

Modeling Consumer Response to EDP Changes

Eric Schulman

September 2017

The goal of this project is to quantify the relationship between a consumer product's everyday price (EDP) and its volume. Quantifying this relationship can help predict how EDP may influence expected future volumes of certain consumer products. I make these predictions using an econometric model inspired by Winer's 1986 paper [4]. In this paper, Winer models the probability buying a brand's product. His model is a function of previous quantity sold, consumer price expectations, competitors price and advertising spending. Winer used data directly from the point of sale, so my model is modified to take advantage of syndicated data from Nielson.

1 Introduction

1.1 Relevant Literature

Intuitively, a trade promotion leverages the relationship between current and future prices. During a promotion the manufacturer cuts the price of a good because they expect consumers to buy more while the price is low. Research into trade promotions often centers around the question of how past prices relate to current purchasing decisions. The channel that past prices use to effect consumers' future decisions involves the relationship between past prices and consumer expectations. When consumers see the past price they form an expectation about what prices should be. These expectations are called reference prices. These expectations effect consumer decisions by influencing how consumers perceive the product's quality. Putler's 1991 paper works out a mathematical model of consumer choice that incorporates reference prices [1]. In the paper, consumers calculate the difference between their reference prices and the actual price. This difference influences purchasing decisions.

Econometricians who try to estimate how promotions will relate to sales often include a variable to capture relationship between past and future price. In this way, the model includes reference prices. The model I take the most inspiration from is Winer's econometric model from 1986. In this model he predicts probability consumers buy from a brand as a function of previous quantity sold, consumers price expectations, competitors' price and advertising spending. [4]. The key difference between my model and his is that I am predicting volume changes and he is predicting purchase probabilities.

Another model worth mentioning is Krishnamurthi and Raj's econometric model which tries to predict the volume of the good purchased. They improve on Winer's model in their 1991 paper using an econometric technique called structural equations [2]. In this paper, Krishnamurthi and Raj separate consumers' decisions about which brand to purchase and the quantity of the brand to buy into separate economic decisions. Each of these decisions is modeled using a separate econometric equations. The fitted values from the first model are calculated and then used as a regressors in the second equation predicting quantity. This model relies on point of sales data to predict brand loyalties using the method proposed in Winer's paper. As a result, I cannot replicate the brand loyalty calculations. However, many of the variables involved with the volume model are also involved with brand loyalty.

The econometric literature does little to explore how consumers expectations about future prices which influence planned purchases. In other words, demand is isolated between periods. An econometric model is poorly suited to explore how past purchases and future expectations might influence future purchases. One model, Ahn, Gumus and Kaminsky create a model for manufacturers facing demand that carries over between periods [3] Their model is focused manufacturing decisions facing residual demand, but their inclusion of residual demand allows demand in past periods to carry over into the future. In their model, consumers wait until a future period for the price to fall. They only buy when price falls below their reservation price. Eventually they buy in the last period.

1.2 Winer's Model

My model takes inspiration from the model in Winer's 1986 paper. Formally, Winer's model describes the probability of purchasing brand i at time t . It takes into account past volumes, consumer price expectations, competitors price and advertising spending. Formally it is given by the equation:

$$Pr(BRAND_i)_{it} = \alpha_0 + \alpha_1 VOL_{it} + \alpha_2 ADV_{it} + \alpha_3 PRICE_REACT_{it} + \alpha_4 \frac{PRICE_{it}}{TOTAL_PRICE_t} + \epsilon_{it}$$

In Winer's model,

$$Pr(BRAND_i)_{it}$$

is a boolean variable representing whether or not the brand was purchased at time t before it is fit using the regression equation. This is a logistic regression. The data comes from the point of sale. If there are j brands then there $j - 1$ more data points are generated for all the brands that were not purchased.

$$VOL_{it}$$

represents the volume at of brand i at time t . Its coefficient represents the relationship between volume and purchase probability.

$$ADV_{it}$$

is a boolean based on the advertising spending of brand i at time t . It is an aggregate metric that takes into account various types of promotional spending.

$$PRICE_REACT_{it}$$

is meant to capture a reaction in the reference price. It is calculated as

$$\frac{PRICE_{it}}{\sum_j PRICE_{jt}} - \widehat{PRICE}_{it}$$

In the model, $PRICE_{it}$ is the price charged by brand i at time t .

The most important part of the reaction is

$$\widehat{PRICE}_{it}$$

which represents the consumers expectation for the price at the current time period. It involves estimating $PRICE_{it}$ as a function of $PRICE_{it-1}$. This process is called two stage least squares. In two stage least squares, you look at the effect of one variable 'through' another variable. In this case we are looking on how past prices effect the current price 'through' the current price. You do this by running a regression using a variable called an instrument (in this case, previous price) on another variable (in this case price). You use the predicted values from the first regression in a second regression. I plan to use this technique as well when estimating my model.

$$\frac{PRICE_{it}}{TOTAL_PRICE_t}$$

represents the ratio between the price of brand i at time t against its competitors. It represents the overall pricing environment. It is the current brand price at time t as a fraction of all the prices.

2 Model Description

My model emulates Winer's model. Although I do not have data on advertising expenditures I include boolean variables that describe the group to proxy for the fixed effects caused by each groups marketing departments. I include terms for previous volume and terms for the ratio between each products price and its competitors prices. Finally, I include past prices as variables. Like Winer, I estimated models that take advantage of two stage least squares. In these models, I include a term that represents how prices effect the volume changes through their relationship with previous prices.

