Modeling Consumer Response to EDP Changes

Eric Schulman

September 21, 2017

1 Introduction

The objective of this research project is predicting the consumer reaction to changes in every day prices (EDP). Predicting how consumers will respond to EDP changes can help firms adjust their everyday prices to maximize sales volume and revenue before and after running promotions. The point of this project is to utilize multiple approaches to get as full a picture of consumer behavior as possible. In particular, this study involves two approaches. The first involves a traditional econometric model. This leverages well understood statistical techniques to describe the relationship between EDP and consumer behavior. However, it does not identify causal relationships between the EDP and consumer behavior. The second is a dynamic program, a recursive maximization problem that models consumer purchasing decisions. The dynamic program better isolates the causal relationship between consumer behavior and EDP changes, however the results are influenced by the assumptions of the model.

1.1 Relevant Literature

Manufacturers run trade promotions because they expect consumers to buy more when prices are low. The way consumers perceive past and current prices are what make promotions effective. As a result, research into trade promotions asks about how past prices relate to current purchasing decisions. This project examines two mechanisms from the literature that describe how consumers consider past prices when making purchasing decisions.

In the first way, consumer use past prices as references when deciding whether or not to buy products during the promotion. Consumers use the difference between past prices as a heuristic for when to purchase. Putler's 1991 paper works out theoretical treatment of this mechanism [2]. More importantly, including the of effects of using past prices as references lends itself well to an econometric model. The most cited paper including reference price effects is Winer's econometric model from 1986. His model predicts the probability purchasing from a brand as a function of previous quantity sold, consumers price expectations, competitors' price and advertising spending. [5]. Krishnamurthi and Raj's econometric model improves on Winer's model by separating the consumers decisions into two separate, and simultaneous economic decisions. Consumers simultameously decide which brand to purchase and the quantity of the brand to buy [3].

In the second mechanism, consumers use past prices to form expectations about future prices. When consumers go to the supermarket, they devise a plan for future purchases based on their expectations for prices in the future. Under this mechanism, the consumer buys more when prices are low because they intend to buy less in the future because they expect prices to go up. An econometric model is poorly suited to capture the effect these expectations because of the intertwined nature of future and present prices and purchases. The recursive framework of a dynamic programming models better captures the way past affects the present. Ahn, Gumus and Kaminsky developed a model consumers will wait until a future period for the price to fall to save money [4] Their model focuses on manufacturers decisions in the face of demand carrying over between periods. More importantly, Gonul and Srinivasan specified a model focused on consumer decisions to wait for prices to fall. In their model, consumers make purchasing decisions by forming expectations about future prices [1]. They estimate the parameters in their model by iterating between solving their model and estimating the parameters using maximum likelihood estimation.

2 Econometric Model

The first approach to understanding consumer responses to EDP involves quantifying the statistical relationship between EDP and volume using a traditional econometric model. The statistical properties econometric models are well established. Even though econometric models poorly identify causality, understanding the correlation between product characteristics and volume can be informative.

2.1 Winer's Model

Winer's 1986 paper inspired the model used in this project. Krishnamurthi and Raj's econometric model is more nuanced because it breaks consumer decisions about brand and purchase quantity into separate econometric equations. Like Winer, their model relies on point of sales data. However, the syndicated data does not lend its self to their approach.

On the other hand, Winer's model can be adapted to the syndicated data. Winer estimates the probability of buying a brands' product as a function of previous quantity sold, consumer price expectations, competitors' prices and advertising spending. 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. It is given by the equation below followed by descriptions of the variables [5].

$$Pr(BRAND_i)_{it} = \alpha_0+\alpha_1 VOL_{it}+\alpha_2 ADV_{it}+\alpha_3 PRICE\_REACT_{it}+\alpha_4 \frac{PRICE_i t}{TOTAL\_PRICE_t}+ \epsilon_{it}$$

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

Winer's model is a logistic regression. The dependent variable is a boolean variable representing whether or not the brand was purchased at time t before fitting. The data comes from the point of sale, so when the consumer makes a purchase, the consumer implicitly did not buy the other brands were not purchased. In other words, if there are j brands then there $j-1$ more data points are generated for all the brands that were not purchased.

$$VOL_{it}$$

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

$$ADV_{it}$$

The above variable 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.

$$PRICE\_REACT_{it}$$

The above variable 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 formula, $PRICE_{it}$ is the price charged by brand i at time t. $\widehat{PRICE}_{it}$ represents the consumers expectation for the price at the current time period 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, we 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. We do this by running running a regression using a variable called an instrument (in this case, previous price) on another variable (in this case price). We use the predicted values from the first regression in a second regression. This project uses this technique as well when estimating the model.

$$\frac{PRICE_{it}}{TOTAL\_PRICE_t}$$

The above variable represents the ratio between the price of brand i at time t against its competitors to capture the over all pricing environment.

2.2 Model Description

The main model in this project emulates Winer's model with several modifications to suit our data set. In order to take advantage of syndicated data, the model predicts percentage changes in the volume sold by group i in consumer trade area (CTA) j at time t. Instead of advertising expenditures we include group characteristic boolean variables to proxy for fixed effects caused by each groups marketing departments. The estimated equations for this model along with selected other models are included in the appendix section.

