Using Predictive Analytics to Reduce Uncertainty in Enterprise Risk Management

Research-in-Progress

Hossein Ghasemkhani
Krannert School of Management,
Purdue University
hossein@purdue.edu

Shachar Reichman
Recanati Business School,
Tel Aviv University and
Sloan School of Management,
Massachusetts Institute of Technology
shachar@mit.edu

George Westerman
Sloan School of Management,
Massachusetts Institute of Technology
georgew@mit.edu

Abstract

Traditional economic and business forecasting about corporate credit has relied on statistics from government agencies, annual reports and financial statements. These statistics are often published with significant delay, which limits their usefulness for predicting changes in creditworthiness. Yet, a delay in responding to changes in a company's credit rating can have significant financial and risk consequences. With the widespread adoption of search engines, social media and related information technologies, it is possible to obtain data on literally trillions of economic decisions almost the instant that they are made. In this study, we investigated the power of these online activity data, combined with data on firms' business ecosystems, to predict the likelihood of counterparty credit downgrade risk. The research offers a novel approach that contributes to the fields of information systems, finance, and social science by providing new insights on the role of these data types on firms' financial risk.

Keywords: Predictive analytics, risk management, financial prediction

Introduction

On October 6th, 2014, GT Advanced Technologies (stock symbol GTAT) filed for bankruptcy in a surprise announcement. The default affected many companies, notably Apple, which only a year earlier had chosen GTAT to be the exclusive supplier of sapphire components for its new product lines. The bankruptcy threatened GTAT's ability to supply components to Apple, thus jeopardizing Apple's ability to meet consumer demand for its products. The GTAT example emphasizes the importance of understanding potential enterprise risk in the business world, especially the link between credit risk and operational risk. It also highlights the interconnectedness of enterprise risk; a firm's performance is dependent on the risk and performance of its business ecosystem partners.

Traditional forecasting of credit risk has relied on statistics from government agencies, annual reports and financial statements. These statistics are often published with significant delay, which limits their usefulness for prediction. In addition, credit risk is a function of internal factors that are not publicly available. The only external signals of internal operations risk are reports produced by ratings agencies, but these are subject to delay and missing information. In fact, one of the common criticisms of credit rating agencies is that they are too slow in updating their rating (Alp 2013).

Now, due to the widespread adoption of search engines, social media and related information technologies, it is possible to obtain data on literally trillions of economic decisions almost the instant that they are made. Furthermore, query technology has made it possible to obtain such information at nearly zero cost, virtually instantaneously and at a fine-grained level of disaggregation. Each time a consumer or business decides to search for a product via the Internet, valuable information is revealed about that individual's intentions to make an economic transaction (Moe & Fader, 2004). In turn, knowledge of these intentions can be used to predict supply and demand (e.g. Choi & Varian, 2012; Goel et al. 2010).

This social information revolution is well underway. It portends a revolution in our ability to make business predictions and ultimately a sea change in business decision making. The change is not a mere difference in degree, but a fundamental transformation of what is known about the present and what can be predicted about the future. The collective intelligence from different social media and ecommerce channels can be used to predict real-world outcomes. For example, Asur and Huberman (2010) have shown that Twitter can be used to predict box-office revenues better than market-based predictions.

In this research, we focus on the important area of predicting enterprise risk. In particular, we investigate the power of search volume data and Wikipedia visit data, combined with data on companies' networks of suppliers and buyers, to predict the likelihood of counterparty credit risk. For the purpose of this study, we focus on the risk that a firm's credit rating is downgraded by a major ratings agency. Downgrades do more than signal risk. They also create risk by restricting a company's access to capital and potentially reducing customer willingness to purchase from the company. Yet ratings announcements are subject to delays and information leakage. Companies that can assess a partner's risk and predict rating changes can take action to reduce risk before a public announcement happens.

The research extends credit prediction models beyond financial indicators to include social indicators of economic decision-making. It also examines the effect of a firm's business ecosystem on its credit risk. Prior studies (e.g, Cohen and Frazzini 2008; Menzly and Ozbas 2010) have shown that supply-chain relationships may influence cross-predictability. Surprisingly, this topic has received scant attention in the extant literature of credit rating. To the best of our knowledge, this research is the first to study the effects of the online activity data of a company's business peers in predicting its default and credit risk.

