Consumer Attention, Engagement, and Market Shares: Evidence from the Carbonated Soft Drinks Market

Yizao Liu
University of Connecticut
yizao.liu@uconn.edu

Huaxia Rui
University of Rochester
huaxia.rui@simon.rochester.edu

Selected Paper prepared for presentation at the 2014 AAEA/EAAE/CAES Joint Symposium: Social Networks, Social Media and the Economics of Food, Montreal, Canada, 29-30 May 2014

Consumer Attention, Engagement, and Market Shares: Evidence from the Carbonated Soft Drinks Market

Yizao Liu
University of Connecticut
yizao.liu@uconn.edu

Huaxia Rui
University of Rochester
huaxia.rui@simon.rochester.edu

April 3, 2014

PRELIMINARY AND INCOMPLETE

Abstract

Social media platforms facilitate a real-time "two-way" communication channels among consumers and between consumers and brands. Social media users can now interact with brands directly through Facebook and Twitter. These new features make social media a very distinctive class of online WOM. In this paper, we formulate a random coefficient discrete choice model of consumer demand to study whether and how consumer engagement and attention on the Internet affect the consumption of carbonated soft drinks (CSDs). We model consumer attention and engagement on social media as goodwill in order to capture the carry-over effects of WOM's impact on demand and combine two types of product level data: monthly CSD sales data and social media data. Our results suggest that the three types of social media messages all have significant and positive effects on CSD demand. In particular, indirect engagement on Twitter has the largest impact on consumer demand, followed by direct engagement on Twitter and consumer engagement on Facebook.

Keywords: Consumer Engagement, Consumer Attention, Social Media, Consumer Demand

1 Introduction

Since the Internet revolution, online Word-of-mouth (WOM) has assumed a prominent role both in people's everyday life. Consumers today rely more and more on various sources of online WOM to decide which brand of consumer products to buy, which restaurants to go, or even which doctors to visit. In response to the popularity of online WOM, numerous studies have examined the effect of online WOM on movie box office revenues (Chintagunta, Gopinath and Venkataraman, 2010; Rui, Liu and Whinston, 2013), on TV ratings (Godes and Mayzlin, 2004), on book sales (Chevalier and Mayzlin, 2006), etc.

In recent years, the landscape of online word of mouth has been significantly changed by the rise of social media platforms such as Facebook and Twitter. Unlike the traditional online product review sites where product reviews are infrequently posted on certain websites and are pulled by consumers when they actively try to find out those reviews, social media users post and pass messages to friends or followers in a much more interactive and realtime fashion. Moreover, social media users also interact with brands directly through Facebook and Twitter. In other words, social media platforms facilitate a real-time "two-way" communication channels among consumers and between consumers and brands. These new features make social media a very distinctive class of online WOM. This is probably particularly important to the carbonated soft drinks (CSD) industry as consumers rely less on product review information to guide their purchase of soft drink.

Another important factor that may reflect consumers' interest in certain brand is their online search behavior which can be captured by the Google trends search index given the dominant role Google plays in the Internet search industry. Google Trends search index measures the frequency of a keyword being searched in certain region over certain period of time and the search index is available almost in real time. Indeed, there are plenty of papers that use Google Trends data to predict some real world phenomena. We call consumers' attention (measured by google search index) a passive form of social media, while consumers' engagement on Twitter and Facebook an active form of social media.

This paper is among the first in the literature to study whether and how consumer engagement and attention on the Internet affect the consumption of CSD. To adapt to the unique features of social media, we depart from the traditional approach of analyzing the effect of online WOM based on its sentiment (e.g, positive or negative). Instead, we distinguish social media messages based naturally on how they are used by users. For example, we identify *direct engagement* messages as social media content directly (and publicly) sent to brands by consumers and we identify *indirect engagement* messgages as social media content containing brand related hashtags. Such consumer engagement social media content (direct or indirect) are complemented by the traditional social media content containing the brand names.

In this paper, we formulate a random coefficient discrete choice model of consumer demand to capture the heterogeneity of consumer preferences. Following Nerlove-Arrow (1962), consumer attention and engagement on social media is modelled as goodwill in order to capture the carry-over effects of WOM's impact on demand. We then constructed our model from the conditional indirect utility of consumer purchasing each type of product. To study the effect of social media on CSD consumption, we combines two types of product level data: monthly CSD sales data and social media data. The monthly data on CSD sales, collected by the Nielsen Company, cover 4 designated market areas (DMAs) from January 2011 to October 2012. The social media data is collected from Twitter, Facebook, and Google for the same period.

