Econ 360 Paper

Daniel Atallah

December 9, 2018

Strength of Internet Presence on Company Performance

The internet and its ability to make networking between individuals has warranted it to be the most efficient data collection platform. Through social networks like Facebook, Instagram, LinkedIn, Twitter, Snapchat, etc., information on our daily and long-term interests is stored. Social platforms use a user's data, and their connections' data, to efficiently choose which advertisements to project to them. By being the best data collection platform, the internet has also quickly become the best marketing platform. As everyone becomes more internet reliant, companies are too, allotting a majority of their advertisement spending (adspend) to digital marketing campaigns. They do this because social networks charge companies a certain amount of money per ad view (the price per view is determined by a varying algorithm) giving businesses the most bang for their buck because, unlike television ads, these ads are targeted to an audience that is more likely interested in their products or services. However, online paid advertisements aren't the only method of digital marketing that companies are pursuing. Many companies are working on a more organic approach called search engine optimization (SEO). Companies with better SEO often appear higher on search engine searches relating to their affiliated webpages. Improving one's SEO is essentially free. Theoretically, companies with more, popular internet content have better SEO. That is why when I began my project I wanted to see how greatly a company's Twitter page improved its performance among consumers. I determined that to prove whether a correlation existed or not, I would regress company share price (dependent variable) on Twitter followers (explanatory variable). Share prices usually increase when a company undergoes improved performance while Twitter followers is a good measure of how many users are engaged to a given company's internet presence.

Upon beginning my research, I found how difficult and expensive it can be to collect Twitter data over time. I used twittercounter.com (a discontinued database as of November 5th, 2018), as my source for Twitter data. I could only get up to 14 days of data since I wasn't paying for a subscription. For this reason, after some discussion with Professor Van Kammen, I decided to make my research more general using Google Trends. With Google Trends I am able to find how often a given company is searched for on Google. While Twitter may have been ideal for researching the effects of social media, Google Trends gives a more holistic idea of a company's SEO. From there I compiled a list of 50 public consumer brands that have a verified account on Twitter (current Twitter follower count is one of my control variables). Using Google Trends I found the total number of times each company was searched for on Google for each of the 12 months of 2017; this makes up for 600 total observations (n = 600). I then used the Nasdaq website to find the monthly closing share price of each company. In other words, I looked for the last day of each month where the stock market was open and collected the closing share price of that day. The table below presents descriptive statistics of these variables.

