report

Introduction

This report aims to provide a detailed analysis of the relationship between guacamole sales and pricing at Chipotle Mexican Grill. The necessity to adjust menu prices due to economic conditions and fluctuations in raw material costs drives this inquiry. The objective is to understand the sensitivity of guacamole sales to pricing changes, aiding in strategic pricing decisions.

Data Overview

The dataset spans from 2020 to the third quarter of 2023, across ten U.S. regions, comprising quarterly records of guacamole orders (in thousands) and their corresponding prices. Preliminary data analysis reveals no missing values, ensuring a robust dataset for analysis.

Methodology

1. OLS Estimation

Objective: To establish a baseline relationship between Sales and Price.
Procedure: Employed the lm function in R for linear regression, modeling Sales as a function of Price, Region, Year and Quarter.
Model: Sales ~ Price + Region + Year + Quarter, with the inclusion of dummy variables for regions, years, and quarters to control for potential confounding factors.
Concern: Potential endogeneity in the price variable, which could lead to biased estimates. To figure this out more reliably, we should use a better way of analyzing it. The Two-Stage Least Squares (2-SLS) Approach is a method that can help do this.

2. Two-Stage Least Squares (2-SLS) Approach

Rationale: To address the possible endogeneity of the price variable in the OLS model.
Methods Applied:

- Manual 2-SLS estimation: Conducted in two stages where the first stage predicts price using an instrumental variable, and the second stage models sales as a function of the predicted Price, Region, Year and Quarter.
- Automated 2-SLS estimation: Employed the “ivreg” package in R, which streamlines the 2-SLS process.

Instrumental Variable: The Global Avocado Harvest Size Index was employed to predict price. This selection is based on the hypothesis that fluctuations in avocado harvest sizes impact guacamole prices due to changes in supply, while having no direct causal relationship with the

quantity of guacamole sold. Thi satisfies the requirements of relevance and exogeneity for an effective instrumental variable.

Estimates of Price's Average Effect on Sales

1. Ordinary Least Squares (OLS) Estimation

- **Coefficient of Price:** The OLS regression model estimated a coefficient of -9.49207 for the price variable.
- **Interpretation:** This implies that for every one-unit increase in price, guacamole sales decrease by approximately 9,492 units. The relationship between price and sales is strongly negative and statistically significant.
- **Model Reliability:** The high R-squared value of 0.9781 indicates that the model explains a significant portion of the variance in sales. However, concerns about potential endogeneity of the price variable may lead to biased OLS estimates.

2. Two-Stage Least Squares (2-SLS) Manual Approach

- **Coefficient of Price:** The manual 2-SLS approach estimated a price coefficient of -8.2999.
- **Interpretation:** This coefficient suggests a slightly lower negative impact of price on sales compared to the OLS estimate. Specifically, a one-unit increase in price is associated with a decrease of about 8,300 units in sales.
- **Considerations:** The 2-SLS method addresses potential endogeneity issues, likely providing a more accurate estimate of the true causal effect of price on sales.

3. Two-Stage Least Squares (2-SLS) Using 'ivreg' Package

- **Coefficient of Price:** The automated 2-SLS estimation via the “ivreg” package estimated a coefficient of -6.9419 for price.
- **Interpretation:** This result suggests that the negative impact of price on sales is substantial but less pronounced than what the OLS and manual 2-SLS models suggest. A one-unit increase in price is associated with a decrease of about 6,942 units in sales.
- **Model Strength**: The statistical tests from the “ivreg” output, including weak instruments and Wu-Hausman tests, affirm the validity of the instrumental variable used and the robustness of the model’s estimates.

Summary: The three approaches yield different estimates for the impact of price on guacamole sales, with the OLS method indicating the most significant negative effect. However, after adjusting for potential endogeneity with the 2-SLS methods, the estimated negative impact of

price on sales is found to be slightly less severe.. The "ivreg" method, in particular, provides a more robust estimate with its comprehensive control for endogeneity and other factors.

