Digital Analytics: Linking Online Data to Purchase Behavior and Firm-, Stock- and Political Performance

Evert de Haan
University of Groningen

Paper presented at the 48th Annual EMAC Conference, Hamburg, May 24-27, 2019.

EMAC
EUROPEAN MARKETING ACADEMY

Digital Analytics: Linking Online Data to Purchase Behavior and Firm-, Stock- and Political Performance

Session Chair:

Evert de Haan (Goethe University Frankfurt)

Included Papers:

1. **Vices Go Together: Product Purchase After Online Compulsion**

   Radek Karpienko (WU Vienna) - Presenter

   Anatoli Colicev (Bocconi University)

   Thomas Reutterer (WU Vienna)

   Koen Pauwels (Northeastern University)

2. **Search Engine Visibility and Shareholder Value: Financial Markets Valuation of a Firm's Visibility in Organic Search**

   Gabriela Alves Werb (Goethe University Frankfurt) - Presenter

3. **Fake News and Real Events: What Drove the 2016 US Election?**

   Koen Pauwels (Northeastern University) - Presenter

   Ginger L. Pennington (Northwestern University)

   Raoul Kubler (Munster)

4. **Satisfaction Surveys or Online Sentiment: Which One Best Predicts Firm Performance?**

   Evert de Haan (Goethe University Frankfurt) - Presenter

All presenters have agreed to register for the conference and to present the paper if the proposal is accepted.

None of the papers has been submitted to other conference tracks, and no paper has previously been presented at EMAC.

Session Abstract

In this session, four studies that focus on getting insights from digital data, from click-stream data to social media and online search data, are brought together. These data are used to explain online purchases, as well as and firm-, stock- and even political performance. The studies provide an overview of novel ways and techniques to get insights from these data in order to explain marketing phenomenon and relevant outcome variables for firms and society at large.

1. Vices Go Together: Product Purchase After Online Compulsion
Radek Karpienko (WU Vienna), Anatoli Colicev (Bocconi University), Thomas Reutterer (WU Vienna), Koen Pauwels (Northeastern University)

Digital addiction lacks clear connections to marketing theory and is supposed to have dramatic consequences on purchase behavior. Using clickstream data, we model purchase behavior after consumption of "vices" on a browsing session level. Our modeling results show that compulsive users generally exhibit a higher purchase probability than households with medium or low compulsion scores

2. Search Engine Visibility and Shareholder Value: Financial Markets Valuation of a Firm's Visibility in Organic Search
Gabriela Alves Werb (Goethe University Frankfurt)

This study provides a theory-based conceptual framework to explain the link between a firm's visibility in organic search and shareholder value. Using a panel of publicly listed U.S. firms and organic search visibility, the author finds that a 1% improvement in a firm's search visibility translates into a .01%-point increase in shareholder value.

3. Fake News and Real Events: What Drove the 2016 US Election?
Koen Pauwels (Northeastern University), Ginger L. Pennington (Northwestern University), Raoul Kubler (University of Münster)

We study the 2016 US election by investing potential driving factors by linking social media data, news articles, sentiment-, topic- and linguistic analyses, and major events to probability poll- and consensus polls data. We find that hate news, not fake news, drove Clinton's polls and the effects of the candidates language on social media has often opposite effects on the probability poll versus the consensus polls.

4. Satisfaction Surveys or Online Sentiment: Which One Best Predicts Firm Performance?
Evert de Haan (Goethe University Frankfurt)

This study investigates if sentiment of online word-of-mouth is a good alternative for survey based satisfaction data in predicting firm performance. Although the initial results indicate that traditional survey data yield better predictions, online sentiment does add incremental value. Combining survey data with sentiment data can improve predictions and help create a better dashboard of the customer base.

Included Papers

1. Vices Go Together: Product Purchase After Online Compulsion

Radek Karpienko (WU Vienna), Anatoli Colicev (Bocconi University), Thomas Reutterer (WU Vienna), Koen Pauwels (Northeastern University)

Compulsion is an important research topic in human behavior and temptation is rifle in today's online media environment. Several vices vie for online consumer attention, from social media to gambling and online games to video watching or adult content consumption. Many of these increasingly popular online behaviors have recently been recognized as addictions, implying patterns of repeatedly reinforcing the reward, motivation and memory circuitry (Olsen 2011, Leeman & Potenza 2013). While gaming addiction seems to be the most widely studied specific form of Internet addiction to date (Kuss & Grifftiths 2012, p. 348), less is known about online compulsion in other areas.