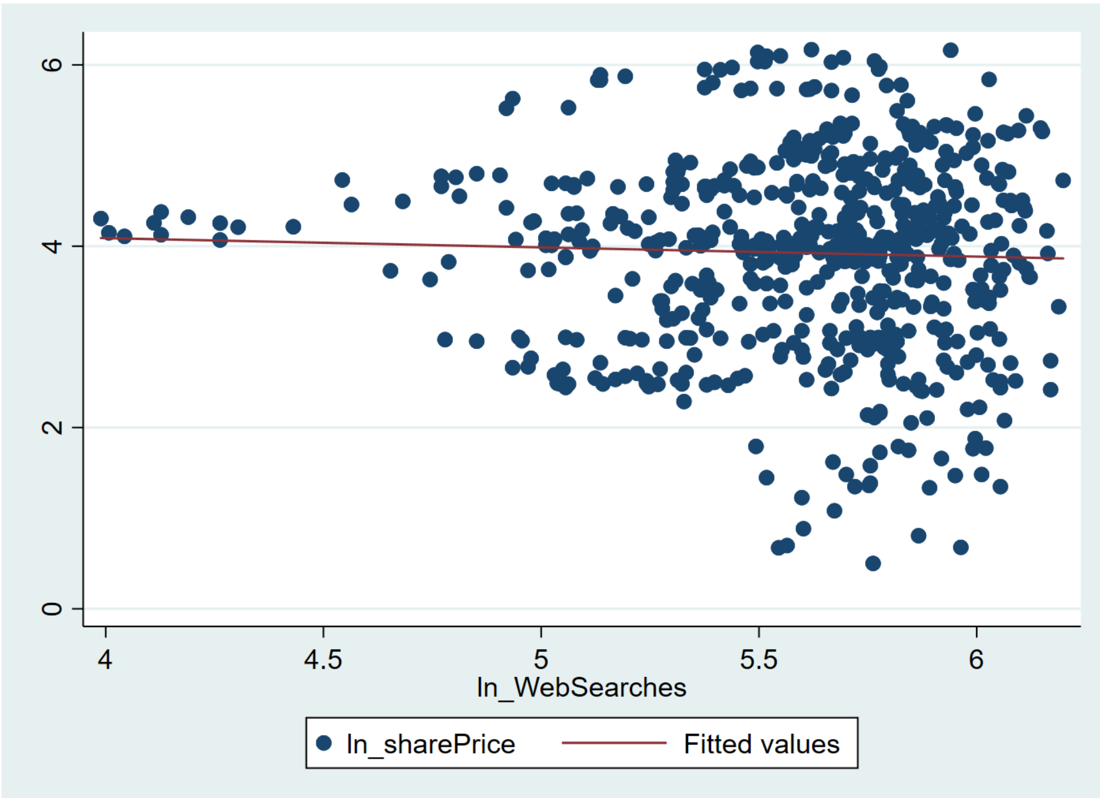
Table 1: Descriptive Statistics

Variable	Obs	Mean	Std. Dev.	Min	Max	Skewness
Share Price (USD)	600	82.94	87.04	1.65	477.35	2.21
Web Searches (count)	600	286.94	88.86	54.00	492.00	-0.22
Adspend (millions of USD)	600	2328.13	4985.24	1.60	24660.00	3.40
Bad Searches (count)	600	181.27	81.02	4.00	447.00	.39
Twitter Followers (millions)	600	2.06	2.54	0.04	11.50	2.07

Caption: Above are the descriptive statistics of my pre-logged variables.

The list of companies represented in the dataset varies greatly in size. Therefore, the values for each variable vary greatly between them. For example, McDonald's' share price is around $150 more than a single share of Wendy's. These sporadic and sometimes large differences are the reason all the variables are logged. This reduces the range between the maximum and minimum values in each category to some number under 6, where Ln(AdSpend) has the greatest range, compared to $22 billion (AdSpend range before the natural log is taken).

Ln(Share Price) is the dependent variable and Ln(Web Searches) is the explanatory variable. Also, Ln(Total Adspend), Ln(Bad Searches), and Ln(Twitter Followers) are all control variables. Similar to the method of finding values for the explanatory variable, "bad searches" was found using Google trends. However, instead of looking for the number of general searches, I specified that I only wanted the number of news searches. By including news searches in my data, I hope to incorporate the effects of any breakout news hits that either hinder or help the company. Similarly, I need to take into account the amount of money each company is spending on general marketing; this includes outlets such as television, radio, newspaper, etc. To do so I searched through the 2017 annual reports for shareholders for all 50 companies. Most companies have a statement on how much they spent on general advertising for the year. While some companies don't have this outright, they include these costs in the "Selling, General and Administrative Expenses" (SG&A) entry. In fact, advertising consumes most of the SG&A budget, making it a reliable number to use for general adspend. Since I did not have adspend divided by month, I attributed each company's annual total to each of the 12 months in 2017; in a sense, I collected 50 observations of adspend. Lastly, to stay true to my original project idea and represent social media engagement, or free marketing, as a factor, I included the current number of Twitter followers for each company's main Twitter accounts. I believe that if a company has a higher number of followers they are more likely to receive a higher amount of Twitter user engagement. If a Twitter account's followers are engaged, they can retweet, or share, content to their followers, easily increasing engagement through a digital means of word of mouth. In a similar scenario with the adspend variable, I collected 50 different observations of Twitter followers.



