Analysis of The United States Domestic Cutlery Market

Dominick DeCanio

Motivation

From October 2018 to February 2020 I launched and ran a cutlery manufacturing venture. The premise of the enterprise was to use the techniques of traditional hand forging to drive innovation in the cutlery industry. Implicit in any business is the need to sell, and implicit in the need to sell is the need for marketing. Thus knowledge of pre-established sales trends in the industry would help me design advertising campaigns to maximize their timing, thereby increasing advertising effectiveness and avoiding ineffective campaign strategies.

Dataset Description

Background Information

The Consumer Price Index (CPI) is a measure of the average change over time in the prices paid by urban consumers for a market basket of consumer goods and services[1]. Consumers spend different proportions of their income on different types of goods. This distribution of expenditure changes over time, which alters the way prices impact the market basket of goods and services consumers are expected to buy. In response, the CPI is calculated by collecting pricing information on goods and services, as well as information on the expenditures of consumers. The data collected on consumer expenditures is used to weight the prices of goods, thereby accounting for this change in expenditure distribution over time. This model needs to be updated and adjusted continually to follow the spending allocations of consumers. Consequently, the Bureau of Labor Statistics (BLS) continually collects data on the expenditures of consumers.

Sample Collection

There are limitations on this investigation. Primarily, the data provided by the Bureau of Labor Statistics is not granular enough to specifically mention "cutlery" as a product category. As a result, I will proceed with my analysis utilizing the product category "Tableware, nonelectric kitchenware," as it is the closest category to cutlery.

The sample data from which I created my data table was sourced from the BLS. The BLS conducts this survey by contacting 5000 addresses each calendar quarter. This outreach results in approximately 3,000 unusable responses per quarter. Survey respondents are consumer units (households) and report data for four consecutive quarters. Each quarter, about 25% of survey respondents are lost and a similar number of respondents are added. This means that there are different numbers of respondents, and different combinations of them, in each quarter.

The data I have analyzed is sourced from the public use micro data (PUMD) database of consumer expenditures. I have compiled a data set which combines this expenditure data within the expenditure category "Tableware, nonelectric kitchenware," with the "Wages and salaries" information for each respondent. This results in a data set which can be used to see the effects of income on expenditures in this segment of goods. Because the BLS compiles this data on a quarterly basis, we will also be able to see the seasonal trends in expenditure for each of these income categories. Because I have held the year constant, and the BLS releases CPI yearly, we can assume that the basket of goods purchased within the year did not change. As a result it is not necessary to standardize or correct the expenditure data.

Explanation of the Rows and Variables

I constructed the data set using the income brackets of the respondents as a bin. This bin forms the rows of the table; "Expenditure_data." The columns of the data set use the categorical variable "Quarter" in order to separate the respondents based off of the quarter in which they were surveyed. I then found the average expenditure for each income level and quarter combination

In constructing my income bracket bins, I used the income brackets that determine federal income taxes as a guide. The income brackets included in the analysis are: 0 - $10,000, $10,001 - $40,000, $40,001 - $85,000, $85,001 - $160,000, and $160,000 and greater. My columns are the quarters in which the expenditure occurred. Each observation in the data table represents the average expenditures of a particular income bracket on non-electric

kitchenware during a particular quarter of 2019. Each observation is measured in U.S. dollars.

Potential Issues

Sample size poses issues for this investigation. The files I began with contained approximately 3,000 responses, but the variable "Tableware, nonelectric kitchenware," was not universally recorded by the respondents. As a result, my cleaned data frames consisted of samples of less than 100 respondents for each quarter. This is small, and the samples became even smaller when I grouped these quarterly results by income bracket. This is a problem because it limits the extent to which we can generalize the conclusions of this analysis. We cannot assume that this data set captures all of the features of consumer expenditure on "Tableware, nonelectric kitchenware," categorized by income, because we have a very small sample in comparison with the population, which is all U.S. domestic consumers.

