Forecasting the Impact of Product-Harm Events on Firm Value by Leveraging Negative Word of Mouth

Bolin Li

A Thesis

in

John Molson School of Business

Presented in Partial Fulfillment of the Requirements

for the Degree of Master of Science in Administration (Marketing) at

Concordia University

Montreal, Quebec, Canada

March 2019

CONCORDIA UNIVERSITY

School of Graduate Studies

This is to certify that the thesis prepared

By：Bolin Li

Entitled: Forecasting the Impact of Product-Harm Events on Firm Value by Leveraging Negative Word of Mouth

and submitted in partial fulfillment of the requirements for the degree of

Master of Science in Administration (Marketing)

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final examining committee:

Dr. Shannon Lloyd	Chair
Dr. Michele Paulin	Examiner
Dr. Kamila Sobol	Examiner
Dr. Michel Laroche	Supervisor

Approved by ______________________________

Chair or Department of Graduate Program Director

____

Dean of Faculty

Date March 22, 2019

Abstract

Forecasting the Impact of Product-Harm Events on Firm Value by Leveraging Negative Word of Mouth

Bolin Li

Product-harm events are always a nightmare for all stakeholders. Analysts believe that defective items may not only provide risks to the general population, but can likewise cause critical monetary and reputational harm to the firms. Since ignoring a problem does not lead to having it go away, more research is needed to shed new light on the way crisis and risk communication should take place once necessary. Prior study has suggested the complexities of consumer word of mouth effects and how to accurately forecast the impacts of product-harm events on firm value as important subjects. This study extracts the sentiments of consumer complaints in the context of product defects and examines if including consumer sentiment in time series models can improve forecasting performance. Authors make an empirical comparison between two multivariate time series forecasting methods: VAR (vector autoregressive model), and deep learning LSTM (long short-term memory model). Unique datasets, containing five-year data of all automobile nameplates for three major manufacturers in the U.S. are analyzed. The one-step rolling forecast approach is applied to validate time series forecasting values. The results of mean RMSE suggest that LSTM outperforms VAR predictive ability of firm value, and on average obtains 59.02% reduction in error rates when compared with error rates of VAR. It is also noticed that adding consumer sentiment in modeling can improve the predictive performance of both LSTM and VAR models; however, VAR-based models make greater progress in predictive error reduction with consumer sentiment. Implications for marketing research and managerial contributions are discussed.

Keywords: Product Harm; Firm Value, Long Short-Term Memory (LSTM); Vector Autoregressive Model (VAR); Text Mining; Consumer Complaints; Communication theory; Word of Mouth.

Acknowledgements

First and foremost, I would like to thank my supervisor, Dr. Michel Laroche who offered me great help in evaluating my research and guiding me when I have questions. It was in Dr. Laroche's Consumer Research Methodology class that I was exposed to text mining techniques for the first time. Dr. Laroche always encouraged me to ask more questions and to explore interesting topics in quantitative marketing research. Without his participation and input, I could not finish my thesis on schedule as expected.

Second, I would like to thank Mr. Hamid Shirdastian who also offered me tremendous help in data collection, research topic selection and research contribution. Mr. Hamid Shirdastian was always willing to share his valuable insights whenever I had questions about my research or writing. Without his generous assistance and enlightenment, I could not finish my thesis as I expected.

And last but by no means least, I would like to thank my family members and all my friends in Concordia University. Thank you for your selfless support and encouragement when I'm pursuing my master degree. It was great honor to be with all of you. Studying at Concordia JMSB is a very precious experience for me. It has become an unforgettable part of my life!

Table of Contents

List of Figures

List of Tables

List of Equations

1. Introduction

In the digital era, the growth of internet-based platforms and channels boosts the number of text data available to researchers. Instead of facing hundreds of thousands of observations, current marketers are more likely to analyze millions of online comments and reviews. Consumer word of mouth is more accessible for companies to measure and monitor than traditional word of mouth because it has become widely available and spontaneous. Online word of mouth is gradually becoming an important strategic issue for marketers, and managers are challenged to understand its influence on many key indicators (Kumar et al. 2016). In one report from Mintel (2015), around 70% of consumers in the U.S. search for online reviews or feedback from others before purchasing. More importantly, managers are cautious about online consumer complaints or negative word of mouth since the high visibility and transparency of negative online chatter can be devastating for companies' future performance such as sales, stock prices, and market share.

Prior research also has shown that negative word of mouth deserves more attention due to its heavy influence on consumer judgments (Herr et al. 1991). One reason is that negative comments and reviews are more informative about company performance and spread fast in social media compared with positive news (Chevalier & Mayzlin 2006; Kwak et al. 2010; Tirunillai & Tellis 2012). Tirunillai and Tellis (2012) also suggest that online word of mouth plays a key role in measuring firm performance, and becomes a leading indicator of stock market value.

One of the negative events that firms are frequently facing is product-harm events. In these events, defective products usually give rise to consumers' complaining that will further affects the reputation of companies. Normally, when one product is suspected of defects, a third-party governmental agency will investigate the event. Even though many investigations end with the product being cleared of suspected defects, there is an increasing number of investigations that lead to product recalls. (Eilert et al. 2017). Managers from various industries such as pharmaceuticals, foods, toys, and automobiles frequently face a large amount of product recalls. According to reports from the National Highway Traffic Safety Administration (NHTSA), around 20 million vehicles were recalled in 2010, but more than 53.2 million passenger vehicles were recalled in 2016 compared to 51.1 million in 2015. The year 2014, 2015 and 2016 are three straight years of record-setting automotive safety recalls of more than 50 million individual vehicles. The volume of one product recall could affect brand equity, spoil consumers' quality perceptions, damage the company's reputation, and lead to the losses of revenues and market share. (Laufer & Coombs 2006; Rhee & Haunschild 2006; Sullivan 1990; Van Heerde, Helsen, & Dekimpe 2007). Moreover, the volume of recalled products could affect investor's confidence in the company, which in turn leads to damaging financial value of companies (Chen, Ganesan, & Liu 2009).

Prior research on product recalls related variables mostly focus on studying the effects of product recall strategy, recall volume, or recall time on firm's performance such as sales and market share. Traditional time series statistical modeling methods are frequently applied to measure and understand the relationships. Tirunillai and Tellis (2012) investigated the relationships between user-generated content in product recall events and stock market performance, but the potential influences from product recall related variables were not considered. Specifically, a close study on forecasting the effect of complaints about defective products in product-harm events is very scarce.