The combination of the different parts of the research offers a novel approach that will contribute to the fields of information systems, economics, and social science by providing new insights on the role of online data and the business ecosystem's online signals on firms' financial risk. The research also offers a practical contribution by providing new methods to evaluate near-real-time financial risk.

Related Literature

The credit rating of firms has a significant role in the financial systems. This rating measures the risk of the firm and the probability of default or delayed payment of bonds and other debt instruments (Bai, 2010). It is one of the key parameters in investment, and financial and operational decision making for lenders and potential customers (e.g. Nini et al. 2009; Bannier 2012). Because of the importance of credit rating downgrades on risk and future performance of companies, estimating and predicting changes in credit rating has received substantial attention in the literature. Most of the studies focus on reported financial measures of the focal firm as they can provide valuable information on the future financial stability of the firm (Aziz and Dar 2006 provide a thorough review of previous studies). Modeling approaches include statistical estimation, AI methods and machine learning techniques (see Huang et al. 2004, and Aziz and Dar 2006, for a thorough review of previous studies and comparison of the predictive models). These studies have examined the importance of particular financial indicators, or sets of indicators, on the estimation or the accuracy improvement of different prediction methods. However, these studies have not included nonfinancial information such as social and online activity and have also overlooked the influence of ecosystem partners on the credit rating of the firm.

Because internal business conditions are important signals or determinants of risk, companies often must depend on intermediaries such as credit rating agencies to assess the risk of other companies. Traditionally, by generating credit scores, rating agencies took the important role of reducing information asymmetries between firms and investors (Bai 2010). With the availability of online data sources, investors today have more options to fill the information asymmetry gap by analyzing online activity data.

Researchers and businesses increasingly use online and social data to explain and predict various economic outcomes. Online reviews, discussion forums and blogs have been utilized to measure and predict product demand and sales (e.g. Godes and Mayzlin 2004; Chevalier and Mayzlin 2006; Liu 2006; Rui et al. 2013). A parallel and very related stream of research used online search engine logs or search trends to provide useful predictions in a wide range of domains including epidemic outbreaks (Ginsberg et al., 2009), movie box office sales and music billboard rankings (Goel et al., 2010), automotive sales (Choi and Varian, 2012; Du and Kamakura, 2012; Geva et al., 2013), home sales (Choi and Varian, 2012; Wu and Brynjolfsson, 2009), unemployment claims (Choi and Varian, 2012), and private consumption (Vosen and Schimdt, 2011). However, this use of big data for predictions could still benefit from the use of other more traditional data sources (Lazer et al. 2014).

In addition, previous research on credit rating estimation and predictions analyzed each firm independently of others (Aziz and Dar 2006). However, firms do not operate separately; they maintain close relationships with suppliers and customers. Risk and performance of a firm's supply chain network or business ecosystem can proliferate to the focal firm, making the interconnected business ecosystem an important element for evaluation (Anggraeni et al. 2007, Iansiti and Levien 2004). For example, Wu and Birge (2014) found that supplier momentum can be used to predict a firm's returns and that a firm's centrality in the network influences its returns. Leung et al. (2013) showed that analyzing the online activity of firms' suppliers and customers can improve prediction of stock performance and could assist in building a trading strategy.

In this study, we extend prior streams of research by combining financial data with online social data, not only for focal firms, but also for their partners. This novel approach enables the creation of models that capture the whole ecosystem of a firm and evaluate the predictive effect of each category of information.

Conceptual model

Conceptually, credit risk is a function of operational performance and financial performance, each of which are functions of strategy, underlying capability, and access to resources. In addition to steady state performance information, risk can also result from shocks to the corporate system such as executive departures, loss or default of a key supplier or customer, or labor unrest. This information typically does not appear on the firm's financial reports until long after it happens. Public information that is much higher velocity than financial or regulation-mandated reports could be useful in predicting credit downgrades, if it were available.