Further study should be done regarding a finite-sample statistical analysis of the distributions of consumer expenditures. Some data sets which comprise the analysis shown in the following table "Expenditure_data," rely on samples too small to apply the central limit theorem and thus conduct these statistical investigations asymptotically. Alternatively, care could be taken in adjusting other years' data from this survey to standardize the dollar values of expenditures. This would increase the validity of the analysis as descriptively assessing the characteristics of quarterly expenditure. However, in doing so the analyst limits their ability to quantify how this expenditure behavior changes over time.

Numerical Representation

Numerical Representation Data Manipulation

```
colnames(Expenditure_data) = c("Quarter_1","Quarter_2","Quarter_3","Quarter_4")
rownames(Expenditure_data) = c("0 - $10,000", "$10,001 - $40,000",
                  "$40,001 - $85,000", "$85,001 - $160,000", "$160,000 and up")
Expenditure_data
```

```
##                     Quarter_1 Quarter_2 Quarter_3 Quarter_4
## 0 - $10,000           1.00000  2.842500   7.02700   6.06000
## $10,001 - $40,000    14.53985 18.367721  13.86987  22.70181
## $40,001 - $85,000    10.30827  7.839218  11.04119  13.13453
```

```
## $85,001 - $160,000  13.31222 17.094311  13.15600  18.72224
## $160,000 and up     20.18534 26.037493  12.39968  49.90285

qrter_avgs = group_by(Expenditure_data, Q1_all=mean(Quarter_1),
                          Q2_all=mean(Quarter_2), Q3_all=mean(Quarter_3),
                          Q4_all=mean(Quarter_4))
qrter_avgs = qrter_avgs[1, c("Q1_all", "Q2_all", "Q3_all", "Q4_all")]
qrter_avgs = matrix(qrter_avgs, nrow=4, ncol=1, byrow=FALSE)
qrter_avgs = data.frame(qrter_avgs)

qrter_avgs = as.matrix(qrter_avgs)
qrter_avgs = as.numeric(qrter_avgs)
qrter_avgs = round(qrter_avgs, digits=2)
heights = qrter_avgs

rownames =  c("Quarter 1", "Quarter 2", "Quarter 3", "Quarter 4")
qrter_avgs = as.data.frame(qrter_avgs)
qrter_avgs = cbind(rownames, qrter_avgs)
colnames(qrter_avgs) = c("Quarter", "Average")
```