This study differs from prior studies in three ways. First, we focus on both sentiments in consumer complaints and product recall volume at the same time. Sentiments in online consumer complaints about product-harm events, one form of user-generated content, is worth studying since it has a close relationship with involved company's stock market performance (Tirunillai & Tellis 2012). On the other hand, many researchers utilize product recall volume to evaluate the magnitude and the severity of product harm events are (Liu, Shankar, & Yun 2017; Liu & Shankar 2015; Borah & Tellis 2016; Eilert, Jayachandran, Kalaignanam, & Swartz 2017; Kalaignanam, Kushwaha, & Eilert 2013). In this study, we leverage both recall volume and sentiments in consumer complaints to forecast the impact on stock market performance. Second, we emphasize the forecasting performances of the Vector Autoregressive (VAR) model and the Long Short-term Memory (LSTM) model and aim to forecast the change of firm's stock market value. Accurate prediction eventually will help managers make marketing strategies and manage customer relationships. This research also compares these two types of multiple time series forecasting methods from the statistical modeling domain and the machine learning domain, investigating which method has a better predictive performance in the context of product-harm events. Lastly, prior studies suggest that the level of consumer complaint influences the company market value in a direct linear way (Luo, 2007; Luo 2009) or a non-linear way (Claro et al. 2014). Complex hidden relationships may exist between consumer complaints, product recall, and stock market performance. We are curious about the change of forecasting values with and without consumer sentiment in predictive models. By applying lexicon and rule-based sentiment analysis tools, we extract sentiment scores from consumer complaints and investigate if VAR-based models and LSTM-based models will improve the predictive performance by adding sentiment data.

In this research, we seek to answer the following three questions:

- Given the complexities in the effect of consumers complaints in product-harm events, how to accurately forecast a firm's stock market performance with VAR and LSTM by leveraging sentiment measures?
- When it comes to multiple time series modeling and forecasting, what are the major differences between a traditional statistical modeling method (VAR) and a machine learning technique (LSTM)?
- Compared with VAR, does LSTM have a better forecasting performance in the context of product-harm events? Which one makes more progress in predictive error reduction by including consumer sentiment in predictive modeling?

2. Literature Review

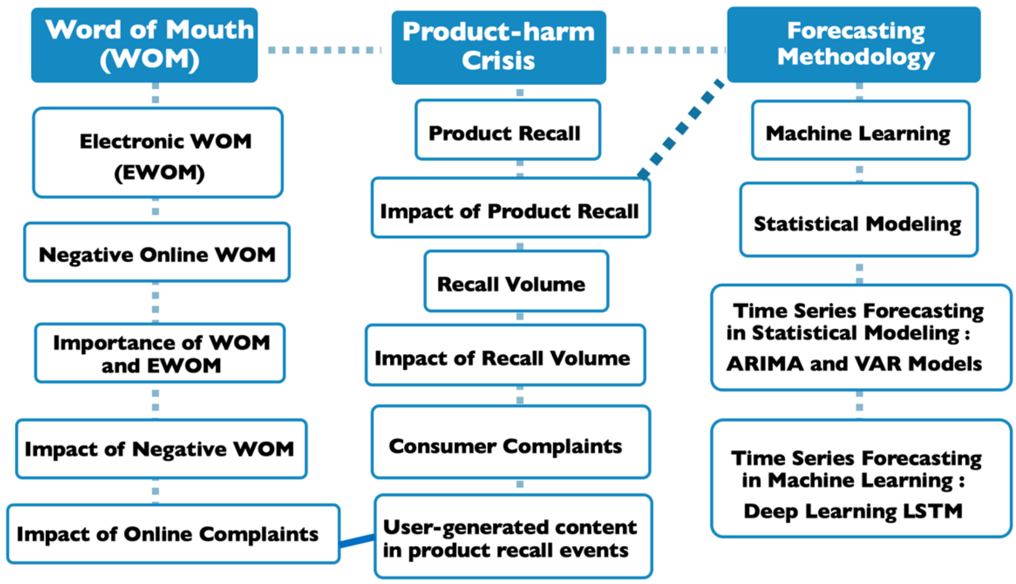


Figure 2.1 Overview of the Literature

In this section, we introduce literature related to three main subjects: word of mouth, product-harm crisis and forecasting methodology. Impacts of focal variables and important time series forecasting methods are also discussed.

2.1 Word of Mouth

2.1.1 Word of Mouth and eWOM

Word of mouth (WOM) was defined by Arndt in 1967 as "face-to-face communication about a brand, product or service between people who are perceived as not having connections to a commercial entity". Word of mouth in marketing is also seen as an unpaid form of promotion in oral or written form. Satisfied or dissatisfied consumers tell others how they like or dislike a business, product, service or event (Naylor 2016). Nowadays, WOM has been increasing in both importance and complexity because many internet-based platforms or channels have been established, and communication has expanded beyond interpersonal channels to the electronic word of mouth (eWOM). One early definition of eWOM is a positive or negative statement made by potential, actual, or former consumers about one product or company, and the statement is available to other people and institutions via the internet (Hennig-Thurau et al. 2004). More and more consumers review a product or a service that they have experienced online, forming a great number of eWOM. Electronic word of mouth (eWOM) contains a significant number of online complaints and compliments about the product or service because of the rapid growth of the internet and e-commerce.

2.1.2 Complex Effect of Negative WOM

Many traditional WOM articles (Brown et al. 2005; De Matos & Rossi 2008; Duhan et al. 1997; Singh 1990) suggest that negative WOM deserves particular attention because of its massive influence on consumer decision making (Herr et al. 1991). Some researchers have demonstrated the negative impact of negative WOM on firm's performance in online environments (Verhagen et al. 2013). However, some researchers question the prevalence of the negativity effect, noting that potential consumers evaluate such negative information through the perspectives of their existing attitudes. Studies suggest that positive attitudes toward brands of products tend to reduce the negativity effect (Ahluwalia 2000, 2002; Kirmani et al. 1999; Roehm and Brady 2007). Wilson et al. (2017) state that a consumer's connection with a brand may go beyond merely reducing the negative effect of eWOM and instead lead to a counterintuitive effect. The results of their studies support the positive effects of negative eWOM for consumers who have a high self-brand connection.

2.1.3 Complex Effect of Online Complaint

One of the typical negative eWOM examples is the online consumer complaint. In some industries such as telecommunication, automobiles, and pharmacy, relevant regulatory agencies public provide online records of consumer complaints and assess companies' performances. Due to defective products or product accidents, product-harm events frequently cause consumer complaints and lead to a product recall. Prior studies found that product recalls also damage companies' reputation in consumers' mind (Dawar & Pillutla 2000). Such damage may lead to consumers complaining online because consumer use of online platforms is mobile and pervasive (Borah & Tellis 2016). There is a growing consensus that that public complaints influence consumers' perceptions about a company and help spread the failure of products and services by describing stories of negative experiences (Luo, 2009; Winchester, Romaniuk, & Bogomolova, 2008). After receiving bad products or services, consumers could easily post negative complaints through online platforms and are more likely to include the details of problems or defects (Bentivegna, 2002; Santos & Fernandes, 2011). Brown and Reingen (1987) mention that a third-party agency's credentials may further intensify the impact of complaints. In this study we look into one typical example of third-party agency in the automobile industry, the National Highway Traffic Safety Administration (NHTSA). As an organization of the Executive Branch of the U.S. government, NHTSA posts product-harm records, consumer complaints and recall information online. Figure 2.1 is a screenshot of one complaint about Toyota Prius 2011 posted on the website of NHTSA.

