Cereal Market Analysis

Daniel Bestard, Akhil Lohia, Michael Cameron

November, 2016

Introduction

This report provides an analysis of the ready-to-eat cereal market in the USA. Each observation in the used dataset corresponds to a specific product (there are a total of 82 products across 7 manufacturers) for each month and store. Therefore, for each row of the dataset we know how many products were sold in a specific month and store. Given the large amount of total products and the great similarities among some of them, our team has decided to perform the analysis by manufacturer. There are a total of seven manufacturers which are: Dominicks, General Mills, Kelloggs, Kraft, Nabisco/Kraft, Quaker Oats and Ralston. Alternate analysis could have been followed, for example, by the store selling the products. However, our team has special interest in understanding the competitive situation among main manufacturers and see how their demand curves look like.

We start this report with some descriptive analysis about the cereal market where variables are grouped by manufacturer as explained before. Once some insights about the cereal market have been gained ,we proceed to study the demand curves of each manufacturer. In the study of demand curves, we use multiplicative demand models and instrumental variables to deal with endogeneity problems. The reason why we rely on instrumental variables in this analysis is because when estimating the demand of a given manufacturer, variables such as price of inputs might affect the price and not quantity demanded. Therefore, it does not make sense to include such a predictor in the demand model. However, cost of inputs (among other variables) can be used to predict the price of a manufacturer and, therefore, remove the correlation between price and the error term. This procedure is called, Two Stage Least Squares, and allows us to estimate demand models with as little bias as possible.