Our results suggest that the three types of social media messages all have significant and positive effects on CSD demand. In particular, indirect engagement on Twitter has the largest impact on consumer demand, followed by direct engagement on Twitter and consumer engagement on Facebook.

The rest of the paper is organized as follows. We first describe all sources of data we used in Section 2. We then introduce the model and estimation in Section 3, and present our empirical findings in Section 4. In Section 5, we present the managerial implications and discussions of this study, and then conclude.

2 Data

This analysis combines two types of product level data: monthly CSD sales data and social media data. The monthly data on CSD sales, collected by the Nielsen Company, cover 12 DMAs from January 2011 to October 2012, including Atlanta, Boston, Chicago, Dallas, Detroit, Hartford, Los Angeles, Miami, New York, San Francisco, Seattle, and Syracuse. These data include market level sales data for supermarkets with more than $2 million annual sales, which consists of dollars sales, volume sales, and prices for 18 diet and regular CSDs. Of these 18 products, 5 are owned by Coca Cola; 7 are owned by Pepsi; and 6 are owned by Dr.Pepper. The dataset contains information on product characteristics (e.g. nutrition content and package), marketing (e.g. price and in-store displays), location, and time of each purchase. The richness of the data allows us to capture price and packaging variation of various national brand and private label soft drinks while controlling time, markets, and product characteristics.

Our social media data is collected from three different sources: Twitter, Facebook, and Google Trends. Twitter data is collected from Twitter public streams, which is a real-time random sample of all tweets generated.[1] From the sample, we match those tweets containing keywords that we are interested in and aggregrate them to monthly level. Google Trends data is downloaded from Google and then transformed so that the search index is comparable across keywords, cities, and time.

[1] https://dev.twitter.com/docs/streaming-apis/streams/public

Brand	Sugar (g/oz)	Sodium (mg/oz)	Caffeine (mg/oz)	Price (cents/oz)	Mkt Shr ‰
Coca Cola					
Coca Cola Diet	0	3.33	3.92	2.89	18.64
Coca Cola Regular	3.25	4.17	2.92	2.83	30.91
Coca Cola Zero Diet	0	3.33	2.92	2.97	5.5
Fanta Regular	3.67	4.58	0	2.61	2.96
Sprite Regular	3.17	5.83	0	2.88	8.39
Pepsi					
Pepsi Diet	0	2.92	2.92	2.65	12.42
Pepsi Regualr	3.42	2.5	3.17	2.54	23.52
Mountain Dew Code Red Reg.	3.75	8.75	4.5	2.7	0.52
Mountain Dew Diet	0	4.17	4.5	2.77	3.44
Mountain Dew Regular	3.83	5.42	4.5	2.81	10.11
Sierra Mist Free Diet	0	3.17	0	2.33	1.06
Sierra Mist Regular	3.25	3.17	0	2.54	2.67
Dr. Pepper					
Dr.Pepper Diet	0	4.58	3.5	2.9	3.21
Dr.Pepper Regular	3.33	4.58	3.5	2.92	6.88
7 Up Diet	0	5.42	0	2.6	1.79
7 Up Regular	3.17	3.33	0	2.53	3.57
Sunkist Regular	4.17	5.83	3.33	2.54	2.54
Diet Rite Pure Zero Diet	0	0	0	2.46	0.4

Table 1: Summary Statistics of Product Characteristics

We also collect data of consumer engagement on Facebook for a brand. Not all brands have a Facebook page. Therefore, in our data, 12 out of 18 brands have its own official Facebook page. For each posts by the page, the consumers can either click the “like” button if they read it and find it interesting, write down their “comments” under the posts, or “share” the post with their friends on their own Facebook pages. Either way, the consumers are actively engaging and interacting with the brands. Specifically, we collect the number of “like”, “comment”, and “share” for all posts by the brand from January 2011 to October 2012, and construct a variable ”facebook response” by adding all three types together since the majority is “like”.