In order to take advantage of syndicated data, my model predicts percentage changes in the volume sold by group i in consumer trade area (CTA) j at time

t . Estimates for this model are included in the preliminary results section:

$$\begin{aligned} \log(VOL_{ijt}) = & \alpha_0 + \alpha_1 PRICE_{ijt} + \alpha_2 EDP_{ijt} + \alpha_3 DAIRY_j + \alpha_4 FLAVOR_j + \\ & \alpha_5 CM_j + \alpha_6 DD_j + \alpha_7 ID_j + \alpha_8 PL_j + \alpha_9 SIZE32_j + \alpha_{10} SIZE64_{ij} + \alpha_{11} SIZE48_{ij} + \\ & \alpha_{12} PRICE_{ijt-1} + \alpha_{13} PRICE_{ijt-2} + \alpha_{14} VOL_RATIO_t + \alpha_{15} PRICE_RATIO_t + \\ & \alpha_{16} VOL_{ijt-1} + \alpha_{17} VOL_{ijt-2} + e_{ijt} \end{aligned}$$

2.1 Description of the Variables

$$\log(VOL_{ijt})$$

This variable can be interpreted as percentage changes in volume. In addition to looking at percentage changes as the main variable of interest, I looked at linear changes (δVOL_{ijt}) and linear volume VOL_{ijt} . Previous volume is overwhelming the most statistically important factor involved with predicting future volumes. This is the main reason for looking at percentage changes in volume. Additionally, looking at the logarithm of volume makes interpreting results easier.

$$PRICE_{ijt}$$

This is the price for CTA j for product group i at time t . The coefficient on price can be interpreted as the expected percentage change in volume corresponding to a 1 dollar increase in price all else equal.

$$EDP_{ijt}$$

This is the everyday price for CTA j in Group i at time t . The coefficient on EDP_{ijt} can be interpreted as the expected percentage change in volume corresponding to a 1 dollar increase in EDP all else equal.

$$DAIRY_j$$

This is a boolean variable that says whether group i contains dairy. The coefficient on $DAIRY_j$ can be interpreted as the fixed effect of being a dairy product on volume changes.

$$FLAVOR_j$$

This is a boolean variable that says whether group i is a flavored creamer. The coefficient on this variable can be interpreted as the fixed effects of brand on volume changes.

$$CM_j, DD_j, ID_j, PL_j$$

These are boolean variable that represent the brand. If all four boolean variables are zero, then the brand is BA. The coefficient on these variables can be interpreted as the fixed effects of brand on volume changes.

$$SIZE32_j, SIZE64_j, SIZE48_j$$

These are boolean variables that represent the size of the product. If all three boolean variables are zero then the product is 16 ounce units. The coefficient on these variables can be interpreted as the fixed effects of size on volume changes.

$$PRICE_{ijt-1}$$

The coefficient on this variable represents the magnitude of a one dollar increase of price for group j in CTA i at week $t - 1$ would be expected to effect percentage changes in volume at week t

$$PRICE_{ijt-2}$$

This variable represents the price for group j in CTA i at week $t - 2$. After experimenting with various models, I chose to include the previous price going back two weeks. Including previous prices essentially 'divides' the coefficient on prices. However, the coefficients are both statistically significance beyond the five percent level. It seemed prudent to include them.

$$VOL_RATIO_t$$

This variable represents the volume of group i in CTA j as fraction of total volume. I ran models that included additional boolean variables to represent the CTA groups. In these models, I used volume as a fraction of CTA volume.

$$PRICE_RATIO_t$$

Represents price as a ratio of average price across all of the CTA groups. The coefficient represents how increasing the ratio is expected to change percentage change in volume between price and volume. I included this term specifically because Winer includes a similar term in his model to represent the contribution of competitors price.

$$VOL_{ijt-1}, VOL_{ijt-2}$$

Represent volume of the previous two weeks. The coefficients represent how changing volume from previous weeks is expected to change percentage changes in volume. Including a representing volume within a product's given market is an important third variable between price and volume.

2.2 CTA and Week Boolean Variables

I estimated two additional models involving boolean variables and included the results in the appendix. The first model includes boolean variables that represent each of the thirty CTAs.

$$CTA_i, WEEK_i$$

Additionally, total volume and average price needed to be adjusted to reflect average price and volume within the CTA. It is particularly important to adjust price to be restricted to the CTA. Without this adjustment, price loses its statistical significance within the model. The second regression includes 156 boolean variable for the weeks.

2.3 Two Stage Least Squares

Winer uses previous prices in his regression to do two stage least squares. This process involves estimating P_{it} as a function of P_{it-1} . In two stage least squares, you look at the effect of one variable 'through' another variable. In this case we are looking on how past prices effect the current price 'through' the current price.

By using this process we are assuming that the past prices can only relate to future volume changes 'through' consumers the current price. This assumption makes sense through the lense of the reference price literature. In the literature, past prices effect current purchasing decisions because its relationship to current prices. Consumers form expectations about prices and call these expectations reference prices. Theoretical models involving reference prices often involve an explicit assumption that previous prices only affect purchasing decisions through the reference price [1].

Using this technique involves running a regression using a variable called an instrument (in this case, previous price) on another variable (in this case price). The predicted values from the first regression are used in the second regression. I include results for the first stage of this process in the Appendix. The second stage results are included in the results section.

Stage One

$$\begin{aligned} PRICE_{ijt} = & \gamma_0 + \gamma_1 EDP_{ijt} + \gamma_2 DAIRY_j + \gamma_3 FLAVOR_j + \gamma_4 CM_j + \\ & \gamma_5 DD_j + \gamma_6 ID_j + \gamma_7 PL_j + \gamma_8 SIZE32_j + \gamma_9 SIZE64_{ij} + \gamma_{10} SIZE48_{ij} + \gamma_{11} PRICE_{ijt-1} + \\ & \gamma_{12} PRICE_{ijt-2} + \gamma_{13} VOL_RATIO_t + \gamma_{14} PRICE_RATIO_t + \gamma_{15} VOL_{ijt-1} + \\ & \gamma_{16} VOL_{ijt-2} + v_{ijt} \end{aligned}$$