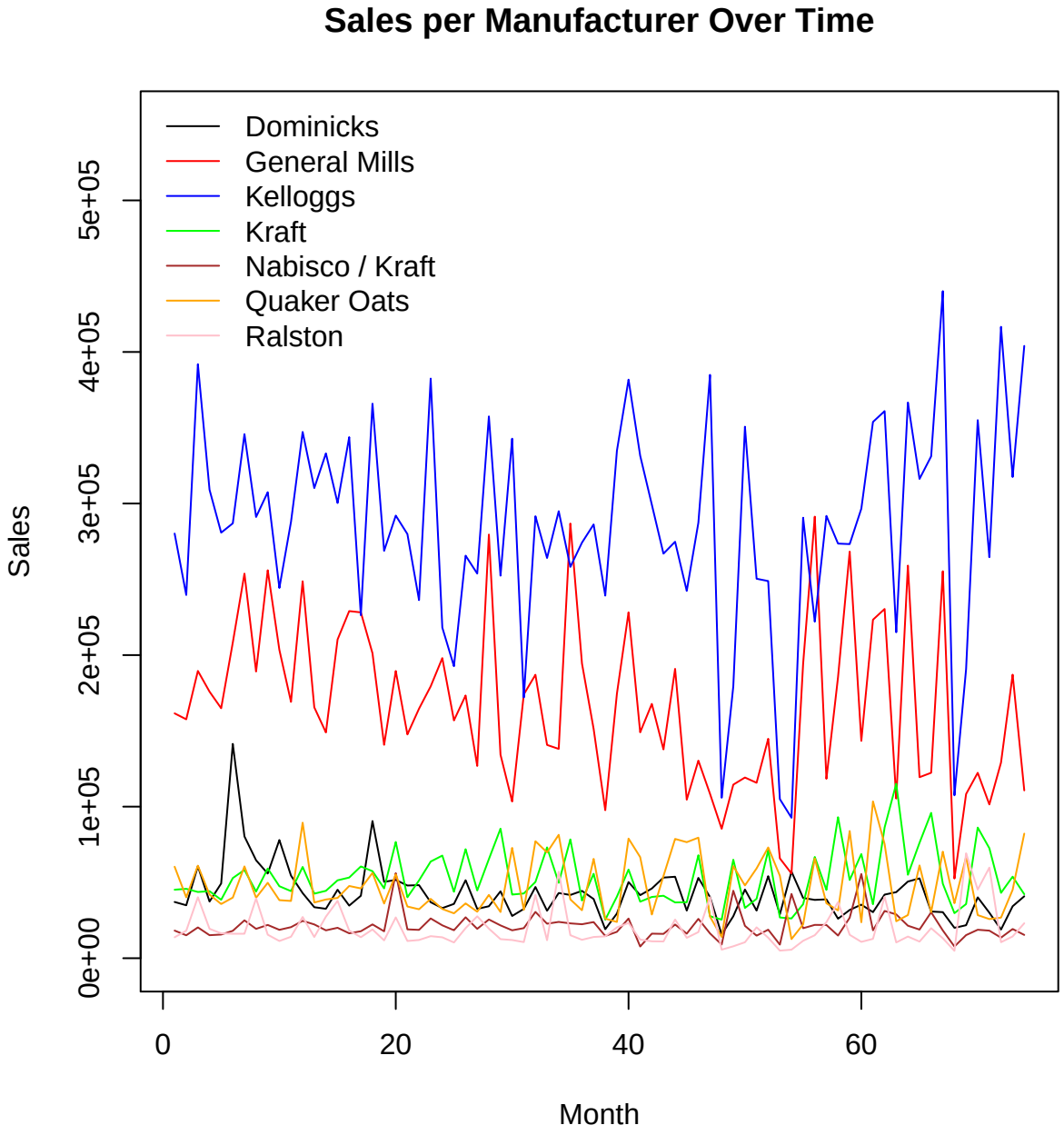
Descriptive statistics

To get an overview of the market, here we display the market share (in percentage) and the mean price (per unit weight over all products) of each of the seven manufacturers.

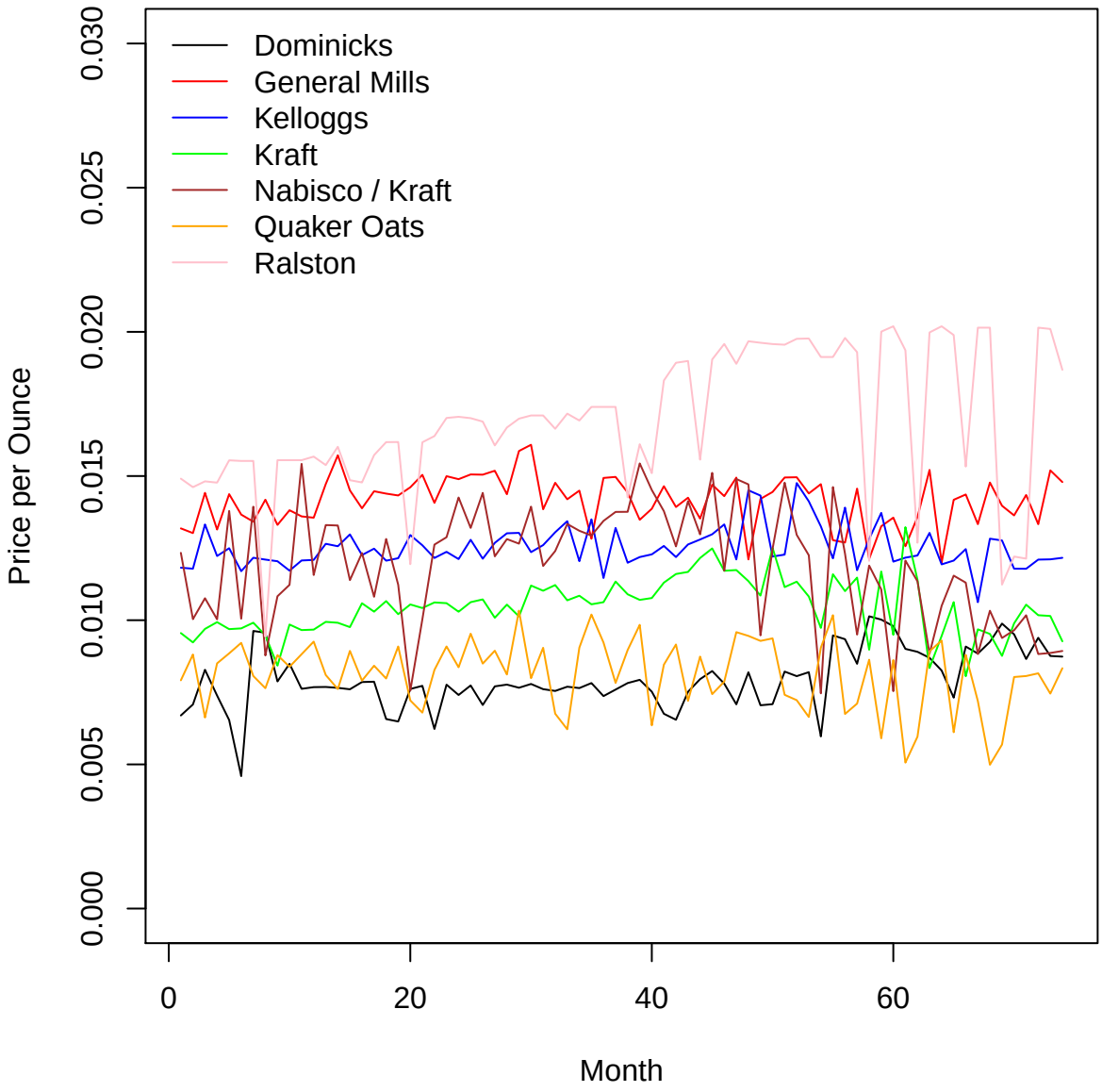
	Dominicks	G.Mills	Kelloggs	Kraft	N/K	Q.Oats	Raltson
Market Share	6.7	26.2	44.8	8.3	3.4	7.5	3.1
Mean Price	12.9	21.6	19.6	17.5	18.3	16.1	21.8