Table 1 reports the summary statistics of the product characteristics of carbonated soft

Brand	Twitter Direct Engagement(@)	Twitter Indirect Engagement (#)	Google Trend Search	Facebook Response
Coca Cola				
Coca Cola Diet	2,845	2,772	8.82	17,591
Coca Cola Regular	2,845	2,772	8.82	192,699
Coca Cola Zero Diet	2,845	2,772	8.82	0
Fanta Regular	0	0	0	5,360
Sprite Regular	178	415	2.08	51,898
Pepsi				
Pepsi Diet	2,789	1,342	4.29	11,495
Pepsi Regualr	2,789	1,342	4.29	175,261
Mountain Dew Code Red Reg.	324	0	1.13	0
Mountain Dew Diet	324	0	1.13	6,621
Mountain Dew Regular	324	0	1.13	83,837
Sierra Mist Free Diet	0	0	0	0
Sierra Mist Regular	0	0	0	14,218
Dr. Pepper				
Dr.Pepper Diet	734	409	0	0
Dr.Pepper Regular	734	409	0	519,119
7 Up Diet	0	0	0	0
7 Up Regular	0	0	0	19,000
Sunkist Regular	0	0	0	40,011
Diet Rite Pure Zero Diet	0	0	0	0

Table 2: Summary Statistics of Social Media Engagement across Brands

drinks brands. We see that the average amount of calories, sugar and caffeine contained in regular/diet private label and national brand CSDs are similar but national brand CSDs do contain more sodium on average. Table 1 also reports the summary statistics of average unit price and market shares for different brands. The average unit prices are calculated from sales transactions recorded in the data in the 12 DMAs over the 2 years so they are actually market-share weighted. In general, the price of private label products are cheaper than national brands. For example, the average price of Coca Cola regular is 2.83 cents per oz while it costs 2.65 cents per oz for Pepsi Diet.

The market shares of various CSD products reported in Table 1 are calculated by volume sold. The market is defined as a general refreshment beverage market (Chan 2006, Lopez and Fantuzzi 2012). The market size in each market is the total volume consumption of CSD, liquid tea, fruit juice, milk, and bottled water, which is calculated as *per capita consumption* $\times$ *population*. Therefore, consumers have outside options of not purchasing CSD products. Among all brands, Coke Classic regular enjoys the largest share per market, followed by Pepsi regular, Coke Diet, and Pepsi Diet. It is clear that Coca Cola and Pepsi dominant the CSD market. Table 2 presents the summary statistics of the social media engagement across brands from different sources. It is clear that Coca Cola, Pepsi, and Dr. Pepper are the leaders in terms of the volume of consumer engagement on all three sites.

3 Model

Assume there are a total number of J carbonate soft drink (CSD) product on the market.Use $j = 1, ...J$ to denote a CSD product (e.g., Coca Cola Regular, Pepsi Diet, or Dr. Pepper Regular), and $j = 0$ to denote the general outside product in the beverage market. We define a market as a city-week combination in this analysis.

Following Berry, Levinsohn and Pakes (1995), we specify the conditional indirect utility of consumer i from purchasing CSD j or an outside product in market m as

$$\begin{aligned} u_{ijm} &= \alpha_i p_{jm} + \phi_{1i} TwitterDE_{jm} + \phi_{2i} TwitterIE_{jm} + \phi_{3i} GoogleE_{jm} \\ &\quad + \phi_{4i} FacebookE_{jm} + \ x_j \beta_i + \xi_{jm} + \epsilon_{ijm} \qquad (1) \\ &= \delta_{ijm} + \mu_{ijm} + \epsilon_{ijm} \qquad (2) \end{aligned}$$

where p_{ijm} is the unit price per oz of a soda drink product j in market m. x_j is a vector of observed nutritional characteristics (sugar, sodium, and caffeine content per oz) of soft drink brands. $Tweet_{jm}$ is the tweet goodwill which capture the effect of the total number of tweets mentioning brand j. $TwitterDE_{jm}$ is the engagement tweet goodwill which captures the effects of total number of tweets with "@ + brand j". $TwitterIE_{jm}$ is the promotion tweet goodwill which captures the effect of total number of tweets with "# + brand j". $GoogleE_{jm}$ is the google engagement goodwill which captures the volume of consumer online searches of brand j on google. $FacebookE_{jm}$ is the Facebook engagement goodwill which reflects the consumer engagement and responses on brand j's Facebook page. ξ_{jm} is the unobserved product characteristics.