Stage Two

$$\log(VOL_{ijt}) = \alpha_0 + \alpha_1 \widehat{PRICE}_{ijt} + \alpha_2 EDP_{ijt} + \alpha_3 DAIRY_j + \alpha_4 FLAVOR_j +$$

$$\alpha_5 CM_j + \alpha_6 DD_j + \alpha_7 ID_j + \alpha_8 PL_j + \alpha_9 SIZE32_j + \alpha_{10} SIZE64_{ij} + \alpha_{11} SIZE48_{ij} + \alpha_{12} VOL_RATIO_t + \alpha_{13} PRICE_RATIO_t + \alpha_{14} VOL_{ijt-1} + \alpha_{15} VOL_{ijt-2} + e_{ijt}$$

3 Results

3.1 Predicting Volume

I looked at models with volume, volume changes, and percent changes in volume as my dependent variable. In all cases, previous volume has the most statistical weight in predicting next weeks volume. As you can see in the table below, using just price, and the previous two weeks of volume explains 88 percent of the variance in volumes. On the other hand, using group characteristics and previous prices only explains 25 percent of the variance in weekly volumes. Below I include the results of a model predicting volumes without using previous volume as a regressor. Results for these models are included in the Appendix.

When looking at changes in volume and linear volume category characteristics do not carry as much weight in predicting volume as previous volumes. The fixed effects generated by product characteristics emerge when estimating the log linear model. Using $\log(VOL)$ (i.e. percentage change in volume) as the regressand term helps distinguish how price and product characteristics contribute to volume changes. Below are preliminary results. In the described above, all coefficients are statistically significant coefficients at the five percent level.

| | | | |
|--------------------------|------------------|----------------------------|-------------|
| Dep. Variable: | log(VOL) | R-squared: | 0.650 |
| Model: | OLS | Adj. R-squared: | 0.650 |
| Method: | Least Squares | F-statistic: | 6835. |
| Date: | Wed, 30 Aug 2017 | Prob (F-statistic): | 0.00 |
| Time: | 16:32:46 | Log-Likelihood: | -1.0449e+05 |
| No. Observations: | 62567 | AIC: | 2.090e+05 |
| Df Residuals: | 62549 | BIC: | 2.092e+05 |
| Df Model: | 17 | | |

| | coef | std err | t | P> t | [0.025 | 0.975] |
|-----------------------|-----------|--------------------------|-----------|-------|----------|----------|
| CONST | 9.1761 | 0.066 | 139.597 | 0.000 | 9.047 | 9.305 |
| PRICE | -1.6814 | 0.047 | -36.009 | 0.000 | -1.773 | -1.590 |
| EDP | 1.2369 | 0.049 | 25.486 | 0.000 | 1.142 | 1.332 |
| DAIRY | 1.0814 | 0.017 | 62.230 | 0.000 | 1.047 | 1.115 |
| FLAVOR | 1.6928 | 0.015 | 109.772 | 0.000 | 1.663 | 1.723 |
| CM | 1.7129 | 0.029 | 60.098 | 0.000 | 1.657 | 1.769 |
| DD | -0.9284 | 0.027 | -34.254 | 0.000 | -0.982 | -0.875 |
| ID | 2.4350 | 0.030 | 81.775 | 0.000 | 2.377 | 2.493 |
| PL | 2.2837 | 0.029 | 79.620 | 0.000 | 2.228 | 2.340 |
| SIZE32 | 0.9108 | 0.025 | 36.283 | 0.000 | 0.862 | 0.960 |
| SIZE64 | 0.1227 | 0.029 | 4.198 | 0.000 | 0.065 | 0.180 |
| SIZE48 | -1.5484 | 0.040 | -39.006 | 0.000 | -1.626 | -1.471 |
| PREV PRICE 1 | 0.0475 | 0.047 | 1.018 | 0.309 | -0.044 | 0.139 |
| PREV PRICE 2 | -0.3911 | 0.041 | -9.607 | 0.000 | -0.471 | -0.311 |
| VOL RATIO | 63.3441 | 2.674 | 23.687 | 0.000 | 58.103 | 68.586 |
| PRICE RATIO | -0.2226 | 0.030 | -7.406 | 0.000 | -0.281 | -0.164 |
| PREV VOL 1 | 1.796e-07 | 1.03e-08 | 17.351 | 0.000 | 1.59e-07 | 2e-07 |
| PREV VOL 2 | 1.548e-07 | 9.7e-09 | 15.954 | 0.000 | 1.36e-07 | 1.74e-07 |
| Omnibus: | 12349.164 | Durbin-Watson: | 0.063 | | | |
| Prob(Omnibus): | 0.000 | Jarque-Bera (JB): | 32280.257 | | | |
| Skew: | -1.076 | Prob(JB): | 0.00 | | | |
| Kurtosis: | 5.784 | Cond. No. | 1.39e+09 | | | |

Below is the table with just EDP as an explanatory variable. Removing the EDP causes most of the coefficients to intensify in magnitude.

| | | | |
|--------------------------|------------------|----------------------------|-------------|
| Dep. Variable: | LOG(VOL) | R-squared: | 0.646 |
| Model: | OLS | Adj. R-squared: | 0.646 |
| Method: | Least Squares | F-statistic: | 7147. |
| Date: | Thu, 31 Aug 2017 | Prob (F-statistic): | 0.00 |
| Time: | 20:41:50 | Log-Likelihood: | -1.0482e+05 |
| No. Observations: | 62567 | AIC: | 2.097e+05 |
| Df Residuals: | 62550 | BIC: | 2.098e+05 |
| Df Model: | 16 | | |