Digital addiction is a challenging research area for a number of reasons. First, recent studies on the neurobiology of addiction show that addicted and non-addicted consumers fundamentally differ in their physical response to addictive stimuli (Volkow et al. 2011). Second, addiction is hard to pinpoint outside of clinical studies, because addicted consumers are typically confronted with social desirability, distress, and shame (Ward & Mann 2000, Grubbs et al. 2015). Lastly, digital addiction lacks clear connections to marketing theory and marketing researchers have only recently started to study patterns of compulsion in rather specific areas such as binge consumption of online television (Schweidel & Moe 2016). Key unanswered questions for marketers are how they should react to observed consumer compulsion. Our research aims at shedding more light into these still open questions.

In the present study, we draw from the extant literature on self-control and hedonic consumption to explain how addictive stimuli affect consumers' purchase decisions (Hofmann et al. 2012). We propose that the actual spill-over of hedonic behavior to subsequent decisions is moderated by factors closely related to digital addiction: Shame and the psychological distress that consumers experience when they give in to "vices" – i.e., when they engage in activities that make them feel guilty. We study the proposed effects in the context of online pornography, an empirical setting that scores arguably high on this dimension, and thus represents a perfect test bed for our research questions.

We pursue a dual strategy to empirically test our research hypotheses by examining both (1) clickstream data obtained from households participating in the ComScore Web Behavior Panel and (2) online choice experiments. Using clickstream data, we model purchase behavior after consumption of vices on a browsing session level by relating product choice and purchase quantity to a number of explanatory and control variables in a Tobit type 2 setting. As a key metric of interest, we propose a compulsiveness score for online pornography and other video consumption. This metric is based on the deviations from evenly distributed inter-visit times, where clumpy patterns (heavy use followed by long breaks) indicate higher distress and addictive behavior.

Our results demonstrate that online pornography is indeed consumed in ways that point to particularly high distress, even compared with other vices (e.g., video streaming, social media, shopping). Furthermore, our modeling results show that compulsive users generally exhibit a higher purchase incidence than households with medium or low compulsiveness scores. Finally, we find the expected interaction effect between the consumption of vices in a browsing session and compulsiveness. Less compulsive users seem more sensitive to vices and are more likely to engage in impulsive buying when a vice is observed. We replicate this interaction in a series of online experiments in which we control for self-reported measures of addiction (Grubbs et al. 2015), and show that consumers' reactions to vices are moderated by how they perceive vices.

We conclude that it's not the vice, it's how consumers react to it. We demonstrate that individual reactions to the very same vice can vastly differ across individuals. While the consumption of regular online videos (e.g. on "YouTube" or "vimeo") is hardly associated with a social stigma, addicted individuals can react with high distress that affects purchase decisions and other unrelated activities. Furthermore, our research is the first study that systematically investigates the effects of online pornography consumption in marketing – an area that is extremely predominant in every-day behavior of consumers, and entirely neglected by marketing scholars and the social sciences in general. We discuss opportunities for further research on online vices and digital addiction, and managerial implications for targeting and consumer protection.

2. Search Engine Visibility and Shareholder Value: Financial Markets Valuation of a Firm's Visibility in Organic Search

Gabriela Alves Werb (Goethe University Frankfurt)

With consumers increasingly relying on search engines during their decision journey, organic search represents the most important source of online traffic for firms in several industries. Prior literature on search engines has addressed the effect of a firm's position in organic search results on click-through rates and conversion rates (Yang & Ghose, 2010), as well as how a firm's visibility in search engines affect its online visibility (Drèze & Zufryden, 2004) and its short-term performance (e.g. Alves Werb, Doppler, & Skiera, 2018).

However, as a firm's visibility in organic search may promote shareholder value not only by enhancing current cash flows but also by reducing the risk of and accelerating future cash flows, its long-term value is less likely to be evident according to accounting measures, which are backward looking in nature (Rust, Ambler, Carpenter, Kumar, & Srivastava, 2004; Srinivasan, Pauwels, Silva-Risso, & Hanssens, 2009). Against this background, this study: (1) develops a theoretical framework to understand the effect of a firm's visibility in organic search on shareholder value, (2) empirically measures the hypothesized effect by analyzing 7 years of weekly data on 50 U.S. publicly listed firms and (3) investigates the moderating role of firm and market characteristics.

The base of the analysis is a unique dataset containing 7 years (2011-2017) of weekly measurements of Sistrix's organic search Visibility Index for 50 U.S. publicly listed firms. This data was matched with stock market and accounting data obtained from the Center for Research in Security Prices and COMPUSTAT, which provide additional independent variables and

shareholder value measures – Tobin's q (Tobin, 1969) and firm-specific excess stock returns derived from the Fama and French (1993) three-factor model. The results indicate that a firm's visibility in organic search positively affects shareholder value ($p < 0.001$), after including finance fundamentals and controlling for firm-specific risk factors, market-wide and industry-specific shocks. In average, a 1% improvement in a firm's visibility in organic search results is associated with a 0.01 percentage point increase in shareholder value. For a typical firm in the analyzed sample, a 1% improvement in visibility in organic search translates into approximately $5.7 million more returns for shareholders in the long-term.