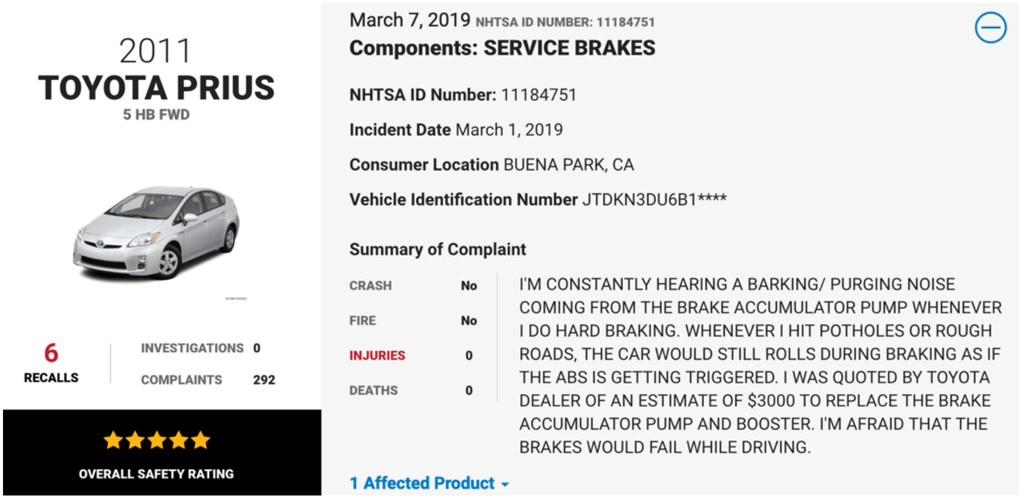


Figure 2.2 Complaint about Toyota Prius 2011

In product-harm events, the emotions in consumer complaints are evident, and the stories from consumers could profoundly influence other consumers in a way that evoke good and evil battle or in the perception of harmed victims (Richins 1983; Laer & Ruyter 2012). Previous evidence suggests that the market value of company performance has a direct linear relationship with the severity of the complaint (Luo 2007; Luo 2009). However, Claro et al. (2014) manifest a nonlinear relationship between consumer complaint level and the company market value because of a tradeoff rationale in the impact of the severity of the complaint. High levels of complaints negatively influence company market value while low levels of complaints positively impact company market value (Claro et al. 2014).

2.2 Product-harm Crises

Companies, especially from telecommunication, automobile, and pharmaceutical industries, often face product-harm crises that may lead to a product recall. The involved companies need to retrieve recalled products from all channels and end consumers. Nowadays, products are becoming more complicated, consumers are more focused on quality and safety, and companies and government agencies have been closely monitoring the products, so that a product recall is more likely to occur (Berman 1999). Many researchers suggest that product recalls have the potential to influence the quality and safety perceptions of products, damage carefully developed brand images, tarnish companies' reputation, and lead to market share or sale losses (Laufer & Coombs 2006; Rhee & Haunschild 2006). Moreover, a product recall could also hurt investors' confidence, which could lead to a further decline in the financial value of a company (Chen, Ganesan, & Liu 2009).

2.2.1 Impact of Product Recall Volume

Product recall volume is one of the most important key indicators to evaluate the magnitude of recall events and could be used to predict the firm value. Prior research has shown that recall volume affects various firm's important factors in both the short-term and the long-term (Liu, Shankar, & Yun 2017). In the short run, the effects are reflected in short-term revenues and costs. The sales revenues of the implicated product decline since recall volume negatively affects consumer preferences for the brand (Liu & Shanker 2015). Consumers are more likely to generate negative online complaints, which can affect others' decision making and product sales of involved companies. The short-term costs are from the product-harm related investigation, notification, repairs, and replacement of defective products (Bromiley & Mareus 1989). Besides, recall volume also affects potential long-term revenues through damage to intangible assets, for example, firm reputation, brand equity, and customer equity (Rhee & Haunschild 2006). Handling large units of defective products means taking a long time to fix the issues and undertaking unexpected costs, such as increased lawsuits and regulatory fines. (Govindaraj, Jaggi, & Lin 2004). Borah and Tellis (2016) state that the larger the recall volume is, the more affected owners may spread negative word of mouth both online and offline, resulting in a further loss in sales.

In product-harm events, we find that investors receive bad news from two typical channels: the first channel is word of mouth from other consumers, and the second one is the recall announcement from the implicated company or the third-party agency. Solely considering product recall volume doesn't fully indicate how severe product-harm events are since sometimes a large volume of units are recalled only for updating the manual of automobiles. Including consumer sentiment of complaints about product-harm events may provide more information related to the impacts of product-harm crisis. We thus study both product recall volume and consumer sentiment in this study and aim to examine if include consumer sentiment will improve the forecasting. Research in finance suggests that bad news may have an impact on investors' trading behavior, and thus the stock market may display a negative drift in the long term (Barberis, Schliefer, & Vishny 1998; Chan 2003). A large recall volume sends a negative signal to investors, giving rise to pessimistic outlooks of one firm's potentiality or development. This negative reaction could further damage the company's market value and prospects. Because of the growing amount of product-harm records and recall volume, it is important for researchers and managers to foresee the impacts of recall volume and understand how it will influence firm's market performance, which will help formulate future strategies and implement remedial actions.

2.3 Marketing Research in the Big Data Era

In the last decade, the growth of internet-based platforms or channels boosts the number of data available to researchers. Instead of facing hundreds of thousands of actual observations, current marketers have the opportunity to analyze millions of online followers, consumer reviews, or complaints. Harnessing big data with appropriate methods in marketing research helps gain precious insights and thus create more value for consumers and companies. Traditionally, marketing researchers apply statistical modeling to predict the value of the target variable. As the amount of data grows further, an increasing number of researchers start to train models with machine learning algorithms. In the business context, firms also start to extract insights with an immense amount of consumer data by applying statistical modeling or machine learning algorithms.