My initial simple regression follows the model: Ln(Share Price) = β_0 + β_1 * Ln(Web Searches) + µ. From Table 1, we see that all the variables are moderately to significantly skewed. While theoretically I can assume normality with a sample size of 600, with the scatterplot above, it is quite graphically noticeable how left skewed the explanatory variable is. Also, we see that for every change in Ln(Web Searches), Ln(Share Price) slightly decreases. In other words, β_1 is slightly negative. This result is rather surprising as it was expected to be a slightly large positive correlation.

The following table shows the numeric output of a simple OLS regression of Ln(Share Price) on Ln(Web Searches). As indicated by the scatterplot above, β_1 is slightly negative. Holding all else constant, a 1% increase in web searches warrants a 0.102% decrease in share price.

Table 2: Regression Estimates of Share Price Regression

	a	b	c	d
Coefficient Estimates by Google Search Count (s.e.) [Robust Errors]	-0.102 (0.117) [0.093]	-0.103 (0.116) [0.093]	-0.040 (0.143) [0.147]	0.059 (0.140) [0.137]
Controls	none	LN(total adspend)	LN(total adspend) LN(bad searches)	LN(total adspend) LN(bad searches) LN(Twitter followers)
Sample Size	600	600	600	600
Adjusted R^2	-0.0004	0.0083	0.0076	0.0596

- *Share price:* The closing prices of each month in 2017, measured in U.S. dollars.
- *Web searches:* The natural log of web searches, measured monthly, using google trends.
- *Total adspend:* The total amount of money spent on ads per month; this data was retrieved from each company's annual report for shareholders.
- *Bad searches:* The number of monthly news searches on a company; measured monthly using google trends.
- *Twitter followers:* Number of Twitter followers of each company's verified main twitter account.

Caption: Above is a table of β_1 for each model with an additional control variable.

Another surprising result is the little change in β_1 when Ln(Total Adspend) is included in the regression. The adjusted R^2 increases from -0.0004 to 0.0083 suggesting that with the addition of marketing, the regression line is now slightly positive. Companies are in fact seeing an increase in consumers as a result of increased adspend, however the relation is very weak and should be taken more so as correlation rather than causation. We see a larger change in β_1

when Ln(Bad Searches) is included in the model is it gets closer to 0. In a sense this is reasonable as most news reports are negative in nature. For example, when Chipotle was in the news in 2017 it was due to their e coli breakout. Therefore, we expect bad searches to have a negative effect on the share price, thus making sense out of web searches being less negative. However the standard error of β_1 gets larger while the adjusted R^2 decreases indicating that the bad search counts weakened our model. Surprisingly, this all improves when Ln(Twitter Followers) is added to the model. β_1 becomes positive and adjusted R^2 becomes a little under 10 times stronger.

Side note, upon writing this paper I learned that when a company name is searched for on Google, the first result is its main webpage and the second is a link to its twitter profile. This might be a good example of SEO having a great effect on a company's performance. It is possible that while someone can search for a company on Google, there is no guarantee that the search will lead anywhere past that first transaction, if that. Conversely, customers that follow a company's Twitter account are inclined to make multiple transactions. It might take those customers a single Google search to become a frequent customer.