| | coef | std err | t | P> t | [0.025 | 0.975] |
|-----------------------|-----------|----------|--------------------------|-----------|----------|----------|
| CONST | 9.9765 | 0.058 | 171.879 | 0.000 | 9.863 | 10.090 |
| PRICE | -1.3185 | 0.045 | -29.497 | 0.000 | -1.406 | -1.231 |
| DAIRY | 1.1773 | 0.017 | 69.040 | 0.000 | 1.144 | 1.211 |
| FLAVOR | 1.7980 | 0.015 | 120.385 | 0.000 | 1.769 | 1.827 |
| CM | 1.7379 | 0.029 | 60.694 | 0.000 | 1.682 | 1.794 |
| DD | -0.8637 | 0.027 | -31.843 | 0.000 | -0.917 | -0.811 |
| ID | 2.4216 | 0.030 | 80.920 | 0.000 | 2.363 | 2.480 |
| PL | 2.0709 | 0.028 | 75.081 | 0.000 | 2.017 | 2.125 |
| SIZE32 | 0.6769 | 0.023 | 28.821 | 0.000 | 0.631 | 0.723 |
| SIZE64 | -0.2144 | 0.026 | -8.187 | 0.000 | -0.266 | -0.163 |
| SIZE48 | -1.9155 | 0.037 | -51.518 | 0.000 | -1.988 | -1.843 |
| PREV PRICE 1 | 0.1972 | 0.047 | 4.241 | 0.000 | 0.106 | 0.288 |
| PREV PRICE 2 | -0.0218 | 0.038 | -0.570 | 0.569 | -0.097 | 0.053 |
| VOL RATIO | 63.9018 | 2.688 | 23.773 | 0.000 | 58.633 | 69.170 |
| PRICE RATIO | -0.2024 | 0.030 | -6.702 | 0.000 | -0.262 | -0.143 |
| PREV VOL 1 | 1.745e-07 | 1.04e-08 | 16.779 | 0.000 | 1.54e-07 | 1.95e-07 |
| PREV VOL 2 | 1.657e-07 | 9.74e-09 | 16.999 | 0.000 | 1.47e-07 | 1.85e-07 |
| Omnibus: | 11885.013 | | Durbin-Watson: | 0.054 | | |
| Prob(Omnibus): | 0.000 | | Jarque-Bera (JB): | 29726.750 | | |
| Skew: | -1.054 | | Prob(JB): | 0.00 | | |
| Kurtosis: | 5.639 | | Cond. No. | 1.39e+09 | | |

Additionally, All of the CTA dummy variables are statistically significant at the five percent levels. When using an F-test for join significance of these variables, they are significant at the five percent level. This means that these variables are related to volume changes and should not be taken lightly. Additionally, when including the CTA dummies, total regional volume is no longer statistically significant. The weekly dummy variables appear to be less important to the overall results. They pass an f-test for joint significant at the 5 percent, but considering the number of data points this is unsurprising. Most of the values do not pass individual t-tests. Due to the size of the table, I include the results of the CTA boolean variables in the appendix.

3.2 Two Stage least squares

Having higher p-values than the group characteristics and previous volume suggests that previous price may not be directly correlated with percentage changes in volume. This suggests it would make a good instrument for volumes, especially when you consider the economic mechanism proposed in the reference price literature. Running the regression causes the coefficient on price to fall. Additionally, the product characteristics are statistically significant after running the two stage model. This suggests that price and product characteristics are highly related. After cleaning price of its endogenous relationship with the other variables in the first stage, these characteristics become statistically relevant in the context of predicting volume changes. Below are the results of

running the two stage regression. Running the two stage regression produces the results in the table below.

| | | | |
|--------------------------|------------------|----------------------------|-------------|
| Dep. Variable: | log(VOL) | R-squared: | 0.642 |
| Model: | OLS | Adj. R-squared: | 0.642 |
| Method: | Least Squares | F-statistic: | 7486. |
| Date: | Wed, 30 Aug 2017 | Prob (F-statistic): | 0.00 |
| Time: | 17:31:18 | Log-Likelihood: | -1.0518e+05 |
| No. Observations: | 62567 | AIC: | 2.104e+05 |
| Df Residuals: | 62551 | BIC: | 2.105e+05 |
| Df Model: | 15 | | |

| | coef | std err | t | P> t | [0.025 | 0.975] |
|--------------------|-----------|----------|---------|-------|----------|----------|
| CONST | 9.1772 | 0.066 | 138.090 | 0.000 | 9.047 | 9.307 |
| FIT PRICE | -2.0288 | 0.086 | -23.493 | 0.000 | -2.198 | -1.860 |
| EDP | 1.1568 | 0.063 | 18.293 | 0.000 | 1.033 | 1.281 |
| DAIRY | 1.0811 | 0.018 | 61.528 | 0.000 | 1.047 | 1.116 |
| FLAVOR | 1.6965 | 0.016 | 108.671 | 0.000 | 1.666 | 1.727 |
| CM | 1.7094 | 0.029 | 58.267 | 0.000 | 1.652 | 1.767 |
| DD | -0.9226 | 0.028 | -33.313 | 0.000 | -0.977 | -0.868 |
| ID | 2.4333 | 0.030 | 80.123 | 0.000 | 2.374 | 2.493 |
| PL | 2.2807 | 0.030 | 76.651 | 0.000 | 2.222 | 2.339 |
| SIZE32 | 0.9101 | 0.026 | 35.606 | 0.000 | 0.860 | 0.960 |
| SIZE64 | 0.1199 | 0.030 | 4.048 | 0.000 | 0.062 | 0.178 |
| SIZE48 | -1.5486 | 0.040 | -38.581 | 0.000 | -1.627 | -1.470 |
| VOL RATIO | 52.2098 | 2.898 | 18.017 | 0.000 | 46.530 | 57.890 |
| PRICE RATIO | -0.1048 | 0.044 | -2.395 | 0.017 | -0.191 | -0.019 |
| PREV VOL 1 | 1.673e-07 | 9.16e-09 | 18.263 | 0.000 | 1.49e-07 | 1.85e-07 |
| PREV VOL 2 | 2.042e-07 | 9.19e-09 | 22.220 | 0.000 | 1.86e-07 | 2.22e-07 |

| | | | |
|-----------------------|-----------|--------------------------|-----------|
| Omnibus: | 12110.341 | Durbin-Watson: | 0.074 |
| Prob(Omnibus): | 0.000 | Jarque-Bera (JB): | 30948.108 |
| Skew: | -1.064 | Prob(JB): | 0.00 |
| Kurtosis: | 5.709 | Cond. No. | 1.49e+09 |