However, big data also brings some challenges to marketing researchers. Larger data analysis needs correspondent and updated technologies for data collection and data storage. Besides, processing large data sets occupies a huge amount of computational power, and thus may take a longer time to estimate sophisticated models or predict target variables. Marketers have been using statistical modeling for a long time, but statistical models are not specifically designed for utilizing a huge amount of data. Since computer scientists have a long history dealing with estimation techniques and modeling approaches in the data-rich environment, marketing researchers are starting to apply machine algorithms from computer science to consumer research (Guttag 2013).

2.3.1 Origin and Definition of Machine Learning

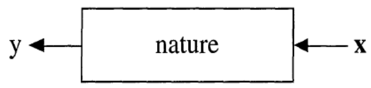
In the mid-1980s, two new algorithms for fitting data became available: neural networks and decision trees. After that, the machine learning community has sprung up. This community consisted of computer scientists, physicists, engineers, and statisticians, and they began to work on complex prediction problems: speech recognition, image recognition, nonlinear time series prediction, handwriting recognition, and prediction in financial markets (Breiman 2001). As a researcher in machine learning, statistics, and artificial intelligence areas, Michael I. Jordan state that numerous ideas in the machine learning community had prehistory in the statistics community such as logistic regression, PCA, canonical correlation, graphical models, K-means, etc. Different from statistical researchers, machine learning practitioners are exceedingly creative at making use of advanced computing architectures, taking ideas across fields, and mixing them to solve challenging problems (Jordan 2014).

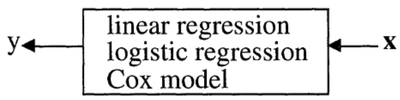
Machine learning is a broad subject, and defining it is relatively hard. One broader definition of machine learning was proposed by electrical engineer and computer scientist Arthur Samuel, which is "one field of study that gives computers the ability to learn without being explicitly programmed." Tom Mitchell (1997) also provides a more engineering-oriented definition: "A computer program is said to learn from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E." Researchers also view machine learning as a field of pattern recognition, prediction algorithms and computational learning theory in artificial intelligence (Choudhury, Kumar, & Nigam 2015). In general, machine learning involves observing a set of examples that represent incomplete information about some statistical phenomenon and then attempting to infer something about the process that generated those examples (Guttag 2013). Machine learning can also be viewed as the science of getting computers to learn from data without explicitly assuming a stochastic data model, and to find a learning algorithm that operates on input variables to predict the response variables based on the predictive accuracy.

2.3.2 Machine Learning and Statistical Modeling

Machine learning and statistical modeling are closely related subjects. They both aim to reach conclusions from analyzing data. Two main objectives in analyzing the data are: predicting what the responses are going to be based on future input and extracting information about the association between response variables and input variables. Kübler et al. (2017) propose that "statistics and machine learning have very much in common and have similar DNA." Shared common methodologies in machine learning and statistical modeling are regression, resampling, classification, and non-linear methods. Logistic regression is one of the most popular modeling methods that's widely applied in both machine learning and statistical modeling (Kübler et al. 2017). However, as the historical backgrounds of Machine Learning and statistics are very different, two fields gradually developed and formed different philosophies. Machine Learning is more result-oriented and emphasizes more prediction accuracy, while statistics modeling is more restrictive and emphasizes more interpretability. Machine learning practitioners frequently face high-dimensional problems and deal with an indefinite number of variables in large data sets. However, statisticians generally study low-dimensional issues and pay more attention to formal statistical inference such as optimal estimators, confidence intervals, and hypothesis testing (Wasserman 2012).

In Leo Breiman's research "Statistical Modeling: The Two Cultures" (2001), he illustrates that both statisticians and machine learning practitioners use modeling, but there are two cultures in applying modeling. He explains that "One culture assumes that the data are generated by a given stochastic data model. The other uses algorithmic models and treats the data mechanism as unknown." As the figure 2.1 shown, Breiman names the two cultures as data modeling and algorithmic modeling, and consider the real relationship between predictive variables (X) and target variables (Y) is in one black box since no one can truly know and see the exact relationship between X and Y. In data modeling culture, researchers first tend to propose one data model, then make assumption for the inside of the black box, and use goodness of fit tests to evaluate how well the data on hand fit with proposed data model. Typical examples are linear regression and logistic regression models. However, in the algorithmic modeling culture, the inside of the black box is considered as unknown, researchers are more interested in applying different machine learning algorithms and finding the better algorithm that can provide great forecasting performance. The machine learning algorithms are designed to learn from data with certain optimization algorithms. A majority of statisticians has been committed to the use of data modeling, while the machine learning community has been developing algorithmic modeling. (Breiman 2001).



Culture		Model Validation
Data Modeling	Assuming a stochastic data model for the inside of the black box. 	goodness-of-fit tests residual examination

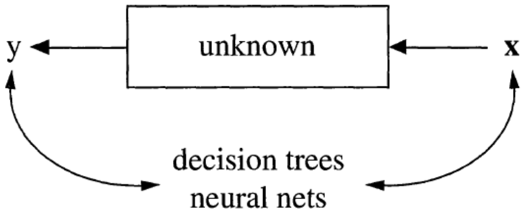
Algorithmic Modeling	Considers the inside of the box complex and unknown. 	Measured by predictive accuracy

Figure 2.3 Two Cultures: Data Modeling and Algorithmic Modeling

2.3.3 Time Series Forecasting in Statistics and Machine Learning

Because of the complex effects of consumer complaint and the diversified impacts from product recall volume, accurately predicting the effects on firm value in the stock market become crucial for managers. Time series modeling and forecasting analysis approaches are widely used for solving this type of questions that could help make marketing strategies and ensure remediation and resolution for managers. The main objectives of time series forecasting are to analyze historical time series data of observations, build models to capture the structures of the data, and predict the future values of target variables.

Traditional statistical time series approaches, ARIMA models, and Vector Autoregressive Models (VARs) are widely used in modeling financial and economic time series (Adebiyi et al. 2014; Alonso & Garcıa-Martos 2012; Banerjee 2005; Khashei & Bijari 2011). However, traditional statistical methods lack flexibility in changing model structure and tuning parameters, and some of the assumptions in statistical modeling are hard to meet. For example, one assumption, a constant standard deviation in errors ARIMA model, sometimes is hard to satisfy in practice. Deep learning neural networks don't require any assumptions of input data and are capable of identifying complex structures by tuning parameters or changing number of hidden layers. One typical technique in deep learning called LSTM has also been widely applied in time-series prediction (Brownlee 2016; Hochreiter & Schmidhuber 1997; Gamboa 2017; Gers et al. 2000; Graves et al. 2009, 2013; Roondiwala 2017; Schmidhuber 2015; Yim 2002; Zhang & Qi 2005). Besides, LSTM-based models have been applied in economic and financial subjects such as predicting the volatility of the S&P 500 (Huck 2009; Kohzadi 1996; Xiong et al. 2015) and measuring the impact of incorporating news for selected companies (Ding et al. 2015; Joshi et al. 2013).