Note that the position of Kelloggs in the cereal market is outstanding given that almost half of the market share belongs to it. General Mills follows Kelloggs with a market share of 26.2% and the rest of manufacturers have market shares smaller than 10%. With respect to the mean price, Ralston and General Mills both have nearly the highest mean price. On the other hand, Dominicks seems to be on average the cheapest manufacturer of the cereal market.

The following plot displays the quantity sold per manufacturer over time.



Note that there seems to be no specific trend over time. However, what can be inferred from this plot is that Kelloggs and General Mills have the highest market share and their demands change accordingly for most part of the period. The evolution of average price per manufacturer over time is displayed in the next plot.



We can infer that Ralston is the most expensive company and General Mills is a close second, which we had already found out before. But in this graph we can see that the variability in price is much larger for Ralston. On the other hand, the strongest manufacturer, Kelloggs, is the one that has less price variability.

Some descriptive data about the nutritional facts for each manufacturer can be found in the following table:

	Dominicks	G.Mills	Kelloggs	Kraft	N/K	Q.Oats	Raltson
Calories	365.8	383.9	372.3	357.0	334.0	415.5	369.7
Total fat (g)	2.0	4.5	1.5	2.2	1.4	9.4	1.8
Saturated fat (g)	0.7	0.5	0.4	0.4	0.2	5.6	0.3
Sodium (mg)	670.2	604.9	577.8	593.0	5.0	399.0	887.3
Carbohydrates (g)	84.3	80.9	85.0	81.9	79.9	78.7	83.9
Fiber (g)	4.4	6.5	3.0	7.8	13.4	4.7	5.3
Sugar (g)	23.7	26.7	29.1	26.3	1.0	34.8	9.1
Protein (g)	6.3	7.8	7.6	9.0	12.5	7.4	7.7
Vitamin A (mg)	1690.5	1608.0	1976.6	1875.3	0.0	77.0	1506.3

Iron (mg)	22.6	23.9	20.2	23.8	4.2	10.9	30.7
Vitamin C (IU)	24.2	41.1	29.3	2.7	0.0	0.1	18.1
Calcium (mg)	34.0	655.8	24.8	36.7	45.0	64.0	301.0

From the previous table it can be concluded that the manufacturer with the worst nutritional composition is Quaker Oats given that much more fat, calories and sugar than the rest of manufacturers. General Mills seems to have an extraordinarily high calcium content in their products as well.