Traditional credit prediction models, which are based on low velocity financial information, have been shown to have strong predictive power for credit ratings (Aziz and Dar 2006). We extend these models in two ways (see Figure 1):

1. **Nowcasting** (Choi and Varian 2012): Online search activity may capture data about current events such as sales or operational disruptions as they happen, as well as sentiment trends regarding that firm. This data may include important indicators of the firm's current operational

or financial conditions that are not yet included in the financial report. The theory underlying this method is that online data reflect cumulative actions performed by a large number of people over time and, as a result, capture longitudinal changes in behavior. This social data may serve as a high velocity signal of information that will later be published in slower-velocity reports.

2. **Risk spillover**: Companies do not operate in an isolated environment. Any shock to a peer might influence the performance and risk of the focal firm. A significant change in the financial stability or operational reliability of a major partner may destabilize the focal firm and increase its credit risk. Adding partner network information to information on the focal firm may identify significant risk spillovers or broader trends affecting an entire industry including the focal firm.

Figure 1 shows the conceptual model driving our research project.

Focal Company	• Quarterly financial report • Credit rating	• Sales volume • Business interruption • Contract cancellation or executive departure • News report
Ecosystem (Risk spillover)	• Partner financials • Partner credit • Industry analyst reports • Ecosystem structure	• Partner sales volume • Partner business interruption • Change in partnership arrangement
	Low velocity information (Traditional reports)	**High velocity information (Nowcasting)**

Figure 1. Conceptual extensions to traditional credit predictions models

Data and Methods

Gathering data about online activity and ecosystem partners required combining information from numerous data sources. We started by collecting data on credit rating announcements of companies on the S&P 500 companies list from any of the three major rating agencies (Moody's, S&P, and Fitch) between 2004 and 2014. Data were collected from the Bloomberg Professional service which provides data on companies' credit ratings and the changes in ratings over time. Because rating actions tend to occur sequentially across rating firms, we noted the first rating action within a two weeks period, and ignored the ones that follow. Overall, we collected data on 441 companies with 4500 rating events, 920 of which are rating downgrades. For each company, we also gathered rating signals, which are indications that rating firms sometimes provide to indicate possible future direction of a rating.

For each of the companies in the dataset we collected three additional data types from different data sources:

1. *Financial data*: we extracted financial data as reported in companies' published financial reports from Compustat. These data include: total assets, cash, long term debt, total inventories, net income, property, plant and equipment total, total receivables, total revenues, market value, and stock price data during the reported financial quarter.
2. *Online Activity Data:*
   a. *Wikipedia data*: the daily number of page visits to these firms' Wikipedia pages.

b. *Google Trends*[1] *data*: Search volume data for each of the firms were collected from Google Trends for searches performed in the US. The data included searches of the firm's name and the combination of the firm's name and 17 financial keywords with negative sentiment. The keywords were chosen using two methods. First, we talked to five risk experts at a large financial services firm to come up with an initial list. Then we used the crowd-squared method (Brynjolfsson et al. 2014), in which a large number of individuals identify additional search terms that are relevant to a specific topic or event.

3. *SPLC data*: we collected data from the Bloomberg supply chain database (SPLC) on each firm's top suppliers and customers. Using the SPLC data we constructed the ecosystem network of the firms and their suppliers and customers, and computed three centrality indices that are commonly used in the literature to characterize network structures and effectiveness: *Betweenness Centrality* of a node, which measures the number of shortest paths between any two nodes in the network (Freeman, 1977); A *Clustering Coefficient* of a node, which quantifies how close a node's neighbors are to being a clique (Watts and Storgatz, 1998); and the PageRank of a node, which measures the node's relative importance within the set of nodes in the network (Brin and Page, 1998). Additionally, we calculated the average credit rating of the focal firm's suppliers and customers.

Table 1 shows the key variables and their descriptive statistics.