3. Empirical Context and Data

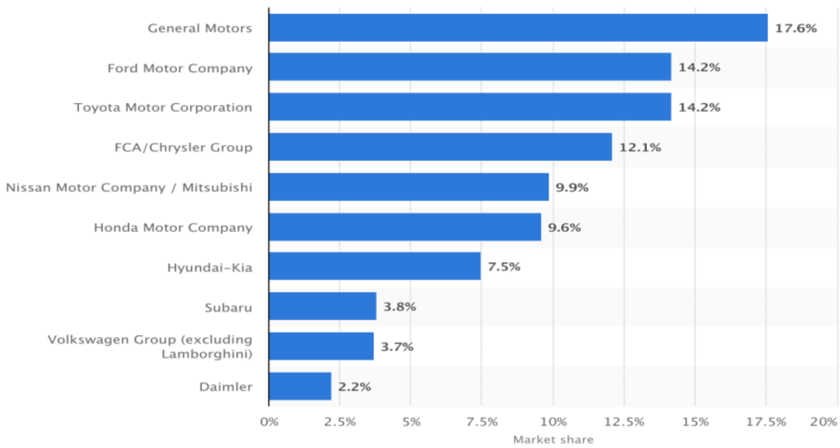


Figure 3.1 The U.S. Leading Automobile Manufacturers Market Share in 2017

We compare statistics, and machine learning forecasting approaches in product-harm events and specifically focus on the U.S. automobile industry. This critical industry has produced around $70 billion gross output in 2015 according to a report from Statista. Studying data from one specific industry could exclude the need for considering other cross-industry factors and improve internal validity. The data for our forecasting comes from two sources: NHTSA (National Highway Traffic Safety Administration) and the Federal Reserve Economic Data website. Specifically, we collect five-year consumer complaints text of defects and product recalls records with timestamps from NHTSA. Top 10 automobile manufacturers in the U.S. are listed in figure 3.1. All automobile nameplates of three selected leading automobile manufacturers (Ford Motor, Honda

Motor, and Toyota Motor) in the U.S. are considered, which account for 38% of the whole market share in 2017. We selected three major manufacturers for two reasons: first, they are multinational manufacturers from three different countries and produce multiple types of automobile products; second, they all went public in NASDAQ from 2005 to 2010, therefore, we can collect public historical stock market data and other related economic data sets. We downloaded stock prices data of the three leading manufacturers and historical S&P index from Yahoo Finance website. Economics time series, inflation rate and treasury rate historical data are also collected from the Federal Reserve Economic Data website. Figure 3.2 shows the line charts of key data sets mentioned above.

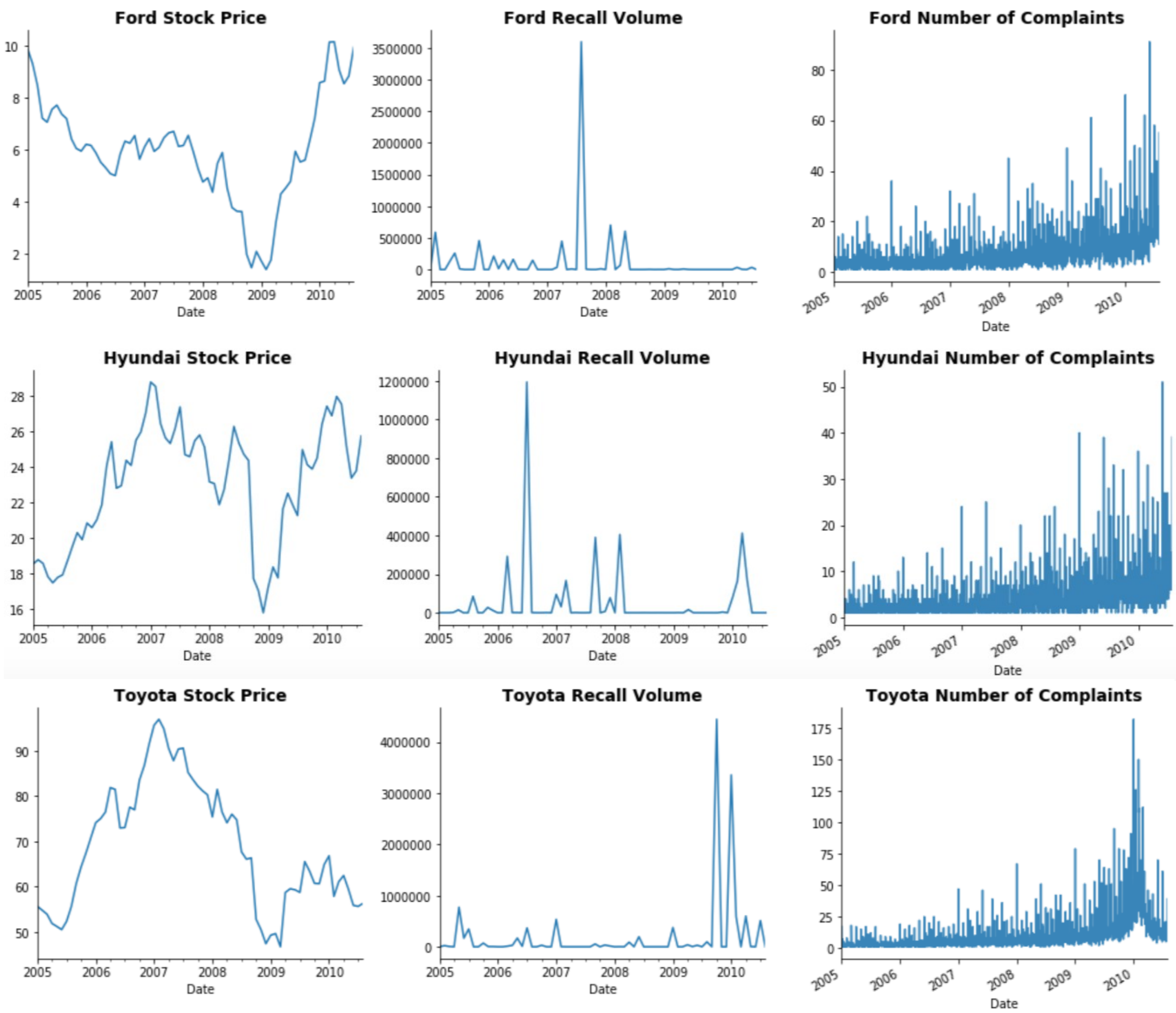


Figure 3.2 Line Charts of Key Data Sets

4. Variables and Measures in Forecasting

This section mainly introduces the target variable and focal predictive variables and elaborates the measures of these variables in our multivariate time series forecasting.