The following table displays the mean content of each manufacturer's products in terms of % of principal ingredients

	Dominicks	G.Mills	Kelloggs	Kraft	N/K	Q.Oats	Raltson
% Corn	33	38	43	17	0	25	33
% Wheat	8	25	31	50	100	0	33
% Rice	25	0	22	0	0	0	33
% Oat	33	50	4	17	0	75	0
% Barley	0	0	0	17	0	0	0

We see that some companies do not use some of the components at all.

The following table shows the average number of coupons per manufacturer.

	Dominicks	G.Mills	Kelloggs	Kraft	N/K	Q.Oats	Raltson
Mean Coupons	0	0.9	0.4	0.8	0	0.5	0.3

Hence, we see that companies such as Dominicks and N/K do not provide discounts in their products. On the other hand, General Mills and Kraft are the companies that provide more discounts.

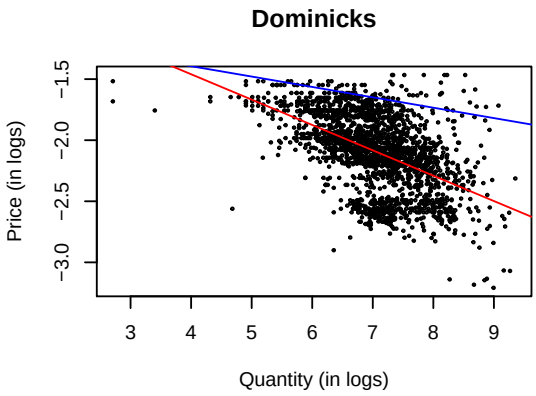
The last descriptive table shows the percentage of adult cereals for each manufacturer:

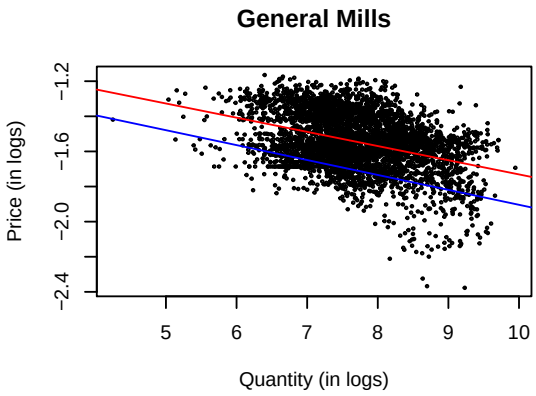
	Dominicks	G.Mills	Kelloggs	Kraft	N/K	Q.Oats	Raltson
% Adult cereal	25	25	22.2	66.7	100	50	0

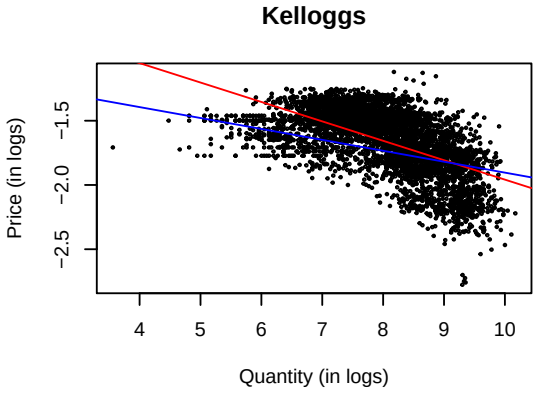
Very interestingly, we see that some manufacturers make their product for specific parts of the population. N/K only produce cereals for adults. On the other hand, Ralston only produces cereals for children. The most important player, Kelloggs, devotes 22.2% of their products to adults.