Table 1 – Descriptive Statistics of a Subset of Variables

Variable	Mean (Std.)	Min	Max
Downgrade (y)	0.29 (0.45)	0	1
Current Rating	9.40 (3.37)	1	24
Rating Signal (Outlook)	-0.13 (0.53)	-1	1
Wikipedia Avg Weekly Visits	620.86 (2117.67)	0	43902.86
Google Trends (Firm name)	33.31 (21.65)	0	100
Stock Price (Quarter High)	54.71 (60.82)	3.02	1378.96
Stock Price (Quarter Low)	42.99 (48.61)	0.21	1103.9
Long Term Debt ($Millions)	9501.13 (27761.52)	0	330067
Cash ($Millions)	2148.17 (5423.88)	0	56008
Net Income ($Millions)	294.04 (857.08)	-5357	13615
Total Assets ($Millions)	46650.38 (135600.2)	1165.739	1199993
Market Value ($Millions)	20515.68 (28154.99)	53.28	287690.8
SPLC PageRank	0.002 (.028)	0.000	0.372
SPLC Betweenness	13027.58 (26453.67)	0	145758.4
SPLC Clustering Coefficient	0.18 (0.06)	0	0.48
Avg SPLC Rating	8.56 (2.26)	3	23

To estimate the suggested methods, we assess the predictive performance of each data source using a discrete-time hazard model:

$$\ln\frac{p(y_{it})}{1-p(y_{it})} = \alpha_0 + \sum_{l=1}^{L}\gamma_l \mathrm{FinancialInfo}_{ilt} + \sum_{j=1}^{j}\alpha_j \mathrm{GoogleTrends}_{ijt} + \sum_{k=1}^{K}\beta_k \mathrm{WikipediaActivity}_{ikt} + \sum_{p=1}^{j}\delta_p \mathrm{SPLC}_{ipt}$$

Here y is a binary variable which is 1 if firm i is downgraded at time t and is 0 otherwise. FinancialInfo is the most updated financial data from the firms' reports; GoogleTrends data is the search volume related to the firm, WikipediaActivity is a measure of the number of visits to the focal firm's Wikipedia page, and SPLC is a measure of the financial indicators and online activity of the firm's ecosystem peers.

[1] Google Trends (http://www.google.com/trends/) is a publicly available product that aggregates billions of search queries and provides information about the relative volume of different search terms.

Analysis and Results

We conducted the analysis in two stages. For variable selection, we ran a random forest model with 1000 trees on the data. Random Forest is an ensemble learning method which combines random feature selection with bagging. It grows a number of trees using a random subset of the data and variables and takes the average of all trees while classifying observations. Next, using variables selected in the first stage, we conducted the second stage of analysis with a discrete-time hazard model.

Stage 1: Variable Selection

Figure 2 shows the ranking of variables suggested by our random forest model. It shows the two figures of merit, mean decrease in accuracy (reduction in accuracy of the model due to excluding the variable), and mean decrease in Gini coefficient (reduction in homogeneity of leaves and nodes) that we used to compare and rank variables. We used the output from the random forest procedure to come up with a list of variables that show up high on the lists based on both measures. As a second step, we also used a stepwise feature selection method which chose variables based on their contribution to the in-sample accuracy of the model. Finally, we chose a subset of variables to include based on repeatedly measuring the out-of-sample performance of different combinations of variables that ranked highly based on those two procedures.