$log(VOL_{ijt}) = \alpha_0 + \alpha_0 PRICE_{ijt} + \alpha_0 EDP_{ijt} + \sum_{k=3}^{11} \alpha_k GROUP_{ijt} + \alpha_{12} PRICE_{ijt-1} + \alpha_{13} PRICE_{ijt-2} + \alpha_{14} VOL\_RATIO_t + \alpha_{15} PRICE\_RATIO_t + \alpha_{16} VOL_{ijt-1} + \alpha_{17} VOL_{ijt-2} + \epsilon_{ijt}$

2.3 Description of the Variables

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

The above variable can be interpreted as percentage changes in volume. We also looked at linear volume (VOL_{ijt}) as a dependent variable. Previous volume is overwhelmingly 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.

$$PRICE_{ijt}, EDP_{ijt}$$

The variable $PRICE_{ijt}$ 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. The variable EDP_{ijt} 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.

$$\sum_{k=3}^{11} \alpha_k GROUP_{ijt}$$

The above variables are boolean variables describing the group characteristics. Their coefficients represent the fixed effects of these variables. The characteristics are whether or not the creamer was flavored, dairy-free, and which of the five brands it belonged too. The coefficient on these variables can be interpreted as the fixed effects of brand on volume changes. Finally, there are variables that represent the size of the product.

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

The coefficient on the above variables represents the magnitude of a one dollar increase of price for group j in CTA i at week $t-1$ and $t-2$ respectively would be expected to effect percentage changes in volume at week t. After experimenting with various models, we chose to include the previous price going back two weeks.

$$VOL\_RATIO_t, PRICE\_RATIO_t$$

The above variable VOL_RATIO_t represents the volume of group i in CTA j as fraction of total volume. We include it because it represents brand penetration. $PRICE\_RATIO_t$ 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. We included the term $PRICE\_RATIO_t$ specifically because Winer includes a similar term in his model to represent the contribution of competitors price.

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

The above variables represent volume of the previous two weeks. The coefficients represent how changing volume from previous weeks is expected to change percentage changes in volume.

$$CTA_i, WEEK_i$$

We estimated additional models involving boolean variables. The first model includes boolean variables that represent each of the thirty CTAs. In the model with CTA, $PRICE\_RATIO_t$ and $PRICE\_RATIO_t$ must be adjusted to reflect average price and volume within the CTA.

$$P\widehat{RIC}E_{ijt}$$

Like Winer, we used previous prices to do two stage least squares. This process involves estimating $PRICE_{ijt}$ as a function of $PRICE_{ijt-1}$ to create $\widehat{PRICE}_{ijt}$. Two stage least squares looks 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.

Using this process assumes that the past prices only relate to future volume changes 'through' consumers the current price. This assumption makes sense through the lenses of the reference price literature. In the literature, past prices effect current purchasing decisions because their relationship to current prices. Papers often invoke an explicit assumption that previous prices only affect purchasing decisions through the reference price [2].

3 Dynamic Program

The second approach to predicting consumer reactions to EDP involves using a dynamic program to model how consumers chose quantities of goods to buy. The problem involves minimizing the cost faced by consumers over multiple periods. The consumer considers past purchases, and expectations for future prices to simultaneously decide on a quantity to purchase and a plan for future purchases. The model parameters are estimated by iterating between solving the consumer's problem and a maximum likelihood estimation routine. Although the statistical properties of this approach are not as documented, changes in consumer behavior within the model can be better attributed to EDP changes.

4 Gonul and Srinivasan's Model

This project emulates Gonul and Srinivasan's approach to predicting consumer behavior [1]. Similar to Winer, Gonul and Srinivasan use point of sale data which tells them which brand each household purchased. The data also implicitly tells them brands that were not purchased.

To predict consumer choice, Gonul and Srinivasan estimate the costs associated with making a purchase or waiting for a promotion. The saving from future promotions factor into the costs of buying or waiting. In other words, consumers anticipate future promotions and make purchases accordingly. Gonul and Srinvasan use a recursive cost function $C(B_t)$ to calculate costs. The argument to the cost function, B_t is boolean variable representing whether the brand was purchased. Consumers minimize costs, which are discounted by δ, over each period.

$$\text{Minimize } E(\sum_t \delta^t C(B_t) + \epsilon_{it})$$

Gonul and Srinivasan's use maximum likelihood estimation to estimate the parameters of $C(B_t)$. They specify a likelihood function in terms of $C(B_t)$. Maximum likelihood estimation maximizes the likelihood function's probability, so parameters in $C(B_t)$ maximize the probability of the observed purchases. In order to build the likelihood function, Gonul and Sriviansan estimate the probability of making purchase. Below they calculate $Pr(B_t = 1)$ and specify the likelihood function. They assume there is a stochastic component ϵ_{it} to the costs. In the calculation, Φ is the cumulative distribution function for the standard normal distribution.