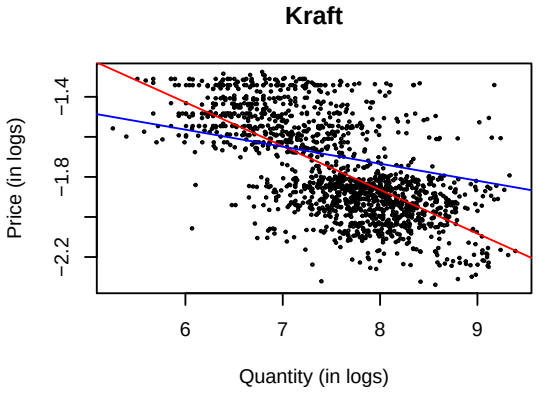
Demand Estimation

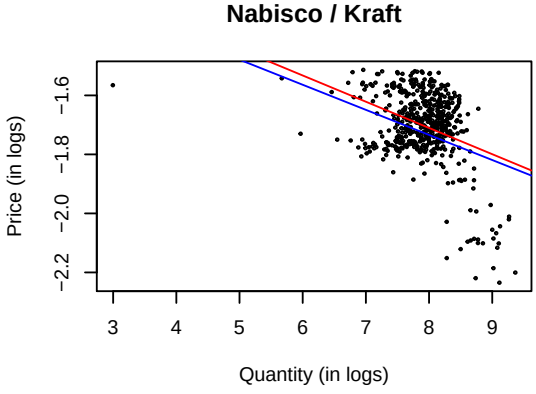
To get an immediate overview of how each manufacturer's demand behaves with respect to the whole cereal market demand, we have built the following plot, where price is plotted against quantity in logarithms for each manufacturer. The red curve is the estimated demand for a specific manufacturer where the only predictor is *its own price*, and the blue curve is the demand of the whole market, using the prices of all manufacturers.

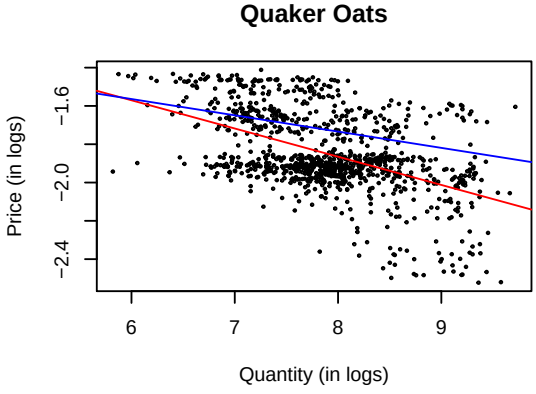

Dominicks
Price (in logs)
−1.5
−2.0
−2.5
−3.0
3
4
5
6
7
8
9
Quantity (in logs)

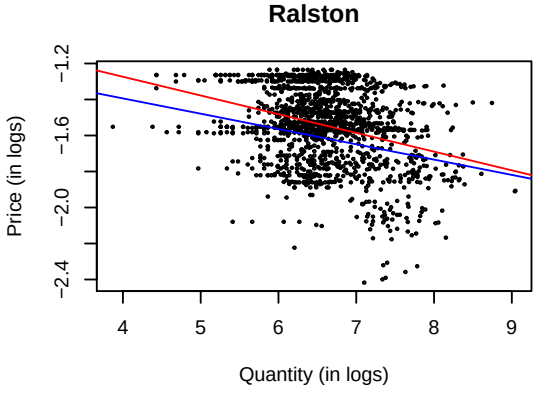

General Mills
Price (in logs)
−1.2
−1.6
−2.0
−2.4
5
6
7
8
9
10
Quantity (in logs)


Kelloggs
Price (in logs)
−1.5
−2.0
−2.5
4
5
6
7
8
9
10
Quantity (in logs)


Kraft
Price (in logs)
−1.4
−1.8
−2.2
6
7
8
9
Quantity (in logs)


Nabisco / Kraft
Price (in logs)
−1.6
−1.8
−2.0
−2.2
3
4
5
6
7
8
9
Quantity (in logs)


Quaker Oats
Price (in logs)
−1.6
−2.0
−2.4
6
7
8
9
Quantity (in logs)


Ralston
Price (in logs)
−1.2
−1.6
−2.0
−2.4
4
5
6
7
8
9
Quantity (in logs)

The demand of the manufacturers Dominicks, Kelloggs, Kraft and Quaker oats are clearly more inelastic than the overall demand of cereals (red line is more steep than blue line). However, as explained before, these demands only use the price of each manufacturer as predictor. Therefore, cross-price elasticities are not computed here which could lead to a change of the slope of the red lines. However, the previous plot provides a quick overview of how the market behaves. With respect to the companies General Mills, N/K and Ralston, it looks like their price elasticities are very similar to the elasticities of the whole market (similar slopes of the red and blue lines).