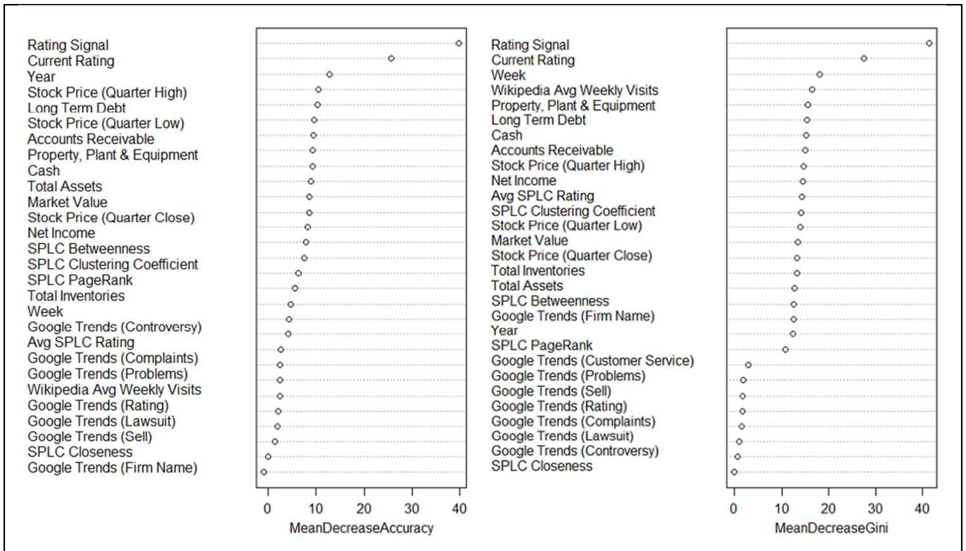

Figure 2. Predictive Power of the Variables According to Random Forest

Stage 2: Predictive Model

In the second stage, we used the variables suggested by our random forest model to estimate a discrete-time hazard model. The time metric in our analysis is any occasion in which a firm is rated, whether the rating is up, down, or stable. The event of interest happens when a firm is downgraded by one of the main three agencies. Therefore, the model estimates the probability of a downgrade conditional on the firm being rated by an agency. A discrete-time model is appropriate for this type of condition. Our model also accounts for censoring at the end of the data collection period. We chose to use the logit link for the hazard model due to its interpretive convenience (Hosmer & Lemeshow, 2000).

Table 2 shows the estimated coefficients of the discrete-time hazard model. The coefficients are in logit scale, meaning positive numbers are associated with an increase in the risk of downgrade, while negative numbers are associated with a decrease in the risk. Here we have included the top-20 variables suggested by the random forest. As 2008 and 2009 were clearly different in terms of the number of downgrades, and also to be able to include the data on Wikipedia visits, which is only available for 2008 and later, we focused our analysis on years 2010-2014.

We then evaluated the predictive power of our model using the area under the receiver operating characteristic (ROC) curve. Area under the ROC combines both sensitivity and specificity of the model, providing a single number to compare the predictive power of various models.

We randomly divided our sample into a calibration sample (70% of the data) and a testing sample (30% of the data). To eliminate the effect of an especially good or bad random sample, we used 500 different bootstrap samples of calibration and testing data. For each combination, we estimated the model using the calibration data and then tested the out-of-sample predictive power using the holdout sample. Our discrete-time hazard model provided an average area under the ROC curve of 0.77 (see Figure 3).

Table 2 – Estimation Results for the Hazard Model

	Variables	Estimated Coefficients (Standard Errors)
	Intercept	362.480** (108.410)
	Year	-0.180** (0.054)
Financial indicators	Rating Signal	-1.922** (0.190)
	Current Rating	-0.172** (0.031)
	Stock Price (Quarter High)	0.005 (0.008)
	Stock Price (Quarter Low)	-0.004 (0.010)
	Long Term Debt	-0.000 (0.000)
	Accounts Receivable	-0.000 (0.000)
	Property, Plant & Equipment	0.000~ (0.000)
	Cash	0.000 (0.000)
	Total Assets	0.000 (0.000)
	Market Value	-0.000* (0.000)
	Net Income	0.001* (0.000)
	Total Inventories	0.000 (0.000)
Online Social Activity	Google Trends (Complaints)	0.014* (0.007)
	Google Trends (Problems)	-0.007 (0.007)
	Wikipedia Avg Weekly Visits	-0.000~ (0.000)
SPLC	SPLC Betweenness	0.000 (0.000)
	SPLC Clustering Coefficient	-0.211 (1.615)
	SPLC PageRank	2.977 (3.287)
	Avg SPLC Rating	-0.022 (0.042)
	Observations	852
	LL	-394.5