References

- [1] Daniel S. Putler. Incorporating reference price effects into a theory of consumer choice. *Marketing Science*, 11(3):287–309, 7 1992.
- [2] Lakshman Krishnamurthi , S. P. Raj. An empirical analysis of the relationship between brand loyalty and consumer price elasticity. *Marketing Science*, 10(2):172–183, 3 1991.
- [3] Hyun soo Ahn , Mehmet Gumus , Philip Kaminsky. Pricing and manufacturing decisions when demand is a function of prices in multiple periods. *Operations Research*, 55(6):1039–1057, 11 2007.

- [4] Russell S. Winer. A reference price model of brand choice for frequently purchased products. *Journal of Consumer Research*, 13(2):250–256, 9 1986.

A Appendix Volume Results

This table below shows how previous volume can be used to predict future volumes.

| | | | | | | |
|-------------------|------------------|---------------------|--------------|-------|-----------|-----------|
| Dep. Variable: | VOL | R-squared: | 0.886 | | | |
| Model: | OLS | Adj. R-squared: | 0.886 | | | |
| Method: | Least Squares | F-statistic: | 1.626e+05 | | | |
| Date: | Thu, 31 Aug 2017 | Prob (F-statistic): | 0.00 | | | |
| Time: | 19:59:20 | Log-Likelihood: | -9.2080e+05 | | | |
| No. Observations: | 62567 | AIC: | 1.842e+06 | | | |
| Df Residuals: | 62563 | BIC: | 1.842e+06 | | | |
| Df Model: | 3 | | | | | |
| | | | | | | |
| | coef | std err | t | P> t | [0.025 | 0.975] |
| const | 3.103e+05 | 8792.933 | 35.286 | 0.000 | 2.93e+05 | 3.28e+05 |
| PRICE | -1.685e+05 | 5117.634 | -32.926 | 0.000 | -1.79e+05 | -1.58e+05 |
| PREV VOL 1 | 0.5407 | 0.004 | 148.953 | 0.000 | 0.534 | 0.548 |
| PREV VOL 2 | 0.4112 | 0.004 | 113.149 | 0.000 | 0.404 | 0.418 |
| | | | | | | |
| Omnibus: | 62854.779 | Durbin-Watson: | 2.246 | | | |
| Prob(Omnibus): | 0.000 | Jarque-Bera (JB): | 19005091.109 | | | |
| Skew: | 4.361 | Prob(JB): | 0.00 | | | |
| Kurtosis: | 87.936 | Cond. No. | 1.13e+07 | | | |

The table below shows how group characteristics can predict volumes. Without taking into account percentage changes in volume, they appear to be a poor predictor.

| | | | |
|--------------------------|------------------|----------------------------|-------------|
| Dep. Variable: | VOL | R-squared: | 0.257 |
| Model: | OLS | Adj. R-squared: | 0.257 |
| Method: | Least Squares | F-statistic: | 1546. |
| Date: | Thu, 31 Aug 2017 | Prob (F-statistic): | 0.00 |
| Time: | 20:39:15 | Log-Likelihood: | -9.7952e+05 |
| No. Observations: | 62567 | AIC: | 1.959e+06 |
| Df Residuals: | 62552 | BIC: | 1.959e+06 |
| Df Model: | 14 | | |

| | coef | std err | t | P> t | [0.025 | 0.975] |
|-----------------------|-------------|--------------------------|-------------|-----------------|---------------|---------------|
| CONST | -3.749e+06 | 7.64e+04 | -49.063 | 0.000 | -3.9e+06 | -3.6e+06 |
| PRICE | -1.624e+06 | 5.03e+04 | -32.273 | 0.000 | -1.72e+06 | -1.53e+06 |
| EDP | 1.295e+06 | 5.73e+04 | 22.619 | 0.000 | 1.18e+06 | 1.41e+06 |
| DIARY | 1.005e+06 | 2.02e+04 | 49.818 | 0.000 | 9.65e+05 | 1.04e+06 |
| FLAVOR | 6.009e+05 | 1.81e+04 | 33.171 | 0.000 | 5.65e+05 | 6.36e+05 |
| CM | 2.195e+06 | 3.26e+04 | 67.421 | 0.000 | 2.13e+06 | 2.26e+06 |
| DD | 4.15e+05 | 3.21e+04 | 12.939 | 0.000 | 3.52e+05 | 4.78e+05 |
| ID | 2.503e+06 | 3.38e+04 | 74.167 | 0.000 | 2.44e+06 | 2.57e+06 |
| PL | 3.046e+06 | 3.17e+04 | 96.155 | 0.000 | 2.98e+06 | 3.11e+06 |
| SIZE32 | 2.045e+06 | 2.86e+04 | 71.614 | 0.000 | 1.99e+06 | 2.1e+06 |
| SIZE64 | 9.073e+05 | 3.44e+04 | 26.351 | 0.000 | 8.4e+05 | 9.75e+05 |
| SIZE48 | 4.081e+05 | 4.7e+04 | 8.679 | 0.000 | 3.16e+05 | 5e+05 |
| PRICE RATIO | 4.56e+05 | 4.82e+04 | 9.468 | 0.000 | 3.62e+05 | 5.5e+05 |
| PREV PRICE 1 | 6.346e+04 | 4.28e+04 | 1.482 | 0.138 | -2.05e+04 | 1.47e+05 |
| PREV PRICE 2 | -4.06e+04 | 3.52e+04 | -1.153 | 0.249 | -1.1e+05 | 2.84e+04 |
| Omnibus: | 55090.052 | Durbin-Watson: | 0.161 | | | |
| Prob(Omnibus): | 0.000 | Jarque-Bera (JB): | 2220030.171 | | | |
| Skew: | 4.165 | Prob(JB): | 0.00 | | | |
| Kurtosis: | 30.968 | Cond. No. | 63.5 | | | |