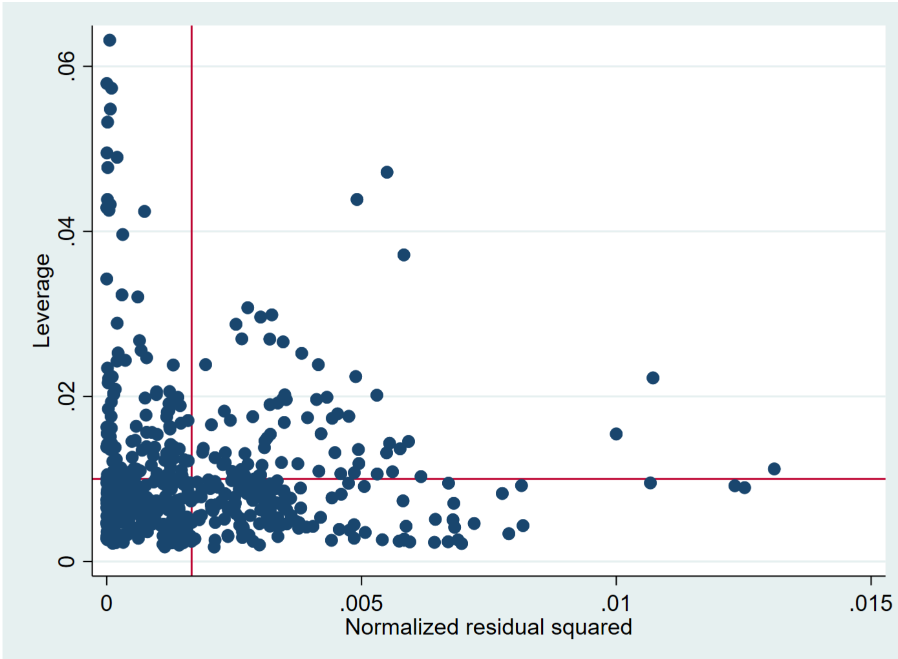
On the other hand, all marketing efforts aside, Twitter followers could just be a representative of how popular the company already is. To further verify whether or not Ln(Twitter Followers) is important, I created the following table with the marginal effects for Ln(Twitter Followers) on β_1 and Ln(AdSpend) on β_1. It seems that Ln(Twitter Followers) has a far stronger effect on Ln(Web Searches) as at each interval the coefficient increases by at least a factor of 0.4 whereas Adspend's by around 0.04 each time. This demonstrates that the effects of Twitter followers is 10 times the magnitude of AdSpend.

Marginal Effects for Ln(Twitter Followers)	Coefficient (S.E.)
At 11	-0.545 (0.355)
At 14	0.062 (0.140)
At 16	0.467 (0.261)

Marginal Effects for Ln(Total AdSpend)	Coefficient (S.E.)
At 14	-0.028 (0.454)
At 20	0.060 (0.140)
At 23	0.105 (0.265)

Caption: Left table is marginal effect of Ln(Twitter Followers); right is Ln(Total Adspend).

While this project has brought unique realizations, I can't confidently trust my model if my data is heteroscedastic. After running Breusch-Pagan, Cook-Weisberg, and White's tests for heteroskedasticity, I found that my model is highly heteroskedastic as my p-values are very small. By consequence, I included the robust errors on my second table on page 5, but it appears that this inclusion makes no improvement on the significance of my variables. Below is the leverage residual plot. It can be noted that there are a lot of outliers; this is probably the source for the high standard errors. After running a new regression that excludes the outliers, I am left with a significantly smaller sample size of n = 423, and significantly larger β_1 = 6.418, and a larger adjusted R^2 of approximately 0.1028. Lastly, when running the RESET test, I get a large value for p meaning I fail to reject the null hypothesis. In other words, I do not need to worry about omitted variables in my model.



In my search to find the effectiveness of businesses' evolving marketing efforts, I collected data from Google trends, Nasdaq, Twitter, and multiple company annual reports. I regressed a company's share price on the number of Google searches it receives. I found that with completely sporadic data, and a significantly heteroskedastic model, there is slight upward correlation between the two variables when Twitter follower count, company adspend, and news search counts are added as control variables to the model. I found β_1 to be smaller than expected, however the twitter follower data seemed to have a larger positive influence than expected. I guess this is a slight indication on how social media marketing is an increasingly used branding method. While I don't think I could have done anything better with my explanatory variable data, I want to try running this same regression with *good* Twitter data as my explanatory variable. I definitely feel that if years of twitter data were available to me I could determine the growing effects of social media on brand performance.

Works Cited

Google (2018). Trends. Retrieved November 21st, 2018, from https://trends.google.com/trends/

Nasdaq (2018). Company Share Price Data. Retrieved November 21st, 2018, from Nasdaq.com database (Contact me for a saved list of company stock names. Nasdaq also has annual reports under each company's profile.)

https://twitter.com/ (Contact me for a saved list of Brands' Twitter handles.)