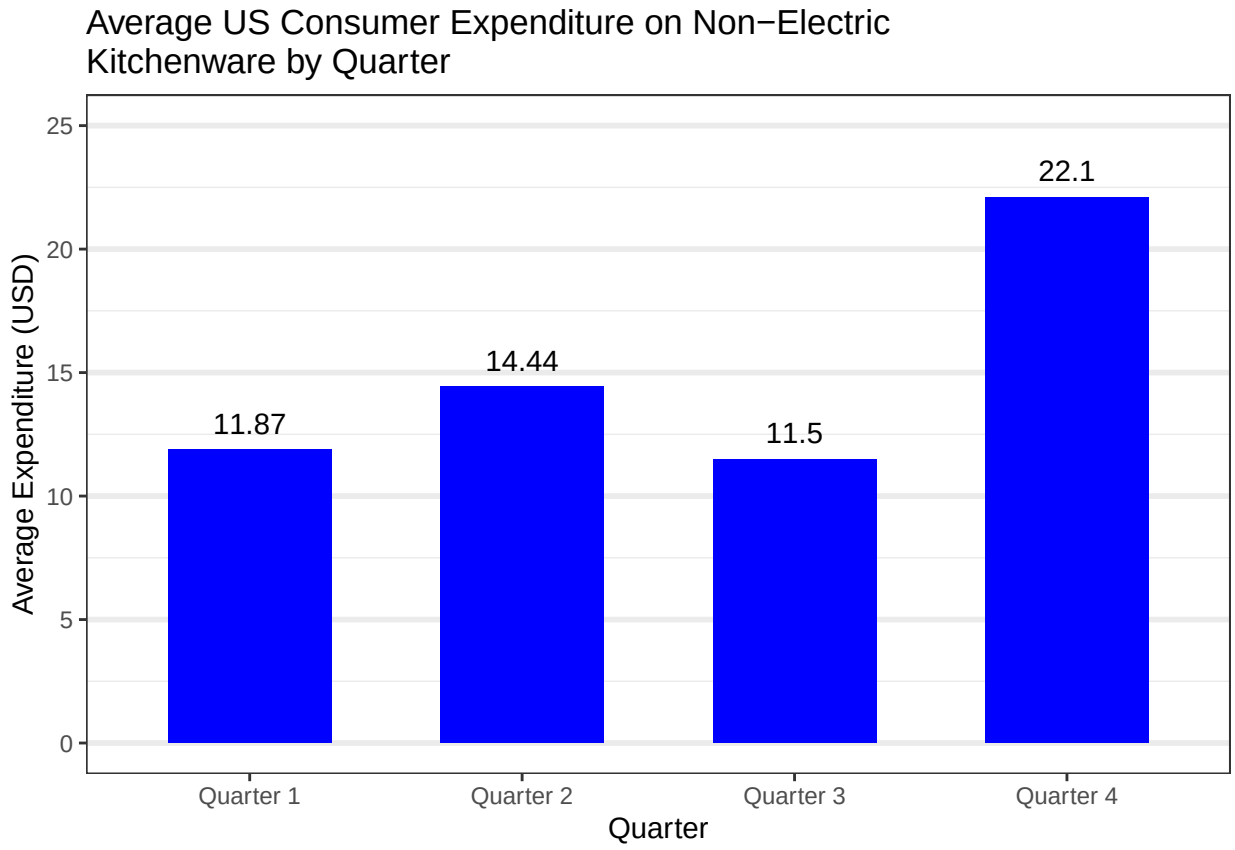
In the above section of code, I began with the data frame "Expenditure_data" and manipulated it into a form that can be easily displayed by my intended numerical summary. To do so, I took the average of every income bracket's average expenditure for a given quarter, and assigned this new value to the "Average" column for that quarter. Doing this for each quarter gives us a sense of the overall spending trends exhibited across all of the income brackets represented. After the data manipulation step is over, we can see that the resulting data frame "qrter_avgs" shows us the data which I will express in my numerical summary.

Numerical Representation

```
theme_set(theme_bw())
numerical_summary = ggplot(qrter_avgs, aes(x=Quarter, y=Average)) +
  geom_bar(stat="identity", fill="blue", width=0.6) +
  labs(title="Average US Consumer Expenditure on Non-Electric
Kitchenware by Quarter", y="Average Expenditure (USD)") +
  stat_summary(geom="text", aes(label=Average), vjust=-.75) + ylim(0,25) +
```

```
  theme(panel.grid.major.x=element_blank(),
        panel.grid.major.y=element_line(size=1))
numerical_summary
```

```
## No summary function supplied, defaulting to ‘mean_se()‘
```



Numerical Representation Conclusions

The above graphic displays the average expenditure on "Tableware, nonelectric kitchenware" by quarter, irrespective of the income level of the respondent.

From this graphic we can see that there is a spike in expenditures in quarter 4. This is certainly due to the holiday season. The holiday season can be attributed to causing the quarter 4 spike in expenditures on "Tableware, nonelectric kitchenware," because the holiday season increases demand for this product category in two ways. First, people may purchase these items as gifts. Second, there is an overall greater need for this type of good during this time of year. Quarter 4 encompasses nearly the entire holiday season, and with events

like Thanksgiving, Christmas, and Hanukkah, which traditionally result in large gatherings of people, tableware is a more necessary and functional commodity.