Comparative Analysis of Methodological Results

OLS Estimation:

- Coefficient of Price: -9.49207
- Intercept: 160.14726
- Standard Error of Price: 0.24288
- R-squared: 0.9781
- Residual Standard Error: 4.549
- F-statistic: 371.5 on 16 and 133 DF, p-value: < 2.2e-16

2-SLS Manual Approach:

- Coefficient of Price (D_hat): -8.2999
- Intercept: 176.0445
- Standard Error of D_hat: 1.38986
- R-squared: 0.7747 (Note: This R-squared is for the second-stage regression)
- Residual Standard Error: 14.59
- F-statistic: 28.58 on 16 and 133 DF, p-value: < 2.2e-16

2-SLS Using 'ivreg' Package:

- Coefficient of Price: -6.9419
- Intercept: 137.8604
- Standard Error of Price: 0.5501
- R-squared: 0.96
- Residual Standard Error: 6.151
- Weak Instruments Diagnostic Test: p-value = 2.13e-14
- Wu-Hausman Test: p-value < 2e-16

Key Differences:

Coefficient Estimates:

The OLS coefficient (-9.49207) is more negative compared to the 2-SLS coefficients (-8.2999 and -6.9419), indicating a higher sensitivity of sales to price changes in the OLS model.

Standard Errors:

The OLS model exhibits a lower standard error (0.24288) compared to the 2-SLS models (1.38986 for manual and 0.5501 for 'ivreg'), reflecting more precision. This variation is due to the additional uncertainty introduced by adjusting for endogeneity in 2-SLS methods.

R-squared Values:
The OLS model has a higher R-squared (0.9781), suggesting a better fit but potentially misleading if endogeneity is present. The 'ivreg' 2-SLS model also shows a high R-squared (0.96), while the manual 2-SLS has a lower R-squared (0.7747), indicating a possibly more robust approach to endogeneity.

Residual Standard Error:
The OLS model's lower residual standard error (4.549) suggests tighter clustering of residuals around the regression line compared to the 2-SLS models (14.59 for manual and 6.151 for 'ivreg').

Diagnostic Tests ('ivreg' Method):
The 'ivreg' method includes important diagnostic tests like the Weak Instruments and Wu-Hausman tests, providing additional validity checks for the instrument's strength and the model's appropriateness, which are not present in the OLS or manual 2-SLS methods.

Summary:
While the OLS model indicates a higher sensitivity of sales to price changes, its potential bias due to endogeneity makes the 2-SLS methods, particularly the 'ivreg' approach, more reliable despite higher standard errors and varying R-squared values.

Uncertainty Associated with Estimated Price Elasticities

1. Ordinary Least Squares (OLS) Approach:

- Standard Error: The OLS model estimated the price coefficient with a standard error of 0.24288. This relatively low standard error indicates a high level of precision in the OLS estimate of the price coefficient.

- Confidence Interval: The 95% confidence interval for the price coefficient in the OLS model ranges from -9.9724692 to -9.0116740. This narrow interval suggests a high level of certainty about the estimate within this range.

2. Two-Stage Least Squares (2-SLS) Manual Approach:

- Standard Error: The manual 2-SLS approach resulted in a larger standard error of 1.38986 for the price coefficient. This increase reflects greater uncertainty in the estimate, likely due to the additional complexity of addressing endogeneity.

- Confidence Interval: The 95% confidence interval for the price coefficient in the manual 2-SLS method ranges from -11.02402 to -5.575873. This wider interval indicates more uncertainty in the estimated impact of price on sales than the OLS model.

Two-Stage Least Squares (2-SLS) Using 'ivreg' Package:

- Standard Error: The 'ivreg' 2-SLS method shows a standard error of 0.5501 for the price coefficient, which is lower than the manual 2-SLS but higher than the OLS approach. This indicates a moderate level of uncertainty.

- Confidence Interval: The 95% confidence interval for the price coefficient using this method ranges from -8.0300357 to -5.8537027. This interval is narrower than the manual 2-SLS but wider than the OLS, suggesting moderate certainty in the price coefficient estimate.

Summary: The OLS approach shows the least uncertainty in its estimates, as indicated by the lowest standard error and the narrowest confidence interval. However, due to potential endogeneity issues, this precision might be misleading. The 2-SLS approaches, both manual and 'ivreg', exhibit higher uncertainty in their estimates, as seen in larger standard errors and wider confidence intervals. Despite this, they offer a more accurate reflection of the true causal relationship by addressing endogeneity concerns, especially the 'ivreg' package approach.

Implication of using the erroneous OLS price elasticity model coefficients

OLS model:
The coefficient for price is -9.49207. For every $1 increase in the price of guacamole, there would be a decrease in sales by 9,492 units.

Manual 2SLS model:
The coefficient for price is -8.2999. For every $1 increase in the price of guacamole, there would be a decrease in sales by approximately 8,300 units.