The table below combines previous volumes with product characteristics to predict future volumes.

| | | | |
|--------------------------|------------------|----------------------------|-------------|
| Dep. Variable: | VOL | R-squared: | 0.972 |
| Model: | OLS | Adj. R-squared: | 0.972 |
| Method: | Least Squares | F-statistic: | 1.270e+05 |
| Date: | Wed, 30 Aug 2017 | Prob (F-statistic): | 0.00 |
| Time: | 17:30:49 | Log-Likelihood: | -8.7714e+05 |
| No. Observations: | 62567 | AIC: | 1.754e+06 |
| Df Residuals: | 62549 | BIC: | 1.754e+06 |
| Df Model: | 17 | | |

| | coef | std err | t | P> t | [0.025 | 0.975] |
|-----------------------|------------|--------------------------|--------------|-------|-----------|-----------|
| CONST | -1.188e+05 | 1.52e+04 | -7.835 | 0.000 | -1.49e+05 | -8.91e+04 |
| PRICE | -2.707e+05 | 1.08e+04 | -25.121 | 0.000 | -2.92e+05 | -2.5e+05 |
| EDP | 1.21e+05 | 1.12e+04 | 10.803 | 0.000 | 9.9e+04 | 1.43e+05 |
| DAIRY | -2.024e+04 | 4009.879 | -5.049 | 0.000 | -2.81e+04 | -1.24e+04 |
| FLAVOR | 2.377e+04 | 3558.452 | 6.680 | 0.000 | 1.68e+04 | 3.07e+04 |
| CM | 4513.4702 | 6576.924 | 0.686 | 0.493 | -8377.314 | 1.74e+04 |
| DD | -1.657e+04 | 6254.244 | -2.650 | 0.008 | -2.88e+04 | -4315.086 |
| ID | -5294.8228 | 6870.938 | -0.771 | 0.441 | -1.88e+04 | 8172.229 |
| PL | 7.491e+04 | 6618.551 | 11.318 | 0.000 | 6.19e+04 | 8.79e+04 |
| SIZE32 | 5.316e+04 | 5792.548 | 9.178 | 0.000 | 4.18e+04 | 6.45e+04 |
| SIZE64 | 4.562e+04 | 6742.034 | 6.767 | 0.000 | 3.24e+04 | 5.88e+04 |
| SIZE48 | 3.222e+04 | 9159.842 | 3.517 | 0.000 | 1.43e+04 | 5.02e+04 |
| PREV PRICE 1 | 3.881e+05 | 1.08e+04 | 36.064 | 0.000 | 3.67e+05 | 4.09e+05 |
| PREV PRICE 2 | 9.708e+04 | 9393.262 | 10.335 | 0.000 | 7.87e+04 | 1.15e+05 |
| VOL RATIO | 2.234e+08 | 6.17e+05 | 362.095 | 0.000 | 2.22e+08 | 2.25e+08 |
| PRICE RATIO | -4.027e+05 | 6934.906 | -58.065 | 0.000 | -4.16e+05 | -3.89e+05 |
| PREV VOL 1 | 0.1537 | 0.002 | 64.360 | 0.000 | 0.149 | 0.158 |
| PREV VOL 2 | 0.1028 | 0.002 | 45.893 | 0.000 | 0.098 | 0.107 |
| Omnibus: | 46260.040 | Durbin-Watson: | 0.982 | | | |
| Prob(Omnibus): | 0.000 | Jarque-Bera (JB): | 10318654.001 | | | |
| Skew: | 2.585 | Prob(JB): | 0.00 | | | |
| Kurtosis: | 65.701 | Cond. No. | 1.39e+09 | | | |

The table below estimates volume changes using the variables from the previous model. It is more successful at predicting volumes, however not as successful as using $LOG(VOL)$ as the dependent variable.

| | | | |
|--------------------------|------------------|----------------------------|-------------|
| Dep. Variable: | ΔVOL | R-squared: | 0.282 |
| Model: | OLS | Adj. R-squared: | 0.282 |
| Method: | Least Squares | F-statistic: | 1637. |
| Date: | Thu, 31 Aug 2017 | Prob (F-statistic): | 0.00 |
| Time: | 19:24:17 | Log-Likelihood: | -9.1784e+05 |
| No. Observations: | 62567 | AIC: | 1.836e+06 |
| Df Residuals: | 62551 | BIC: | 1.836e+06 |
| Df Model: | 15 | | |