Graphical Representation Data Manipulation

In order to properly express this data set graphically, more data manipulation is needed. When creating my source data frame, I made the table "Expenditure_data" in a way that is easily read by people. Because the data frame "Expenditure_data" is in wide form, I convert it into long form in the following segment of code in order to create a graphical representation thereof.

```
income_1 = matrix(c(Expenditure_data[1,]), nrow=4, ncol=1, byrow=TRUE)
income_2 = matrix(c(Expenditure_data[2,]), nrow=4, ncol=1, byrow=TRUE)
income_3 = matrix(c(Expenditure_data[3,]), nrow=4, ncol=1, byrow=TRUE)
income_4 = matrix(c(Expenditure_data[4,]), nrow=4, ncol=1, byrow=TRUE)
income_5 = matrix(c(Expenditure_data[5,]), nrow=4, ncol=1, byrow=TRUE)

avg_expend_data_long = matrix(rbind(income_1, income_2, income_3, income_4,
                                    income_5), nrow=20, ncol=1, byrow=TRUE)
avg_expend_data_long = as.numeric(avg_expend_data_long)

income_levels = rep(c("0 - $10,000", "$10,001 - $40,000", "$40,001 - $85,000",
                      "$85,001 - $160,000", "$160,000 and up"), each=4)
time = rep(c("Quarter 1", "Quarter 2", "Quarter 3", "Quarter 4"), 5)
```

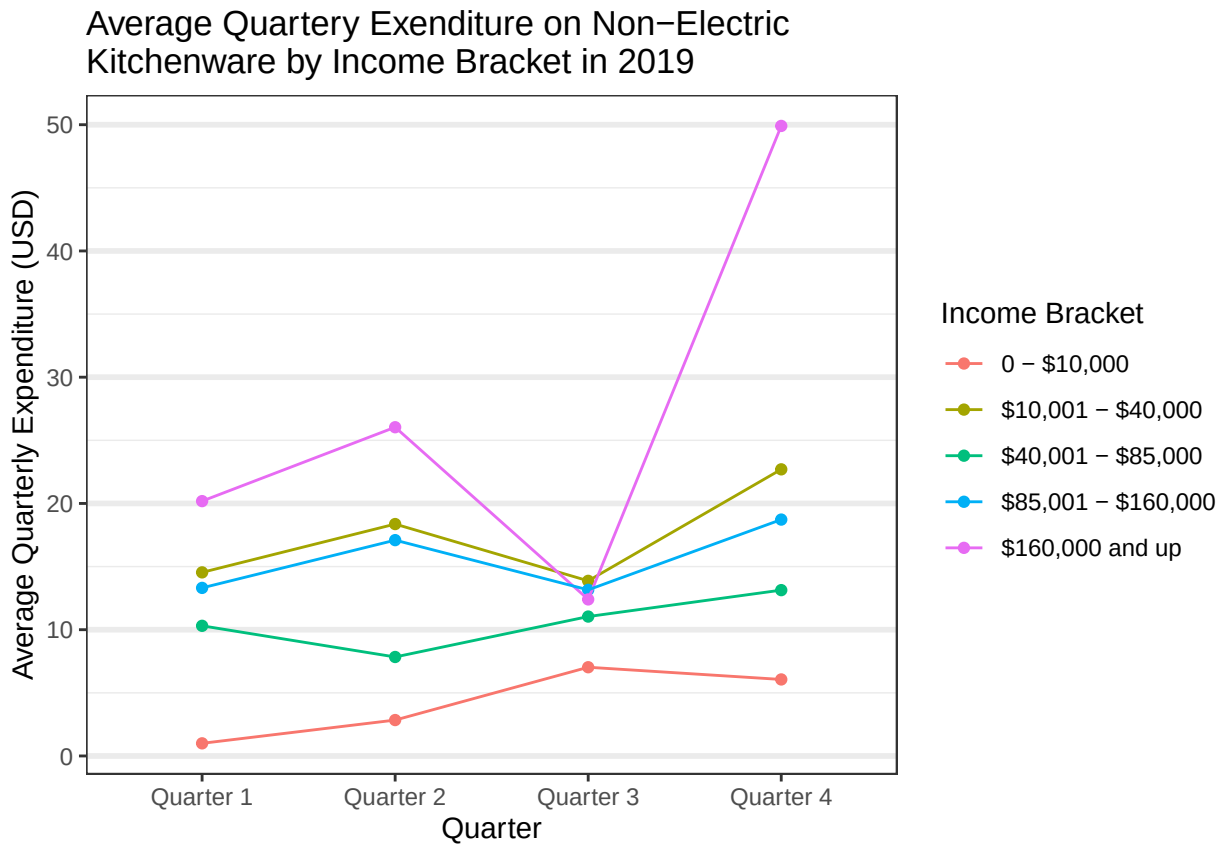
The result is the data frame "expenditure_data_long," a segment of which is visible below.

```
expenditure_data_long = data.frame(income_levels, avg_expend_data_long, time)
head(expenditure_data_long)
```

```
##       income_levels avg_expend_data_long      time
## 1       0 - $10,000              1.00000 Quarter 1
## 2       0 - $10,000              2.84250 Quarter 2
## 3       0 - $10,000              7.02700 Quarter 3
## 4       0 - $10,000              6.06000 Quarter 4
## 5 $10,001 - $40,000             14.53985 Quarter 1
## 6 $10,001 - $40,000             18.36772 Quarter 2
```