For marketing managers, these findings present hard evidence that a firm's visibility in organic search is a valuable intangible asset and can help to justify investments in Search Engine Optimization (SEO). For investors, the findings suggest that all else being equal, they should rate downward the stocks of firms that are affected by visibility losses in organic search.

3. Fake News and Real Events: What Drove the 2016 US Election?

Koen Pauwels (Northeastern University), Ginger L. Pennington (Northwestern University), Raoul Kubler (Munster)

The 2016 US presidential election outcome came as a surprise to many, and much has been suggested about the reasons – from the nature of press coverage to the influence of fake news and the messaging of the candidates themselves. In this paper, we code a wide range of explanations and analyze to what extent each drove the candidate's daily fortunes in the 4 months from their party nomination until the presidential election. Our focus lens is rich information from social media, which became the main marketing weapon in the Trump campaign due to its frequency of presentation, immediacy, simplicity and its ability to facilitate feelings of social cohesion among like-minded individuals.

For our main performance variable, we use the USC/LA Times Dornsife *probability* poll, which continuously surveyed people about how likely they were to vote and for which candidate. This poll was administered online, with internet access provided to participants who would otherwise be unable to access it. This poll was able to predict the election outcome with far greater accuracy than traditional polls. We contrast the effect of drivers on this poll with their effect on the '*consensus poll*', composed of the standard weighting of the traditional polls.

Our data on the potential drivers of voter opinion is drawn from a variety of sources: activity on candidates' Facebook pages, news coverage of events, campaign donations, candidate advertising spending, and candidates' own statements. We derive from this information the potential driving variables, which come in 4 groups:

(1) Candidates' Facebook page likes, shares, user comments and sentiment analysis;
(2) Coding of news articles shared on these pages (e.g., identification of fake / biased news)
(3) Sentiment, topic and linguistic analysis of candidates' statements;
(4) Occurrence of major events such as conventions, debates, FBI statements, and leaks

First, we scraped both candidates' Facebook pages for 3 years leading up to the election, collecting daily likes, shares, campaign statements (the official posts), and all the comments and

shared news by Facebook users. We used a Support Vector Machine trained with specifically context-dependent human coded social media data (40,000 positive and 40,000 negative observations) to derive the daily sentiment (number of positive and number of negative comments). A hold out validation with a classification accuracy of over 85% indicates that the SVM is performing well.

Second, we coded all the news articles shared by users on this Facebook Pages. *Fake news*, i.e. false stories that appear to be news, usually created to influence political views spreads faster, further and broader than real news because of its *novelty* (Aral et al. 2017). The human impulse is to click on and share novelty. This impulse is even stronger for political news, as compared to urban legends, disasters, etc. In our analysis, we distinguish fake news from *biased news*, which is factually correct, but strongly opinionated in favor or against a candidate. We further distinguish clickbait, conspiracy theories, hate, rumors, junk science, satire, and clickbait sites following Zimdar (2016). Our operationalization procedure proceeded in 2 steps: (1) we flagged the *sources* of potentially fake and biased stories based on widely available lists of biased websites and news sources (e.g. Breitbart, CNN, Fox News, NYTimes, OccupyDemocrats), (2) we recruited M-Turkers as blind coders of the specific news stories, achieving a high intersubjective agreement.

Third, with the help of Facebook's API, we coded each candidate's Facebook posts by *elements* (e.g. photo, video). Further, we used R's text mining package to code *use of language* (e.g. I, you, authenticity, gender), *emotions* (e.g. anger, sadness), and *temporal focus* (past, present, future). Last, we used Train's moral emotions dictionary to measure *morality* (e.g. fighting for or against authority: see Haidt and Graham 2001).

Fourth, we included specific events that received extensive coverage in the press and/or on social media, such as debates and conventions, the leaking of specific candidate statements (e.g. Clinton's 'Deplorables' and Trump's 'Grab that Pussy' public comments), Wiki Leaks of documents, FBI director Comey's statements and New York Times stories on 'No Evidence of Russia involvement', Clinton's emails and major endorsements of each candidate. Our analysis proceeds in two steps: (1) Granger causality tests shows *which* of the many potential drivers temporally cause the election polls and *when* each driver did so and (2) Dynamic regression modeling shows *how much* each driver affected the election polls.