Following Nerlove-Arrow(1962)'s exponential decay goodwill model, social media engagement for each brand is modeled as goodwill in order to capture the carry-over effects of the engagement's impact on demand. Specifically, product j's Tweet direct engagement goodwill stocks in period t is derived in a distributed lag form:

$$TwitterDE_{jt} = \theta TwitterDE_{j,t-1} + \sqrt{direct_{jt}} \qquad (3)$$

where θ is the carryover coefficients for brand tweets engagement and the square root captures diminishing effects (Erickson 1992). tw_{jt} represents the total number of engagement tweets with "@ + brand j" mentioning the CSD brand j at time t and t and k denote time periods.

In this analysis, Tweet direct engagement goodwill enters the utility functions directly.

Other social media engagement goodwill variables are modeled in a similar way.

$$
\begin{aligned}
TwitterIE_{jt} &= \theta TwitterIE_{j,t-1} + \sqrt{indirect_{jt}} && (4)\\
GoogleE_{jt} &= \theta GoogleE_{j,t-1} + \sqrt{google_{jt}} && (5)\\
FacebookE_{jt} &= \theta FacebookE_{j,t-1} + \sqrt{facebook_{jt}} && (6)
\end{aligned}
$$

where $indirect_{j,t-k}$, $google_{jt}$, and $facebook_{jt}$ are the total number of promotion tweets with "#+brand j" at time t, the total number of google searches of brand j at time t, and the total number of Facebook responses on brand j's Facebook page at time t, respectively.

To capture the heterogeneity of consumer preferences, we model the distribution of consumers' taste parameters, $\theta_i = (\alpha_i, \beta_i, \phi_{1i}, \phi_{2i}, \phi_{3i}, \phi_{4i})$, as multivariate normal distributions.

$$\theta_i = \theta + \Sigma\nu_i \quad (7)$$

where Σ is a scaling matrix and ν_i is the unobserved household characteristics, which is assumed to have a standard multivariate normal distribution. Let

$$
\begin{aligned}
\delta_{jm} &= \alpha p_{jm} + \phi_{1i} TwitterDE_{jm} + \phi_{2i} TwitterIE_{jm} + \phi_{3i} GoogleE_{jm}\\
&\quad + \phi_{4i} FacebookE_{jm} + x_j\beta_i + \xi_{jm} && (8)\\
\mu_{ijm} &= (p_{jm}, TwitterDE_{jm}, TwitterIE_{jm}, GoogleE_{jm}, FacebookE_{jm}, x_j)' * (\Sigma\nu_i), && (9)
\end{aligned}
$$

then the indirect utility U_{ijm} can be decomposed into three parts: a mean utility term δ_{jm}, which is common to all consumers; a brand-specific and consumer-specific deviation from that mean, μ_{ijm}, which includes interactions between consumer and product characteristics; and idiosyncratic tastes, where ϵ_{ijm} is a mean zero stochastic term distributed independently and identically as a type I extreme value distribution.

Assuming an i.i.d type I extreme value distribution of ϵ_{ijm}, we have a closed form solution of the probability a consumer i choosing soft drink j in market m:

$$Pr_{ijm} = \frac{\exp(\delta_{jm} + \mu_{ijm})}{1 + \sum_J exp(\delta_{jm} + \mu_{ijm})}$$

Aggregating over consumers, we can generate the market share of the brand in market m at time t:

$$s_{jm}(\theta) = \int I(\nu_i, \epsilon_{ijm}) : U_{ijm} \geq U_{ihm}, \forall h = 0, ..., JdG(\nu)dF(\epsilon) \tag{10}$$

where θ is a vector of consumer taste parameters as defined previously; $h = 0$ denoted the outside goods, and $G(\nu)\text{and}F(\epsilon)$ are the cumulative density functions for the indicated variables, which is assumed to be independent from each other. Matching the predicted market shares with data, We can solve for $(\alpha, \beta, \phi_1, \phi_2, \phi_3, \phi_4, \Sigma)$ using GMM.