$$Pr(B_t = 1) = Pr(C(B_t = 1) + \epsilon_{1t} < C(B_t = 0) + \epsilon_{0t})$$

$$= Pr(\epsilon_{0t} + \epsilon_{1t} < C(B_t = 0) - C(B_t = 1)) = \Phi(C(B_t = 0) - C(B_t = 1)))$$

They combine these probabilities in the likelihood function below. Gonul and Srinivasan initialized values for parameters in cost function and then iterated between the likelihood function and the dynamic program until the parameters in the cost function converged. The likelihood function is given below.

$$\text{Maximize} \prod_t Pr(B_t = 1)^{B_t} Pr(B_t = 0)^{1-B_t}$$

4.1 Model Specification

The model in this paper modifies Gonul and Srinivasan's model to use syndicated data. Instead of modeling purchases as a binary decision, purchases are based on volumes. Consumers have a target consumption of $\bar{x}$ in the cost function. They incur a penalty for consuming more or less. Having a consumption target allows the consumer to deviate from this target based on prices and purchase more during promotions and less afterwards, as expected. The consumer's cost function is given as:

$$C(x_0) = \alpha_1(x_1 + x_0 - \bar{x})^2 + p_1 x_1$$

Like in Gonul and Srinivasan's model, the consumer makes purchases by trying to minimize the future costs associated with that purchase. When calculating these costs, the consumer must consider their future costs based on their current stockpiles, prices, and their expectations for future prices. To calculate future costs, the consumer implicitly forms a plan for future purchases. This leads to the recursive formulation below. Consumers try to minimize costs, which are discounted by δ in each period.

$$\text{Minimize }_x \sum_t \delta^t (x_t + x_{t-1} - \bar{x})^2 + px$$

The two key take aways of the dynamic program are (1) Consumers minimize costs over future periods. They can carry over goods from period to period, so they can 'stock up' on goods during promotions. (2) Consumers do not know future prices. If consumers knew the future prices, then they would know their future costs with certainty. They would always follow the optimal purchase plan. Because they do not know prices, consumers must adapt their purchases to unexpected promotions. Of course, The way consumers expectations will influence the model's results.

5 Preliminary Regression Results

At this stage, we have estimated various versions of the econometric model. The coefficient on EDP is persistently positive through all of the estimated models. Perhaps a higher EDP makes promotions more effective by signaling a brand's quality. Consumers may feel that they get a better value when the price starts out high. We estimated a model without EDP as an explanatory variable. Removing this variable causes the coefficients on group characteristics to slightly increase in magnitude. This supports the theory that EDP carries some information about the quality of the product. If EDP contained information about the product quality, the group characteristics would exhibit upward bias without it, increasing the value of the group coefficients. Removing EDP also causes the coefficient on price to decrease in magnitude.

In all versions of the model, past volume explains most of the variance in future volumes. As we can see in the appendix, using just price, and the previous two weeks of volume explains 88 percent of the variance in volumes. This is not very informative about what firms can do to increase volumes. Estimating a log linear model (i.e. $log(VOL)$) reveals more about the relationship between the other characteristics and volume. This variable can be interpreted as percentage change in volume. It helps distinguish how price and product characteristics contribute to volume changes.

Finally, the mechanism proposed in the literature regarding reference prices may be correct. Previous prices may only be correlated with purchase 'through' their relationship with present prices. Depending on the model, previous prices may not be statistically significant at the five percent level. Running the two stage regression causes the coefficient on price to fall and the p-values on the product characteristics to increase.

6 Conclusion

This project utilizes multiple approaches for a full picture of consumer behavior in response to EDP changes. The econometric model describes the relationship between various factors and volumes using established statistical techniques. It builds on the intuition that consumers use prices decreases as a psychological heuristic for making purchases. It also builds on the large literature of econometric models trying to predict how price changes affect sales outcomes. Counter-intuitively, EDP is consistently correlated with a positive increase in volume. A higher EDP may make promotions more effective. Additionally, past volumes are the most important statistical factor in explaining future volumes.

The dynamic programming model is a more involved approach toward predicting consumer behavior. The dynamic program establishes more causality between EDP changes and consumer behavior than the econometric model. It captures the intuition that consumers buy more because they expect future prices to increase the next time they visit the store. However, the dynamic program makes more assumptions and uses less established techniques to make predictions. The next steps for involve estimating the dynamic programming model and extending it to include multiple brands. Additionally, exploring other mathematical tools like high dimensional model representation can aid this project.

References

[1] Fusun Gonul and Kannan Srinivasan. Estimating the impact of consumer expectations of coupons on purchase behavior: A dynamic structural model. *Marketing Science*, 15(3):262–279, 1996.

[2] Daniel S. Putler. Incorporating reference price effects into a theory of consumer choice. *Marketing Science*, 11(3):287–309, 7 1992.

[3] 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.

[4] 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.

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

Appendices

A Past Volumes Regression Results

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

B Volumes Regression Results

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

C Log(VOL) Regression Results

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

D No EDP Regression Results

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

E 2 Stage Least Squares Regression Results

E.1 Stage 1

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

E.2 Stage 2

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

F CTA Boolean Variables Regression 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