| | coef | std err | t | P> t | [0.025 | 0.975] |
|-----------------------|------------|--------------------------|--------------|-------|-----------|-----------|
| CONST | 2.188e+05 | 2.9e+04 | 7.542 | 0.000 | 1.62e+05 | 2.76e+05 |
| PRICE | -1.901e+06 | 1.9e+04 | -100.107 | 0.000 | -1.94e+06 | -1.86e+06 |
| EDP | 1.051e+05 | 2.14e+04 | 4.904 | 0.000 | 6.31e+04 | 1.47e+05 |
| DAIRY | -6.31e+04 | 7679.756 | -8.216 | 0.000 | -7.82e+04 | -4.8e+04 |
| FLAVOR | -3.198e+04 | 6812.641 | -4.694 | 0.000 | -4.53e+04 | -1.86e+04 |
| CM | -1.141e+05 | 1.26e+04 | -9.072 | 0.000 | -1.39e+05 | -8.95e+04 |
| DD | -1.899e+04 | 1.2e+04 | -1.585 | 0.113 | -4.25e+04 | 4500.494 |
| ID | -1.431e+05 | 1.31e+04 | -10.892 | 0.000 | -1.69e+05 | -1.17e+05 |
| PL | -1.532e+05 | 1.26e+04 | -12.146 | 0.000 | -1.78e+05 | -1.29e+05 |
| SIZE32 | -1.324e+05 | 1.11e+04 | -11.974 | 0.000 | -1.54e+05 | -1.11e+05 |
| SIZE64 | -5.736e+04 | 1.29e+04 | -4.443 | 0.000 | -8.27e+04 | -3.21e+04 |
| SIZE48 | -2.194e+04 | 1.76e+04 | -1.250 | 0.211 | -5.63e+04 | 1.25e+04 |
| PREV PRICE 1 | 2.281e+06 | 1.8e+04 | 126.860 | 0.000 | 2.25e+06 | 2.32e+06 |
| PREV PRICE 2 | -4.751e+05 | 1.6e+04 | -29.738 | 0.000 | -5.06e+05 | -4.44e+05 |
| VOL RATIO | 1.888e+07 | 4.48e+05 | 42.106 | 0.000 | 1.8e+07 | 1.98e+07 |
| PRICE RATIO | -4.587e+04 | 1.32e+04 | -3.486 | 0.000 | -7.17e+04 | -2.01e+04 |
| Omnibus: | 32047.211 | Durbin-Watson: | 2.745 | | | |
| Prob(Omnibus): | 0.000 | Jarque-Bera (JB): | 16022654.435 | | | |
| Skew: | -1.135 | Prob(JB): | 0.00 | | | |
| Kurtosis: | 81.364 | Cond. No. | 766. | | | |

A.1 Appendix CTA Boolean Variables Results

| | | | |
|--------------------------|------------------|----------------------------|-----------|
| Dep. Variable: | log(VOL) | R-squared: | 0.742 |
| Model: | OLS | Adj. R-squared: | 0.742 |
| Method: | Least Squares | F-statistic: | 3823. |
| Date: | Wed, 30 Aug 2017 | Prob (F-statistic): | 0.00 |
| Time: | 17:31:09 | Log-Likelihood: | -94973. |
| No. Observations: | 62567 | AIC: | 1.900e+05 |
| Df Residuals: | 62519 | BIC: | 1.905e+05 |
| Df Model: | 47 | | |

| | coef | std err | t | P> t | [0.025 | 0.975] |
|--------------|-----------|----------|---------|-------|----------|----------|
| CONST | 8.7230 | 0.065 | 134.763 | 0.000 | 8.596 | 8.850 |
| PRICE | -0.9790 | 0.036 | -26.955 | 0.000 | -1.050 | -0.908 |
| EDP | 0.8578 | 0.043 | 19.828 | 0.000 | 0.773 | 0.943 |
| DAIRY | 1.2511 | 0.016 | 79.503 | 0.000 | 1.220 | 1.282 |
| FLAVOR | 1.4695 | 0.014 | 104.185 | 0.000 | 1.442 | 1.497 |
| CTA 5 | 1.5271 | 0.026 | 58.271 | 0.000 | 1.476 | 1.579 |
| CM | -0.9021 | 0.024 | -38.006 | 0.000 | -0.949 | -0.856 |
| DD | 2.0167 | 0.027 | 74.198 | 0.000 | 1.963 | 2.070 |
| ID | 1.8490 | 0.026 | 71.102 | 0.000 | 1.798 | 1.900 |
| PL | 0.7108 | 0.023 | 30.959 | 0.000 | 0.666 | 0.756 |
| SIZE32 | 0.1031 | 0.027 | 3.883 | 0.000 | 0.051 | 0.155 |
| SIZE64 | -1.1738 | 0.036 | -33.034 | 0.000 | -1.243 | -1.104 |
| SIZE48 | -0.0677 | 0.039 | -1.742 | 0.081 | -0.144 | 0.008 |
| PREV PRICE 1 | -0.4059 | 0.035 | -11.623 | 0.000 | -0.474 | -0.337 |
| PREV PRICE 2 | 6.0870 | 0.055 | 111.150 | 0.000 | 5.980 | 6.194 |
| VOL RATIO | -0.5738 | 0.022 | -26.309 | 0.000 | -0.617 | -0.531 |
| PRICE RATIO | 7.056e-08 | 8e-09 | 8.821 | 0.000 | 5.49e-08 | 8.62e-08 |
| PREV VOL 1 | 1.112e-07 | 7.86e-09 | 14.135 | 0.000 | 9.57e-08 | 1.27e-07 |
| PREV VOL 2 | 1.0804 | 0.038 | 28.239 | 0.000 | 1.005 | 1.155 |
| CTA 1 | 0.9538 | 0.036 | 26.746 | 0.000 | 0.884 | 1.024 |
| CTA 2 | 0.8537 | 0.038 | 22.674 | 0.000 | 0.780 | 0.927 |
| CTA 3 | 1.0946 | 0.036 | 30.672 | 0.000 | 1.025 | 1.165 |
| CTA 4 | -0.5691 | 0.040 | -14.305 | 0.000 | -0.647 | -0.491 |
| CTA 5 | 1.0351 | 0.036 | 28.825 | 0.000 | 0.965 | 1.105 |
| CTA 6 | 0.8199 | 0.036 | 23.007 | 0.000 | 0.750 | 0.890 |
| CTA 7 | 0.8199 | 0.036 | 23.007 | 0.000 | 0.750 | 0.890 |
| CTA 8 | -1.1479 | 0.037 | -30.739 | 0.000 | -1.221 | -1.075 |
| CTA 9 | 0.7188 | 0.039 | 18.566 | 0.000 | 0.643 | 0.795 |
| CTA 10 | 1.1997 | 0.035 | 33.999 | 0.000 | 1.131 | 1.269 |
| CTA 11 | 0.8047 | 0.035 | 22.913 | 0.000 | 0.736 | 0.874 |
| CTA 12 | 0.6434 | 0.035 | 18.613 | 0.000 | 0.576 | 0.711 |
| CTA 13 | 0.7021 | 0.035 | 20.121 | 0.000 | 0.634 | 0.771 |
| CTA 14 | 0.4214 | 0.035 | 12.079 | 0.000 | 0.353 | 0.490 |
| CTA 15 | 0.4001 | 0.035 | 11.473 | 0.000 | 0.332 | 0.468 |
| CTA 16 | 0.7808 | 0.034 | 22.671 | 0.000 | 0.713 | 0.848 |
| CTA 17 | 0.8074 | 0.035 | 22.990 | 0.000 | 0.739 | 0.876 |
| CTA 18 | 0.5809 | 0.035 | 16.634 | 0.000 | 0.512 | 0.649 |
| CTA 19 | 1.1948 | 0.035 | 34.109 | 0.000 | 1.126 | 1.263 |
| CTA 20 | 0.5117 | 0.035 | 14.651 | 0.000 | 0.443 | 0.580 |
| CTA 21 | 0.6243 | 0.035 | 17.828 | 0.000 | 0.556 | 0.693 |
| CTA 22 | 0.8015 | 0.035 | 23.037 | 0.000 | 0.733 | 0.870 |
| CTA 23 | -0.0692 | 0.035 | -1.974 | 0.048 | -0.138 | -0.000 |
| CTA 24 | 1.2644 | 0.036 | 35.565 | 0.000 | 1.195 | 1.334 |
| CTA 25 | 1.3575 | 0.036 | 38.136 | 0.000 | 1.288 | 1.427 |
| CTA 26 | 1.1837 | 0.035 | 33.434 | 0.000 | 1.114 | 1.253 |
| CTA 27 | 1.4746 | 0.032 | 45.407 | 0.000 | 1.411 | 1.538 |
| CTA 28 | 1.0682 | 0.032 | 32.930 | 0.000 | 1.005 | 1.132 |
| CTA 29 | 1.9507 | 0.034 | 57.428 | 0.000 | 1.884 | 2.017 |