In our model, we assume the product characteristics are exogenously determined, but the prices are correlated with unobserved product characteristics or demand shocks. To control for this endogeneity issue, we use several sets of exogenous instrumental variables following Nevo (2000). The first set of instruments is cost shifters, such as raw sugar price, aluminum price, TV advertising cost per second in each city/month etc. We also use Hausman (1996) type instruments, which are prices of the same brand in other markets. The intuition behind is that the prices of the same brand in different markets are correlated due to the common production cost, but uncorrelated with market specific demand shocks.

4 Results

We present the estimates of the demand coefficients in Table 3. Columns 2 and 3 give the means and standard errors of the parameters denoting the mean preference of consumers, or

$\alpha, \beta, \phi_1, \phi_2, \phi_3$and ϕ_4. Columns 4 and 5 provides the standard deviations of the random coefficients that capture the heterogeneity of consumer preferences. As expected, the estimate of the price coefficient is negative (-1.336) and strongly significant. Consumers' preferences of nutritional factors are also given. On average, consumers have a significantly positive valuation of sugar, and hence high calories, of carbonated soft drinks. Besides sugar, consumers also prefer caffeine intake but generally dislikes sodium intake.

The second panel of the table show the impact of all types of online engagement with brands on consumers' demand. Overall, all types of engagement on Twitter, Google, and Facebook have positive and significant impacts on demand, with varying magnitudes. Since the google trend search volume is normalized across all brands and time period, the coefficient of google trend search is not directly comparable with other three indexes. Among other three indexes, Twitter indirect engagement (# + brand name) has the highest impact, followed by Twitter direct engagement (@ + brand name) and Facebook responses.

5 Conclusion

In this paper, we formulate a random coefficient discrete choice model of consumer demand to study the impact of consumer attention and consumer engagement on social media on CSD demand, by combing monthly CSD sales data and social media data. Consumer attention and engagement on social media is modelled as goodwill in order to capture the carry-over effects of WOM's impact on demand, following Nerlove-Arrow (1962). Our results suggest that the three types of social media messages all have significant and positive effects on CSD demand. In particular, indirect engagement on Twitter has the largest impact on consumer demand, followed by direct engagement on Twitter and consumer engagement on Facebook.

	Mean Preference		Std. Deviation	
	Mean	Std.Err	Mean	Std.Err
Product Characteristics				
Price	-1.336	0.124	0.030	0.101
Sugar	0.165	0.022	0.000	0.195
Sodium	-0.110	0.024	0.104	0.028
Caffeine	0.104	0.025	0.025	0.068
Social Media Engagement				
Twitter Direct Engagement(@)	8.211	2.928	0.711	24.588
Twitter Indirect Engagement (#)	24.343	6.224	-7.348	7.889
Google Trend Search	0.041	0.013	0.006	0.038
Facebook Response	1.389	0.178	0.285	0.690
Others				
Constant	-3.221	0.340	-0.101	0.345
Coca Cola	0.715	0.124		
Pepsi	-0.273	0.080		
Month Dummies	Yes			

Table 3: Demand Estimates of Consumer Preferences in the CSD Market

References

Berry, S., J. Levinsohn, and A. Pakes, 1995. Automobile prices in market equilibrium. *Econometrica*, **63**, 841-89.

Chevalier, J., and D. Mayzlin. 2006. The Effect of Word of Mouth on Sales: Online Book Reviews. *Journal of Marketing Research*, **43(3)**, 345-354.

Chintagunta, P. K., S. Gopinath, and S. Venkataraman. 2010. The Effect of Online User Reviews on Movie Box Office Performance: Accounting for Sequential Rollout and Aggregation Across Local Markets. *Marketing Science*, **29(5)**, 944-957.

Godes, D., and D. Mayzlin. 2004. Using Online Conversations to Study Word-of-Mouth Communication. *Marketing Science*, **23(4)**, 545-560.

Nerlove, M. and K.L. Arrow. 1962. Optimal Advertising Policy under Dynamic Conditions. *Economica*, **29(114)**, 129-142.

Rui, H., Y. Liu, and A.B. Whinston. 2013. Whose and What Chatter Matters? The Impact of Tweets on Movie Sales. *Decision Support Systems*, **55(4)**, 863-870.