Instrumental variables

As explained in the introduction, in this report we use instrumental variables for predicting the price of each manufacturer in order to account for endogeneity factors. Given that a good instrument requires a non-zero correlation between the instrument and the price variables, in order to choose proper instruments we started studying these types of correlations. The chosen instruments are:

- Cost of inputs: the inputs of the cereals taken into account are barley, wheat, sugar, corn, rice and oats. For a given observation (product) we computed the sum of the expenditure per 100g of each of these inputs. Hence, we obtain an overall variable that contains most of the cost of inputs.
- Cost of gasoline when transporting: again, the cost of transportation should affect the price but not the quantity (at least directly). We use the cost of gasoline from the factory to the retailer.
- Cost of salaries of transportation: this refers to the fact that the higher the salaries the higher the cost of transportation and, therefore, higher prices.
- Margin per sold product: this has a very strong positive correlation with price, which makes it a good instrument.
- Average price in all zones: this is the most important instrument given that the correlation with price is extremely high and positive. This variable accounts for the fact that prices among areas are not equal. Here we make the assumption that if an area is generally more expensive than another one, then individuals do not move to a cheaper area to buy their products. This assumption is necessary given that if it does not hold then it would affect quantity demanded as well as price.
- Average price in all zones: this is the most important instrument given that the correlation with price is extremely high and positive. This variable account for the fact that prices among areas are not equal. Here we make the assumption that consumers do not travel to cheaper neighbourhoods just to purchase cereals. This assumption is necessary given that if it does not hold then it would affect quantity demanded as well as price.

The output of the regression is:

```
Call:
lm(formula = price_mon ~ distance_gasoline + cost_inputs + earnings_tradetransport +
    retailprofperquant_mon + price_instru_zone, data = cereals)

Residuals:
      Min        1Q    Median        3Q       Max
-0.094308 -0.004153 -0.000729  0.003789  0.053008

Coefficients:
                           Estimate Std. Error t value Pr(>|t|)
(Intercept)              8.900e-03  1.252e-03   7.107 1.23e-12 ***
distance_gasoline        7.409e-06  4.801e-07  15.432  < 2e-16 ***
cost_inputs             -7.693e-08  1.170e-08  -6.576 4.98e-11 ***
earnings_tradetransport -8.728e-04  1.175e-04  -7.426 1.18e-13 ***
```