4.1 Target Variable

The target variable in this study is the abnormal stock return of automobile companies from Jan-2005 to July-2010. Each stock price time series data has a set of variables: [1]Open Price, [2]High Price, [3]Low Price, [4]Close Price, [5]Adjusted Close Price and [6]Volume. Since Adjusted Close or adjusted closing price takes corporate actions such as stock splits, dividends, and rights offerings into consideration, we use historical stock returns with adjusted closing prices. We calculated the Beta coefficient and the real risk-free rate with Treasury Constant Maturity Rate, inflation rate historical data, stock adjusted closing price and S&P index. The abnormal return, the difference between actual return and expect rerun. We use the capital asset pricing model (CAPM) to calculate expect return of one stock. The formula of CAPM is written as the following function:

$$ER_i = R_f + \beta_i \left(ER_m - R_f\right) \quad (4.\ 1)$$

ER_i is the expected return of the investment or the stock, R_f is risk-free rate, β_i = Beta of the investment, ER_m is the expected return of the stock market, and $(ER_m - R_f)$ is the stock market risk premium.

4.2 Focal Predictive Variables

In this study, we scrape product recall records and negative word of mouth or consumer complaint from the NHTSA from Jan-2005 to July-2010. We measure the product volume variable as one predictive variable by summing the monthly number of the recalled automobiles for all brands or nameplates in one manufacturer. When it comes to the sentiment measurements of complaints about one manufacturer, we

[1] Open price is the price when a stock first trades upon the opening of a given period of time.
[2] High price is the highest price of a stock in a given period of time.
[3] Low price is the lowest price of a stock in a given period of time.
[4] Close price is the price when a stock last trades upon the closing of a given period of time.
[5] Adjusted closing price is a stock's closing price on any given day of trading that has been amended to include any distributions and corporate actions during a given period of time.
[6] Volume is the number of shares or contracts traded during a given period of time.

apply the lexicon and rule-based sentiment analysis tool named VADER (Valence Aware Dictionary and Sentiment Reasoner) to capture sentiments in consumer complaints. Two sentiment variables: negative consumer sentiment and compound consumer sentiment are used in our forecasting. Besides, we also take S&P index, treasury rate and inflation rates into consideration, because they contain the information of the market trends and fluctuations.

The reason why we apply VADER is because it doesn't require training data and has a generalizable, valence-based, human-curated gold standard sentiment lexicon (Hutto & Gilbert 2014). Researchers suggest that VADER is attuned explicitly to sentiments expressed in microblog-like contexts and outperforms individual human rater in terms of accessing emotions (Hutto & Gilbert 2014). In VADER, each of the words in the lexicon is rated with positive, natural, and negative scores. When analyzing with VADER, informal writing: multiple punctuation marks, acronyms, and an emoticon and other things such as word context, punctuation and so on are also considered. There are four sentiment metrics from word ratings in VADER: positive, neutral, and negative, and compound. Positive, neutral and negative scores represent the proportion of the text that falls into those categories. Negative score is calculated by only summing negative score within all complaints and then standardizing the sum of scores in each complaint. The compound score is computed by summing all three types of lexicon ratings: negative score, positive score and neutral score in each complaint, and then standardized the sum to a range. Negative score solely reflects the total negative sentiment without considering other sentiments, while the compound score represents the relative sentiment score by consider three sentiment scores together. In this study we use both the negative sentiment score and compound sentiment score as two predictive variables of consumer sentiments.

5. Time-series Forecasting Methodology

Forecasting can be simply defined as predicting the future values of the series by analyzing historical data sets. To compute the future forecasting values, historical records of the series of target variables and exogenous variables are fitted into different statistical models or machine learning algorithms. A statistical model or machine learning algorithm can be viewed as "a rule" to analyze the historical time series. In time-series statistical models, the univariate time series model uses the past values in one series to predict the future values of the series. One popular time series statistical model is the autoregressive moving average model (ARMA) that can be written as the following functions:

(5. 1) $$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} \dots + \phi_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} \dots + \theta_q \varepsilon_{t-q}$$

(5. 2) $$y_t = c + \sum_{i=1}^{p} \phi_i y_{t-i} + \varepsilon_t + \sum_{i=0}^{q} \theta_i \varepsilon_{t-i}$$

Where $\phi_i \neq 0$, $\theta_i \neq 0$ and $\sigma^2 > 0$. The parameter p and q are the orders of Autoregression (AR) and Moving Average respectively. ARMA captures the unidirectional relationship within the series itself.

5.1 Vector Auto-Regressive Models (VAR)

Compared with univariate time series models, multivariate time series models not only can model the dynamics of each series itself but also capture the interdependency among these series. Multivariate time series modeling is widely used in many real-world domains such as weather forecasting, healthcare, and financial market prediction. Auto-Regressive Model (VAR) is one of the most prevalent models in analyzing the economic and financial time series (Melnyk et al. 2016; Taylor 2007). For example, researchers have applied VAR to estimate the relationships between negative online chatter in social media and companies' sales (Borah & Tellis 2016), and to detect anomalies in aviation system (Melnyk et al. 2016). Del Negro and Schorfheide (2001) describe Vector Autoregressive Models (VARs) as "At first glance, VARs appear to be straightforward multivariate generalizations of univariate autoregressive models. At second sight, they turn out to be one of the key empirical tools in modern macroeconomics". VAR is designed to capture the joint dynamics of multivariate linear time-series and mainly used for two purposes: prediction and structural analysis. We focus on predicting with the reduced form of VAR in this study. To illustrate VAR, let Y_t be a vector of n variables at time t:

(5. 3) $$Y_t = [\, Y_{1,t} + Y_{2,t} + \dots + Y_{n,t} \,]'$$

A p-order vector autoregressive process generalizes a p-order univariate autoregression or AR(p) process to n variables:

(5. 4) $$Y_t = G_0 + G_1 Y_{t-1} + G_2 Y_{t-2} + \dots + G_p Y_{t-p} + e_t$$

The VAR model describes the vector Y_t as a function of its past values: Y_{t-1}, Y_{t-2} … Y_{t-p} and a vector of stochastic error term e_t . G_0 is an (n X 1) vector of constants, G_j is a (n X n) matrix of coefficients, and e_t is an (n X 1) vector white noise. The estimation of the VAR model is performed with an ordinary least squares (OLS) estimator.