Graphical Representation

```
graphical_summary = ggplot(expenditure_data_long, aes(x=time,
                           y=avg_expend_data_long, color=income_levels,
                           group=income_levels)) +
  geom_point() + geom_line() +
  labs(title="Average Quartery Exenditure on Non-Electric
Kitchenware by Income Bracket in 2019", y="Average Quarterly Expenditure (USD)",
       x="Quarter", color="Income Bracket") +
  scale_color_discrete(limits=c("0 - $10,000", "$10,001 - $40,000",
       "$40,001 - $85,000", "$85,001 - $160,000", "$160,000 and up")) +
  theme(panel.grid.major.x=element_blank(),
        panel.grid.major.y=element_line(size=1))
graphical_summary
```



Graphical Representation Conclusions

Both income and time of year affect the expenditures of consumers. Generally, those with a higher income spent more money. The peaks in expenditure in Quarter 4 are not surprising, but we can also see that there are smaller peaks in Quarter 2 for some income brackets. This is a result I did not anticipate. These Quarter 2 peaks could be caused by summer cleaning, and people replacing worn out kitchenware as they clean.

On first glance it would appear that the highest three income brackets exhibit different behavior than the lower two brackets, based on this curious Quarter 2 peak. Upon inspection, however, the viewer will notice that the second lowest income bracket-$10,000 to $40,000-holds the second highest expenditures of any income bracket. Furthermore, the pattern of expenditure is almost precisely identical to that of the the second highest income bracket-$85,001 to $160,000.

I believe this phenomena is caused by the different motivators for demand. The lowest income bracket might reasonably not have a high demand for kitchenware, as there is strong a connection between low income and consumption of pre-prepared foods. The second lowest bracket, however, might have a higher demand for these kitchen items, but are forced to purchase non-durable goods because of cost limitation. As a result, they will spend more monetarily for a comparable amount of kitchen-ware utility, because the goods they must purchase do not last. This logic continues to apply for the other income brackets. For the middle income bracket, there is still high demand for kitchen goods, but because of their ability to spend more, they acquire lasting equipment, thereby lowering overall costs. In the second highest bracket, I presume that their goods are not readily wearing out, but the price premium of the incrementally better wearing kitchenware evens out the consumer surplus of their durability. Finally, for the highest income bracket, their expenditures are dramatically different over time because cost is no longer a real consideration. Spikes in expenditure become severe because purchases are made based on timing rather than necessity. Evidence of this is that the highest income bracket had lower expenditures in Quarter 3 than 2 other income brackets, with its inter-Quarter variation being the most severe.

References

1. Bureau of Labor Statistics, https://www.bls.gov/cex/pumd_data.htm.