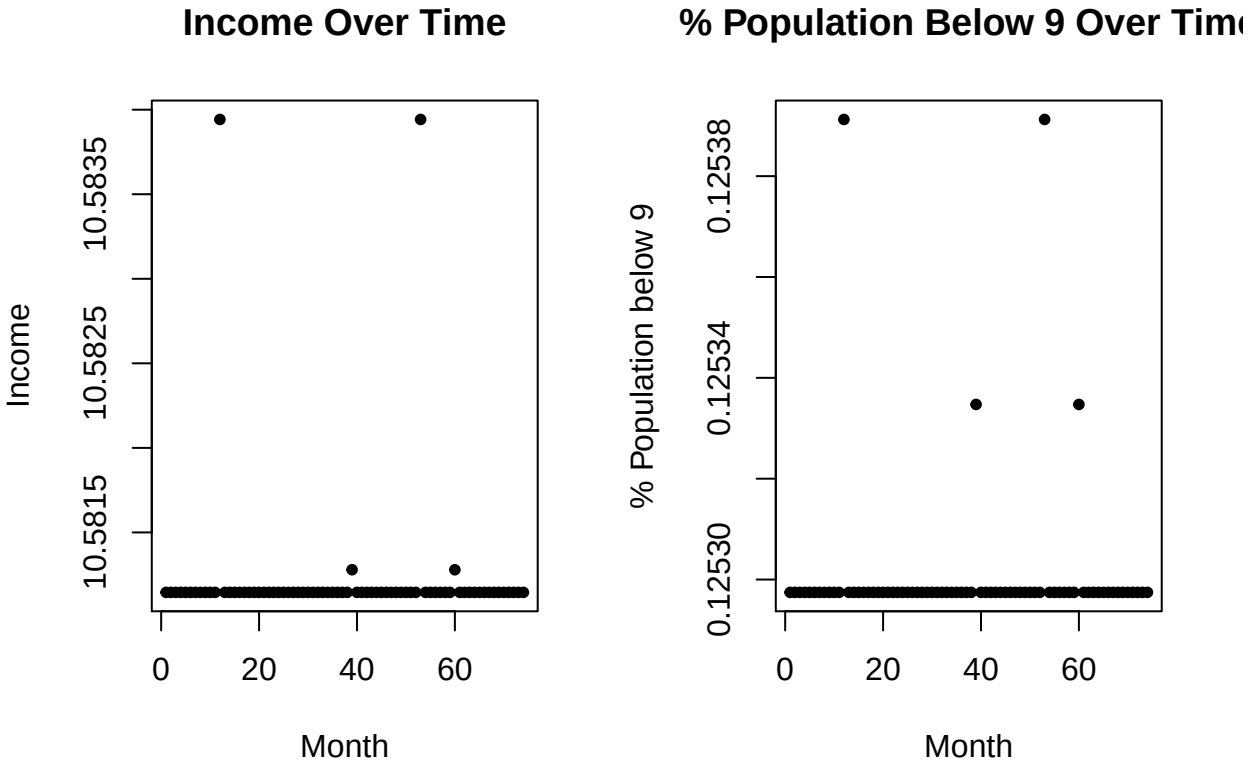
```
retailprofperquant_mon   3.886e-01  5.790e-03  67.118  < 2e-16 ***
price_instru_zone        9.456e-01  1.754e-03 539.086  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.007836 on 15530 degrees of freedom
Multiple R-squared:  0.9746,    Adjusted R-squared:  0.9745
F-statistic: 1.189e+05 on 5 and 15530 DF,  p-value: < 2.2e-16
```

Note that all variables are statistically significant at a 0.05 levels of significance. However, the effect of gasoline cost when transporting, the effect of input costs and the effect of salaries of transportation is very close to zero. The only instruments that seem to have a significant effect on price are average price in all zones and margin per sold product.

So far, we have performed the first step of the Two Step Least Squares Procedure. The next step is to predict the price variable using the model above and use this predicted variable to fit the demand curves. The following subsection contains the demand results for each company:

Demands for Each Manufacturer

When estimating the demands for each manufacturer we used the estimated price using the instrumental variables. Once the price variable is predicted we computed the mean price per manufacturer and month and use these data to estimate the demands. The variable quantity was also grouped by manufacturer and month and used as the dependent variable. It is very important to note that we do not use other exogenous variables in our model because the most significant ones do not change over time. For example, it makes sense to think that income and percentage of population below 9 years old will affect the quantity demanded but the following plot displays these changes over time.



Note that almost no change is happening over time, which means that adding this variable as predictor would mean to basically add a constant. The same phenomenon occurs when using other potential exogenous

variables.

The following subsections show the estimated demand of each manufacturer where each coefficient can be interpreted as cross-price elasticities. Note that all the effects of changes in prices of all manufacturers on the quantity demanded of a given one are not statistically significant at a 0.05 level of significance. This basically means that we did not find statistical evidence to say that when a manufacturer changes its prices, it also affects the quantity demanded of other manufacturers. Of course, there is statistical evidence that at a 0.05 level of significance the changes in prices of a given manufacturer affects its own demanded quantity, and this relationship is negative in all cases.