| | | | |
|-----------------------|-----------|--------------------------|-----------|
| Omnibus: | 17479.756 | Durbin-Watson: | 1.885 |
| Prob(Omnibus): | 0.000 | Jarque-Bera (JB): | 77054.962 |
| Skew: | -1.309 | Prob(JB): | 0.00 |
| Kurtosis: | 7.765 | Cond. No. | 8.52e+07 |

B Appendix Stage One Results

| | | | | | | |
|-------------------|------------------|---------------------|------------|-------|----------|----------|
| Dep. Variable: | PRICE | R-squared: | 0.946 | | | |
| Model: | OLS | Adj. R-squared: | 0.946 | | | |
| Method: | Least Squares | F-statistic: | 6.803e+04 | | | |
| Date: | Wed, 30 Aug 2017 | Prob (F-statistic): | 0.00 | | | |
| Time: | 17:29:40 | Log-Likelihood: | 49278. | | | |
| No. Observations: | 62567 | AIC: | -9.852e+04 | | | |
| Df Residuals: | 62550 | BIC: | -9.837e+04 | | | |
| Df Model: | 16 | | | | | |
| | coef | std err | t | P> t | [0.025 | 0.975] |
| CONST | 0.0118 | 0.006 | 2.090 | 0.037 | 0.001 | 0.023 |
| EDP | 0.3169 | 0.004 | 80.072 | 0.000 | 0.309 | 0.325 |
| DAIRY | 0.0019 | 0.001 | 1.280 | 0.200 | -0.001 | 0.005 |
| FLAVOR | 0.0080 | 0.001 | 6.061 | 0.000 | 0.005 | 0.011 |
| CM | 0.0454 | 0.002 | 18.665 | 0.000 | 0.041 | 0.050 |
| DD | 0.0368 | 0.002 | 15.892 | 0.000 | 0.032 | 0.041 |
| ID | 0.0328 | 0.003 | 12.889 | 0.000 | 0.028 | 0.038 |
| PL | 0.0526 | 0.002 | 21.493 | 0.000 | 0.048 | 0.057 |
| SIZE32 | -0.0287 | 0.002 | -13.348 | 0.000 | -0.033 | -0.024 |
| SIZE64 | -0.0212 | 0.003 | -8.462 | 0.000 | -0.026 | -0.016 |
| SIZE48 | 0.0027 | 0.003 | 0.791 | 0.429 | -0.004 | 0.009 |
| PREV PRICE 1 | 0.4339 | 0.004 | 120.639 | 0.000 | 0.427 | 0.441 |
| PREV PRICE 2 | -0.0428 | 0.003 | -12.300 | 0.000 | -0.050 | -0.036 |
| VOL RATIO | -23.5398 | 0.209 | -112.761 | 0.000 | -23.949 | -23.131 |
| PRICE RATIO | 0.3349 | 0.002 | 152.394 | 0.000 | 0.331 | 0.339 |
| PREV VOL 1 | 5.657e-08 | 8.57e-10 | 66.024 | 0.000 | 5.49e-08 | 5.82e-08 |
| PREV VOL 2 | 1.851e-08 | 8.28e-10 | 22.363 | 0.000 | 1.69e-08 | 2.01e-08 |
| Omnibus: | 16672.641 | Durbin-Watson: | 1.498 | | | |
| Prob(Omnibus): | 0.000 | Jarque-Bera (JB): | 164979.945 | | | |
| Skew: | -0.999 | Prob(JB): | 0.00 | | | |
| Kurtosis: | 10.700 | Cond. No. | 1.27e+09 | | | |