Ivreg 2SLS model:
The coefficient for price is -6.9419. For every $1 increase in the price of guacamole, there would be a decrease in sales by approximately 6,942 units.

Based on the OLS regression results, there would be a decrease in sales by 9,492 units for every $1 increase in the price of guacamole. However, due to endogeneity issues, the estimates provided by the OLS model are biased and inconsistent. This could lead to misguided business decisions if the management were to use the erroneous OLS model. For instance, if the true

effect of price on sales is smaller in magnitude than the OLS estimate (as per the automated 2-SLS model coefficient of -6.9419), then the OLS model is overestimating the negative impact of a price increase on sales. Consequently, management might avoid increasing prices due to an overestimated fear of losing sales, potentially missing out on additional revenue.

Therefore, the implication of using the erroneous OLS coefficients is that management might either miss out on potential profits by not increasing prices enough (the OLS estimate is larger in absolute value than the true effect). It underscores the importance of addressing endogeneity and using a more reliable estimation method, such as 2-SLS ('ivreg' approach), which provides an adjusted coefficient (-6.9419) indicating a somewhat smaller impact of price on sales than the OLS model suggests. This could lead to more informed and effective pricing strategies.

Assignment 4A-2: Sweeteners

Analytical Report: Sales Impact of Shelf Repositioning for Splenda

Objective

The report aims to analyze the impact of an unplanned change in the shelf positioning of a Splenda product from lower to eye-level shelving in a retail store located in Boston.

Data and Methodology

The data comprised two years of weekly sales from 167 U.S. stores. Initially, all stores displayed Splenda on the bottom shelf. In the following year, store number 205512 moved Splenda to an eye-level shelf. The methodology employed was the Difference-in-Differences (DiD) regression model. This approach was chosen due to its effectiveness in evaluating causal relationships in observational studies. By comparing the before-and-after effects of the intervention (shelf repositioning) in the treatment group (store 205512) and comparing it with a control group (other stores), DiD helps isolate the effect of the intervention from other confounding variables.

The model incorporated variables like time, promotional activities, market, closure status, price per unit, etc. These controls were crucial for adjusting for factors that could influence sales independently of the shelf change, such as general market trends, seasonal effects, or store-specific characteristics.

Findings

The analysis of the data indicates that moving the Splenda product to eye-level shelving resulted in an average increase of approximately 10.59 units sold per week for the store that implemented this change. This figure represents the change in sales attributable specifically to the new shelf placement after accounting for other variables in the model.

Potential Impact of Adopting the Eye-Level Shelf Strategy in all U.S

For Future Years:

Annual Impact per Store: Projecting the weekly increase of 10.59 units over a year yields 550.68 units/year/store.

Total Annual Impact: Assuming uniformity of this effect across all stores, with 165 operational stores, the projected increase is 90,828.47 units/year.

Conclusion

Adopting eye-level shelving across all stores could potentially increase annual unit sales by approximately 90,828.47 units. This projection is based on the assumption that the observed effect in the Boston store would be replicated in each store nationwide.

Implications of Price and Promotions

Price per Unit (PricePerUnit): The coefficient for PricePerUnit is -4.5033471, which is highly significant ($p < 2.2e-16$). This suggests that for every unit increase in the price of Splenda, the units sold decrease by approximately 4.5 units. This significant negative coefficient aligns with basic economic principles, indicating that higher prices tend to reduce demand.

Promotional Features (F): The coefficient of 5.1799546 ($p < 2.2e-16$) for the promotional feature variable indicates that when Splenda is featured in store advertisements, there's an average increase of about 5.18 units sold. This positive and significant coefficient suggests that advertising features are effective in increasing sales.

Display Prominence (D1): The coefficient for display prominence is 14.464327 ($p < 2.2e-16$). This indicates a substantial positive impact on sales when Splenda has additional in-store display prominence, with an increase of approximately 14.46 units sold. It underscores the importance of visual merchandising in driving product sales.

Price Promotions (PR1): The coefficient of 2.214424 ($p < 2.2e-16$) for price promotions implies that when there are price promotions, the units sold increase by around 2.21 units. This positive coefficient suggests that customers respond favorably to price promotions.

Conclusion

Implementing eye-level shelving across all stores could potentially enhance annual unit sales significantly. The projection, based on the Boston store's experience, indicates a substantial potential increase in sales. The analysis also highlights the significant impact of pricing and promotional strategies on sales, offering valuable insights for strategic decision-making in retail merchandising.