The following subsections show the estimated demand of each manufacturer where each coefficient can be interpreted as cross-price elasticities. Note that all the effects of changes in prices of all manufacturer on the quantity demanded of a given one are non-statistically significant at a 0.05 level of significance. This basically means that we did not find statistical evidence that when a manufacturer changes its prices there is no effect on the quantity demanded of other manufacturers. Of course, there is statistical evidence that at a 0.05 level of significance the changes in prices of a given manufacturer affects its demanded quantity, and these relationship is negative in all cases.

Dominicks

	Estimate	Pr(>|t|)
(Intercept)	-1.10	0.84
log(Dominicks_price)	-1.14	0.00
log(Kelloggs_price)	-0.69	0.33
log(General.Mills_price)	0.44	0.43
log(Kraft_price)	0.02	0.96
log(Nabisco...Kraft_price)	-0.16	0.52
log(Quaker.Oats_price)	0.02	0.94
log(Ralston_price)	-0.30	0.26

General Miller

	Estimate	Pr(>|t|)
(Intercept)	-4.63	0.42
log(General.Mills_price)	-1.59	0.01
log(Kelloggs_price)	-1.38	0.06
log(Dominicks_price)	0.08	0.80
log(Kraft_price)	0.09	0.87
log(Nabisco...Kraft_price)	0.25	0.34
log(Quaker.Oats_price)	-0.01	0.97
log(Ralston_price)	-0.38	0.16

Kelloggs

	Estimate	Pr(>|t|)
(Intercept)	3.71	0.41
log(Kelloggs_price)	-2.75	0.00
log(Dominicks_price)	0.49	0.06
log(General.Mills_price)	0.57	0.22
log(Kraft_price)	0.48	0.26
log(Nabisco...Kraft_price)	0.12	0.54
log(Quaker.Oats_price)	0.10	0.64
log(Ralston_price)	-0.16	0.47

Kraft

	Estimate	Pr(>|t|)
(Intercept)	6.48	0.23
log(Kraft_price)	-0.97	0.06
log(Dominicks_price)	0.28	0.38
log(General.Mills_price)	0.23	0.68
log(Kelloggs_price)	0.54	0.44
log(Nabisco...Kraft_price)	-0.13	0.58
log(Quaker.Oats_price)	0.00	0.99
log(Ralston_price)	-0.23	0.39

N/K

	Estimate	Pr(>|t|)
(Intercept)	12.16	0.04
log(Nabisco...Kraft_price)	-0.80	0.00
log(Dominicks_price)	-0.01	0.97
log(General.Mills_price)	0.57	0.33
log(Kelloggs_price)	0.42	0.57
log(Kraft_price)	0.64	0.24
log(Quaker.Oats_price)	0.37	0.19
log(Ralston_price)	-0.27	0.32

Quaker Oats

	Estimate	Pr(>|t|)
(Intercept)	9.24	0.14
log(Quaker.Oats_price)	-1.37	0.00
log(Dominicks_price)	-0.01	0.97
log(General.Mills_price)	0.79	0.22
log(Kelloggs_price)	-0.06	0.94
log(Kraft_price)	0.94	0.11
log(Nabisco...Kraft_price)	0.27	0.33
log(Ralston_price)	-0.18	0.55

Ralston

	Estimate	Pr(>|t|)
(Intercept)	5.79	0.41
log(Ralston_price)	-2.27	0.00
log(Dominicks_price)	0.91	0.03
log(General.Mills_price)	1.21	0.10
log(Kelloggs_price)	-1.31	0.15
log(Kraft_price)	0.96	0.15
log(Nabisco...Kraft_price)	-0.01	0.98
log(Quaker.Oats_price)	0.06	0.86

Conclusion

After looking at the market on the manufacturer's level, we ran some basic descriptive statistics to get a feel for the data. This allowed us to find some good choices for instrumental variables and then to check the feasibility of these instruments. We then tested whether a change in price of one manufacturer would cause a change in demand for another. However, the results we found were not statistically significant. So we fail to reject the null hypothesis, that a change in price of one manufacturer causes no change in quantity demanded for another.