5.2 Long Short-Term Memory (LSTM)

However, some studies suggest that VAR-based models have some limitations. VAR is not designed for modeling complex relationships and requires strict assumptions (Namini 2018). Recent significant advances have been made in sequential data using Recurrent Neural Network (RNN). The Long-Short Term Memory (LSTM) algorithm is the new technique that was initially introduced by Hochreiter and Schmidhuber (1979), which is also a kind of RNN with the capability of remembering the historical values for forecasting. Researchers have applied RNN-based models to predict stock returns (Lee & Yoo 2007), but the LSTM algorithm has not been carefully explored in the context of product-harm events. To better understand LSTM, it is important to have a glimpse of what artificial neural network and RNN look like.

5.2.1 Artificial Neural Network (ANN)

Artificial Neural Networks (ANN) are one of the most important tools used in machine learning. Neural networks consist of at least three layers: an input layer, hidden layers, and an output layer. Layers themselves are just sets of nodes that transform the input data in the input layer into the outcomes in the output layer, the number of features or predictive variables determines the number of nodes in the input layer, and the values of nodes in the output layer represent the value of target variables. The inputs and outputs correspond to visible things that can be stored as data, while the hidden layers contain information that are not visible and cannot be saved as data directly. More nodes in hidden layers help the neural network capture more complex interactions.

Synapses are links in a neural network that connect nodes in different layers. Each connection between two nodes has a unique synapse with a weight attached to it. Namin et al. (2018) suggest that the weights in synapses play the role of a decision maker to decide how much information can pass through. The weights of nodes in one hidden layer show the strength or extent of this layer in a neural network, and training a neural network is adjusting the weight for each synapse. Figure 5.1 is an ANN example that contains three input features or predictive variables, two hidden layers, eight nodes in hidden layers, and one target variable.

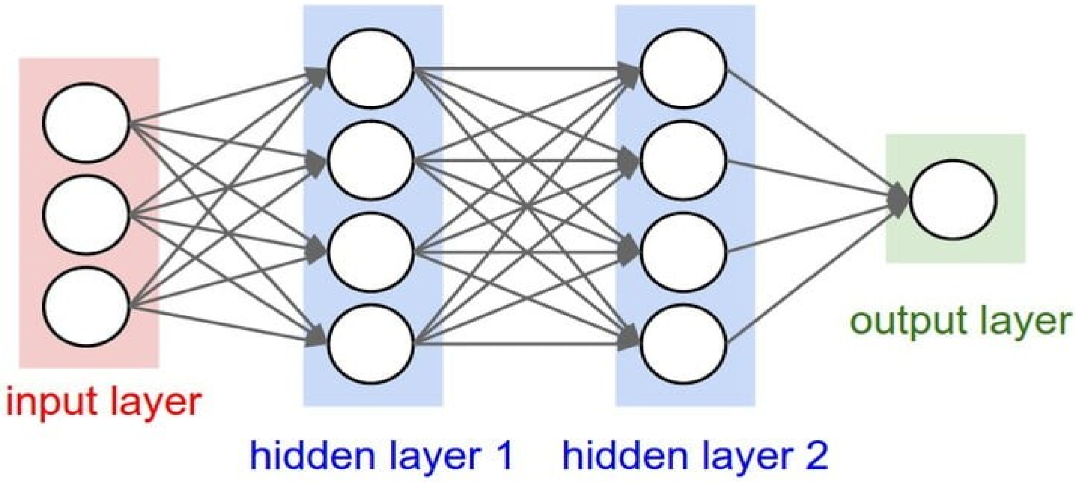


Figure 5.1 Artificial Neural Network ("Neural Network architectures", 2018)

With input data, artificial neural network (ANN) uses forward propagation to generate values through each layer. The data input from the input layer go through nodes in hidden layers, the nodes apply an activation function such as sigmoid and tangent functions on the weighted sum of inputs, which allows a neural network to capture the nonlinearities. In the end, a vector of probabilities is generated in the output layer.

However, making accurate predictions is hard since there are also many values of error corresponding to the many points for which we are forecasting. ANN has a loss function that aggregates all the values of the error to measure the predictive performance. A lower value of the loss function means a better ANN. The goal is to find the weights that give the lowest value for the loss function, and gradient descent is one typical algorithm in machine learning that is used to achieve this goal. Backpropagation is the technique used to calculate weights and optimize complex deep learning models. Backpropagation takes the prediction error from the output layer and propagates it back through the hidden layers towards the input layer. To find the lowest value of the loss function, backpropagation calculates slopes sequentially and updates all weights repeatedly in neural networks with a gradient descent algorithm. The training process of ANN is to find the weights with the minimized value of the loss function.

5.2.2 Recurrent Neural Network (RNN)

Most of the neural networks are feedforward, where the activations flow from input layer toward the output layer. However, Recurrent Neural Network (RNN) is one special type of artificial neural network that has connections pointing backward (Géron 2017). One simple example of RNN is shown in figure 5.2 below. It is composed of only one node receiving inputs and it sends the outputs back to itself. At timestamp t, this RNN receives current input X_t and the input from previous timestamp h_{x-1}. If it is a layer of multiple recurrent nodes, at one timestamp t, every node receives both input vector X_t and the output vector h_{x-1} from the previous timestamp. Two sets of weights are in each node, one set is W_x for input vector X_t, another set is W_h for last output vector h_{x-1}. The output of a single recurrent node is calculated by the following equation:

(5. 5) $$h_t = \phi(X_{(t)}^T \cdot W_x + h_{(t-1)}^T \cdot W_h + b)$$

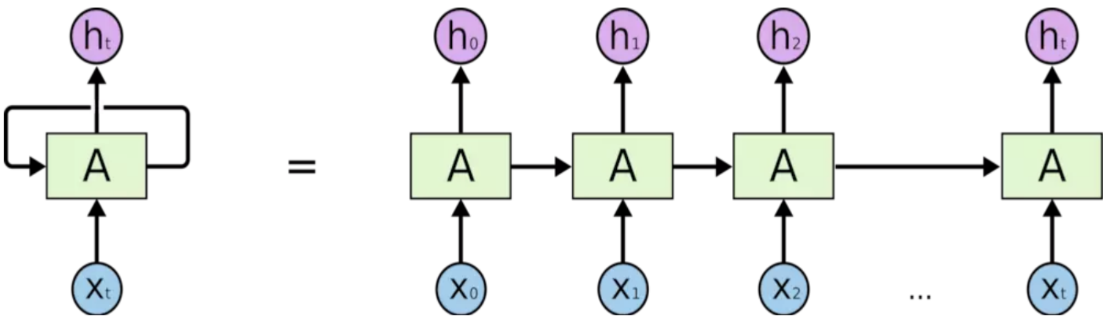


Figure 5.2 Recurrent Neural Network (Zhu, X. 2015)

The h_t is a function of $h_{(t-1)}$ and X_t, h_{t-1} is a function of $h_{(t-2)}$ and $X_{(t-1)}$, h_{t-2} is a function of $h_{(t-3)}$ and $X_{(t-2)}$, and so on. Therefore h_t is a function of all the input since timestamp t = 0, which are X_0, X_1,...and X_t. When t is 0, there are no values previous of outputs, so the outputs are usually assumed as 0. RNN can process a sequence of inputs and generate a sequence of outputs, this kind of structure can be used for time series forecasting such as stock price prediction. For example in Figure 5.3, RNN computes based on the 5-day price inputs (X_0, X_1, X_2, X_3 and X_4), and it outputs the prices shifted by one day into the future (Y_1, Y_2, Y_3 and Y_4). After feeding a sequence of inputs, we can also ignore all outputs except for the last output, one typical application is inputting a sequence of words (X_0, X_1,...and X_3) corresponding to a consumer review to compute a sentiment score (Y_3) (Géron 2017).