Our results reveal several novel insights, including: (1) Hate news, not fake news, negatively drove Clinton's polls, (2) Trump and Clinton's language, expressed emotions and topics had different, often opposite effects on the probability poll versus the consensus polls, and (3) while several events had the expected effect on polls, Clinton *lost* and Trump gained from the Trump 'Grab that Pussy' recordings' release, and from biased news in general.

4. Satisfaction Surveys or Online Sentiment: Which One Best Predicts Firm Performance?

Evert de Haan (Goethe University Frankfurt)

Based on studies conducted in the last few decades we know that customer satisfaction is an important driver and predictor of future customer behavior and firm performance (e.g. Gupta and Zeithaml 2006). It has been shown for instance that customer satisfaction is positively related to a firm's financial performance (Morgan and Rego 2006), its stock return (Fornell et al. 2006), as well as customer retention (De Haan, Verhoef and Wiesel 2015). Many studies which have

looked into customer satisfaction have used data from the American Customer Satisfaction Index (ACSI), which annually measures customer satisfaction across many industries.

A more recent development is to use electronic word-of-mouth (eWOM) to measure the sentiment towards a firm and use this to predict future firm performance. Srinivasan, Rutz, and Pauwels (2016) find for instance that there is a positive relationship between number of likes and sales and Tirunillai and Tellis (2012) find that the amount of online user generated content is positively related to abnormal stock return.

These two streams of literature have up to now been studied isolated from each other. Furthermore, for eWOM most of the studies only focus on one or a small number of firms in a relative short time horizon. For both customer satisfaction and eWOM we do however know that they are an important predictor and driver of firm performance, as highlighted by the empirical generalizations summarized by Hanssens (2015), the strong relation of customer satisfaction and firm performance (e.g. Morgan and Rego 2005), and the high elasticities of eWOM volume and -valance found in the meta-analysis by You, Vadakkepatt and Joshi (2015).

Both data sources have their strengths and weaknesses; in our study we want to compare these two data sources to see which one is strongest related to firm performance. For our empirical study we have collected financial performance and ACSI data for 49 firms from 2010 and 2017. We have combined this with eWOM by scrapping all Tweets send to the official Twitter account of the firm, which resulted in 8,579,904 scrapped Tweets. For each Tweet we calculated the sentiment using the R-package 'qdap', which is used to calculate the overall online sentiment per firm per year, which is matched with the respective financial performance and ACSI data.

Per metric ($Metric_{i,t}$, i.e. ACSI, online sentiment of eWOM, amount of eWOM) we investigate how it predicts various firm performance metrics ($Y_{i,t}$, i.e. revenue, gross margin, cash flow, market share, market value, stock trading volume, and stock return) using equation (1).

$$Y_{i,t} = \beta_0 + \beta_1 \cdot Metric_{i,t-1} + \beta_2 \cdot Y_{i,t-1} + \varepsilon_{i,t} \tag{1}$$

As the table below shows, we find that the ACSI is in most cases significant, in line with Morgan and Rego (2006), while for eWOM (volume and valance) this is in fewer situations the case. In additional analyses, where we also look at relative values of the metrics and the financial performance (i.e. relative to other firms within the industry and relative in terms of growth) the eWOM data are contributing relative better in their predictions. Furthermore, combining ACSI with eWOM information provides significant improvements over just using one of the metrics. For firms, using both eWOM and survey based metrics to monitor the customer base is thus recommended.

	Revenue	Margin	Cash Flow	Market share	Market value	Trading volume	Stock return
ACSI	.002**	.170**	-233.206**	.001[n.s.]	-.050***	.004*	-.007*
Sentiment	.014[n.s.]	1.435[n.s.]	-884.986[n.s.]	-.014[n.s.]	-1.456***	.094**	-.052[n.s.]
Tweets #	.005[n.s.]	.314[n.s.]	-638.431*	.007***	.062[n.s.]	-.005[n.s.]	.013[n.s.]

*** = p. < .01, ** = p. < .05, * = p. < .10, [n.s.] = p. > .10

References

Alves Werb, G., C. Doppler, & B. Skiera (2018). *Visibility-at-Risk: An Approach to Measure a Firm's Risk of Losing Visibility in Organic Search*. Working paper.

De Haan, E., T. Wiesel, & K. Pauwels (2016). The Effectiveness of Different Forms of Online Advertising for Purchase Conversion in a Multiple-Channel Attribution Framework. *International Journal of Research in Marketing*, 33 (3), 491-507.