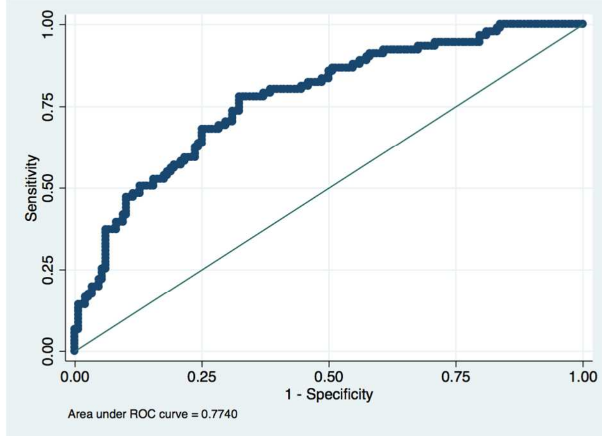

Figure 3. Receiver Operating Characteristic Curve

Concluding Remarks and Future Directions

Understanding the financial stability (or instability) of firms is critical for the business performance of companies (at the micro level) and for policy strategies of the whole market (at the macro level). While many credit risk models use publicly available financial information, new information and technologies provide new data sources with additional predictive power.

Online activity such as search trends and site visits can be considered high velocity proxies for current conditions that have not yet been published in lower-velocity reports. In addition, in an increasingly interconnected economy, this research can shed light on the usefulness of firms' ecosystem online activity data for predicting negative financial events (default or credit downgrade) of focal firms. It is not straightforward to evaluate the extent to which these observations may provide better indicators to predict firms' credit risk. However, in our initial predictive model, a combination of specific financial, social, and ecosystem measures shows promise for predicting credit downgrades, with an area under the ROC curve of 0.7740.

Returning to the paper's opening example, we assessed the risk of GTAT using the discrete-time hazard model and the results are remarkable. The model's estimate of GTAT's downgrade risk in the week before its default announcement was higher than 97% of the companies in the dataset, which means that GTAT was at very high risk compared to other companies in the dataset.

These results demonstrate the importance of the use of online data sources to reduce uncertainty in the business operation of the firm. They also show the potential of the use of non-traditional online data sources in improving predictions.

We plan to extend the research to include a complete picture of each company's ecosystem, and its role in the industry and general business network. We also plan to include additional variables from online sources including news reports and analysts' reports that typically contain additional insights on the firms' financial situation. Using content analysis techniques we will measure these data items and aggregate them to generate comprehensive information on the online appearance of the company. Finally, we will examine different prediction methods beyond the discrete time methods we used in the current model.

The overall vision for this project is to show the power of nontraditional online and ecosystem data to increase accuracy of risk prediction. Our model shows that online activity and ecosystem data can provide indicators of an impending risk downgrade before it is published by a credit agency. Knowing about a trading partner's credit downgrade in advance can allow a company to mitigate risk before the negative implications of a public announcement occur. This method could support decision makers in a wide range of business decisions from supply chain planning to business partnership decisions to investment strategies.

References

Alp, A. 2013. "Structural shifts in credit rating standards," *The Journal of Finance*, *68*(6), 2435-2470.

Anggraeni, E., Den Hartigh, E., and Zegveld, M. 2007. "Business ecosystem as a perspective for studying the relations between firms and their business networks," *ECCON 2007 Annual meeting*.

Asur, S., and Huberman, B. A. 2010. "Predicting the future with social media," *Web Intelligence and Intelligent Agent Technology (WI-IAT)*.

Bai, L. 2010. "On regulating conflict of interests in the credit rating industry," *New York University Journal of Legislation and Public Policy*,*13*, 10-17.

Bannier, C. E., Hirsch, C. W., and Wiemann, M. 2012. "Do credit ratings affect firm investments? The monitoring role of rating agencies," *The Monitoring Role of Rating Agencies (*http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2135483*)*.

Brin, S., and Page, L. 1998. "The anatomy of a large-scale hypertextual Web search engine," *Computer networks and ISDN systems* (30:1), pp 107-117.

Brynjolfsson, E., Geva T. and Reichman, S. 2014. "Using Crowd-Based Data Selection to Improve the Predictive Power of Search Trend Data," *The International Conference on Information Systems (ICIS 2014)*, Auckland, New Zealand.

Chevalier, J. and Mayzlin, D. 2006. "The Effect of Word of Mouth on Sales: Online Book Reviews," *Journal of Marketing Research* 43, pp. 345–354.

Choi, H. and Varian, H. 2012. "Predicting the Present with Google Trends," *Economic Record* (88:s1), pp. 2-9.