Drèze, X., & F. Zufryden (2004). Measurement of Online Visibility and its Impact on Internet Traffic. *Journal of Interactive Marketing,* 18 (1), 20-37.

Fama, E. F., & French, K. R. (1993). Common Risk Factors in the Returns on Stocks and Bonds. *Journal of Financial Economics,* 33 (1), 3-56.

Fornell, C., S. Mithas, F. V. Morgeson III, & M. S. Krishnan (2006). Customer Satisfaction and Stock Prices: High Returns, Low Risk. *Journal of Marketing*, 70 (1), 3-14.

Gupta, S., & V. Zeithaml (2006). Customer Metrics and Their Impact on Financial Performance. *Marketing Science*, 25 (6), 718-739.

Grubbs, J. B., Volk, F., Exline, J. J., & Pargament, K. I. (2015). Internet pornography use: Perceived addiction, psychological distress, and the validation of a brief measure. *Journal of Sex & Marital Therapy*, *41*(1), 83-106.

Haidt, J., & J. Graham (2011). Mapping Moral Domain. *Journal of Personality and Social Psychology*, 101(2), 366–385.

Hanssens, D. M. (2015). Empirical Generalizations about Marketing Impact, 2nd Edition. Marketing Science Institute, Cambridge (MA).

Hofmann, W., Vohs, K. D., & Baumeister, R. F. (2012). What people desire, feel conflicted about, and try to resist in everyday life. *Psychological science*, *23*(6), 582-588.

Kuss, D. J., & M. D. Griffiths (2012). Internet and Gaming Addiction: A Systematic Literature Review of Neuroimaging Studies. *Brain Sciences*, 2, 347-374.

Leeman, R.F., & M. N. Potenza (2013). A Targeted Review of the Neurobiology and Genetics of Behavioural Addictions: An Emerging Area of Research. *Can. J. Psychiatry Rev. Can. Psychiatr.*, 58, 260–273.

Love, Todd, Christian Laier, Matthias Brand, Linda Hatch, & Raju Hajela (2015). Neuroscience of Internet Pornography Addiction: A Review and Update. *Behavioral sciences* 5, no. 3 (2015): 388-433.

Morgan, N. A., & L. L.Rego (2006). The Value of Different Customer Satisfaction and Loyalty Metrics in Predicting Business Performance. *Marketing Science*, 25 (5), 426-439.

Olsen, C.M (2011). Natural Rewards, Neuroplasticity, and Non-Drug Addictions. *Neuropharmacology*, 61, 1109–1122

Rust, R. T., T. Ambler, G. S. Carpenter, V. Kumar, & R. K. Srivastava (2004). Measuring Marketing Productivity: Current Knowledge and Future Directions. *Journal of Marketing,* 68 (4), 76-89.

Schweidel, D. A., & Moe, W. W. (2016). Binge watching and advertising. *Journal of Marketing*, *80*(5), 1-19.

Srinivasan, S., K. Pauwels, J. Silva-Risso, & D. M. Hanssens (2009). Product Innovations, Advertising, and Stock Returns. *Journal of Marketing,* 73 (1), 24-43.

Srinivasan, S., O. J. Rutz, & Koen Pauwels (2016). Paths To and Off Purchase: Quantifying the Impact of Traditional Marketing and Online Consumer Activity. *Journal of the Academy of Marketing Science*, 44 (4), 440-453.

Tirunillai, S. & G. J. Tellis (2012). Does Chatter Really Matter? Dynamics of User-Generated Content and Stock Performance. *Marketing Science*, 31 (2), 198-215.

Tobin, J. (1969). A General Equilibrium Approach To Monetary Theory. *Journal of Money, Credit and Banking,* 1 (1), 15-29.

Volkow, N. D., Wang, G. J., Fowler, J. S., Tomasi, D., & Telang, F. (2011). Addiction: beyond dopamine reward circuitry. *Proceedings of the National Academy of Sciences*, *108*(37), 15037-15042.

Vohs, K. D., & Faber, R. J. (2007). Spent resources: Self-regulatory resource availability affects impulse buying. *Journal of consumer research*, *33*(4), 537-547.

Ward, A., & Mann, T. (2000). Don't mind if I do: Disinhibited eating under cognitive load. *Journal of personality and social psychology*, *78*(4), 753.

Yang, S., & A. Ghose (2010). Analyzing the Relationship Between Organic and Sponsored Search Advertising: Positive, Negative, or Zero Interdependence? *Marketing Science,* 29 (4), 602-623.

You, Ya, Gautham G. Vadakkepatt, & A. M. Joshi (2015). A Meta-Analysis of Electronic Word-of-Mouth Elasticity. *Journal of Marketing*, 79 (2), 19-39.