Cohen, L., and Frazzini, A. 2008. "Economic Links and Predictable Returns," *The Journal of Finance* (63:4), pp. 1977-2011

Du, R. Y., and Kamakura, W. A. 2012. "Quantitative Trendspotting," *Journal of Marketing Research* (49:4), pp. 514-36.

Freeman, L. C. 1977. "A set of measures of centrality based on betweenness," *Sociometry*), pp 35-41.

Geva, T., Oestreicher-Singer, G., Efron, N., and Shimshoni, Y. 2013. "Do Customers Speak Their Minds? Using Forums and Search for Predicting Sales," in *Proceedings of the 2013 International Conference on Information Systems*.

Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., and Brilliant, L. 2009. "Detecting Influenza Epidemics Using Search Engine Query Data," *Nature* (457:7232), pp. 1012-14.

Godes, D., and Mayzlin, D. 2004. "Using Online Conversations to Study Word-of-Mouth Communication," *Marketing Science* (23:4), pp. 545-60.

Goel, S., Hofman, J. M., Lahaie, S., Pennock, D. M., and Watts, D. J. 2010. "Predicting Consumer Behavior with Web Search," *Proceedings of the National Academy of Sciences* (107:41), pp. 17486-90.

Hosmer, D. W., & Lemeshow, S. (2000). *Applied logistic regression*. (2 ed.). New York: Wiley-Interscience Publication.

Howe, J. 2006. "The Rise of Crowdsourcing," *Wired Magazine* (14:6), pp. 1-4.

Huang, Z., Chen, H., Hsu, C. J., Chen, W. H., and Wu, S. 2004. "Credit rating analysis with support vector machines and neural networks: a market comparative study," *Decision support systems*,*37*(4), 543-558.

Iansiti, M., and Levien, R. 2004. "Strategy as ecology," *Harvard business review*,*82*(3), 68-81.

Lazer , D., Kennedy, R., King, G., and Vespignani, A. 2014. "The Parable of Google Flu: Traps in Big Data Analysis" *Science* 343 (6176), 1203-1205

Leung A., Agarwal, A., and Konana, P. 2013. "Co-Searching and Stock Cross-Predictability," *2013 Workshop on Information Systems and Economics*, Milan, Italy.

Liu, Y. 2006. "Word of Mouth for Movies: Its Dynamics and Impact on Box Office Revenue," *Journal of Marketing* (70:3), pp. 74–89.

Manning, C. D., Raghavan, P., and Schütze, H. 2008. "Introduction to information retrieval". Cambridge: Cambridge university press.

McAfee, A., and Brynjolfsson, E. 2012. "Big Data: The Management Revolution," *Harvard Business Review* (October): pp. 2-9.

Menzly, L., and Ozbas, O. 2010. "Market Segmentation and Cross-Predictability of Returns," *The Journal of Finance* (65:4), pp. 1555-1580.

Moe, W. W., and Fader, P. S. 2004. "Dynamic Conversion Behavior at E-commerce sites," *Management Science* (50:3), pp. 326-335.

Nini, G., Smith, D. C., and Sufi, A. 2009. “Creditor control rights and firm investment policy,” *Journal of Financial Economics*, *92*(3), 400-420.

Rui, H., Liu, Y., and Whinston, A. 2013. "Whose and What Chatter Matters? The Effect of Tweets on Movie Sales," *Decision Support Systems* (55:4), pp. 863-70.

Vosen, S., and Schmidt, T. 2011. "Forecasting Private Consumption: Survey-Based Indicators vs. Google Trends," *Journal of Forecasting* (30:6), pp. 565-78.

Watts, D. J., and Strogatz, S. H. 1998. "Collective dynamics of ‘small-world’networks," *nature* (393:6684), pp 440-442.

Wu, J., and Birge, J. R. 2014. “Supply Chain Network Structure and Firm Returns,” Available at SSRN http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2385217.

Wu, L., and Brynjolfsson, E. 2009. "The Future of Prediction: How Google Searches Foreshadow Housing Prices and Quantities," in *Proceedings of the 30th International Conference on Information Systems, Phoenix, Arizona*.