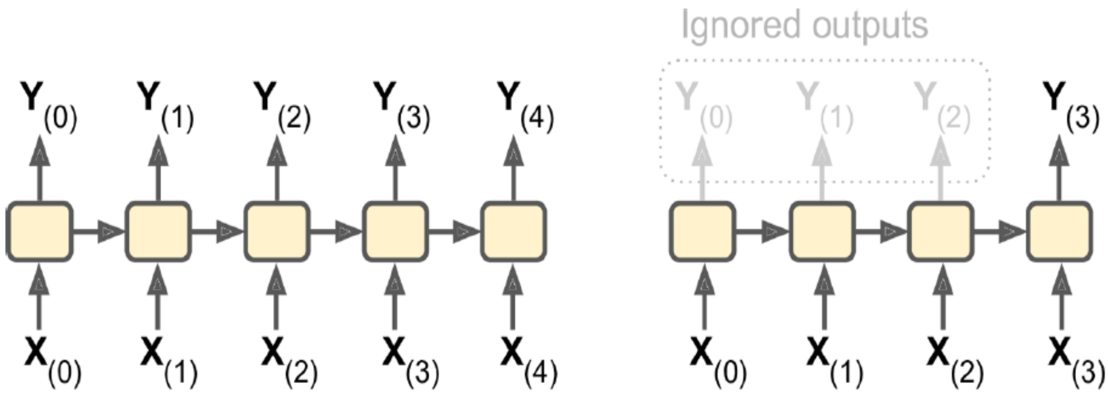


Figure 5.3 Recurrent Neural Network Time Series Forecasting (Géron 2017)

When training an RNN on long sequences, we may need to run the model over many times and suffer from the vanishing or exploding gradient problems, and the training will be slow. Also, very deep RNN face the fact that the memory of the first inputs gradually fades away, therefore some essential information may be lost after multiple time steps. In other words, the RNN can gradually forget the early inputs, which may completely misinterpret the data (Géron 2017).

5.2.3 Long Short-term Memory (LSTM)

The Long Short-Term Memory (LSTM) was proposed in 1997 by Hochreiter and Schmidhuber, working as an extension of RNN to solve the problems in RNN (Hochreiter & Schmidhuber 1997). The LSTM has gradually been improved over the years by researchers (Sak et al. 2014; Zaremba et al. 2015), and it can learn what to store as long as it is needed, what to drop, and what to extract whenever it is required.

In the LSTM, there are two vectors of data: long-term state $c_{(t)}$ and short-term state $h_{(t)}$. As the Figure 6 shows, long-term state $c_{(t-1)}$ transport in the upper line of each cell from left to right. As figure 5.4 shows. the $c_{(t-1)}$ first goes through a forget gate dropping some memories and then pass an addition operation adding new memories from input gate, and thus $c_{(t)}$ is generated and can be sent out to the next cell. To calculate short-term vector $h_{(t)}$ at timestamp t, we need to feed the input vector $x_{(t)}$ and the previous short-term vector $h_{(t-1)}$ into four different fully connected layers (FC). Four layers consist of one main layer and three gate controllers (Géron 2017).

- The main layer is $g_{(t)}$. It processes $x_{(t)}$ and $h_{(t-1)}$. A basic cell or node in the neural network, there is nothing else than a $g_{(t)}$. The output of $g_{(t)}$ in an LSTM is partially stored in the long-term output $c_{(t)}$.
- $f_{(t)}$ controls the forget gate that drops needless memories in the long-term state.
- $i_{(t)}$ controls the input gate and determine which parts of $g_{(t)}$ needs to be added in the long-term state.
- $o_{(t)}$ controls the output gate that controls which parts of the long-term state should be stored in the current short-term state $h_{(t)}$ and output $y_{(t)}$.

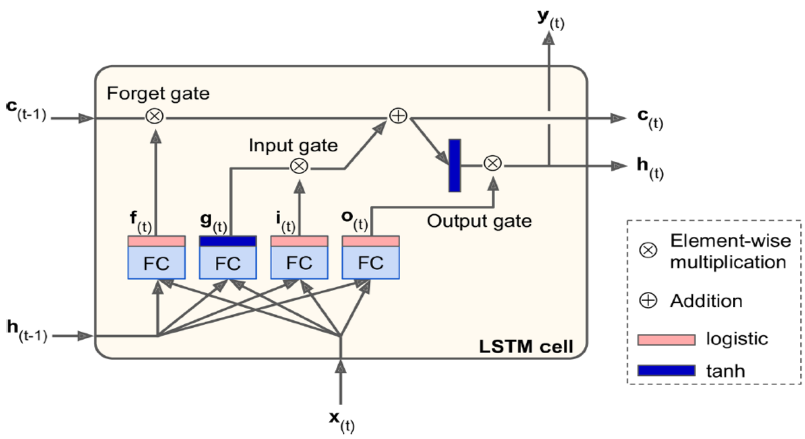


Figure 5.4 Long Short-term Memory Cell (Géron 2017)

6. Experimental Setup

In this section, we introduce the rolling forecast approach for time series cross-validation and the assessment metric RMSE used to conduct forecasting evaluation.

We then present experiments setup and implement experiments for comparisons lastly.

6.1 Time Series Cross-validation

Time series cross-validation is different from standard cross-validation in traditional machine learning problems because of the dependency between two consecutive observations. In this study, we conduct time series cross-validation based on "Rolling Forecasting Origin" (Hyndman & Athanaspoulos 2014); it is also called "walk-forward model validation". We split first 80% of the observations in time series (1st, 2nd, …, Kth observation) as the minimum training set, and the rest works as the primary test set. To perform a one-step rolling forecast in this research, we first select the K+i observation as one test set and estimate the model with the 1st, 2nd, …, K observations. The predictive error of the K+i observation is calculated with predictive value and actual value. Repeat the above steps for i = 0, 1, …., T - k where T is the total number of observations. With the rolling forecast approach, the overall predictive error of forecasting models then can be calculated.

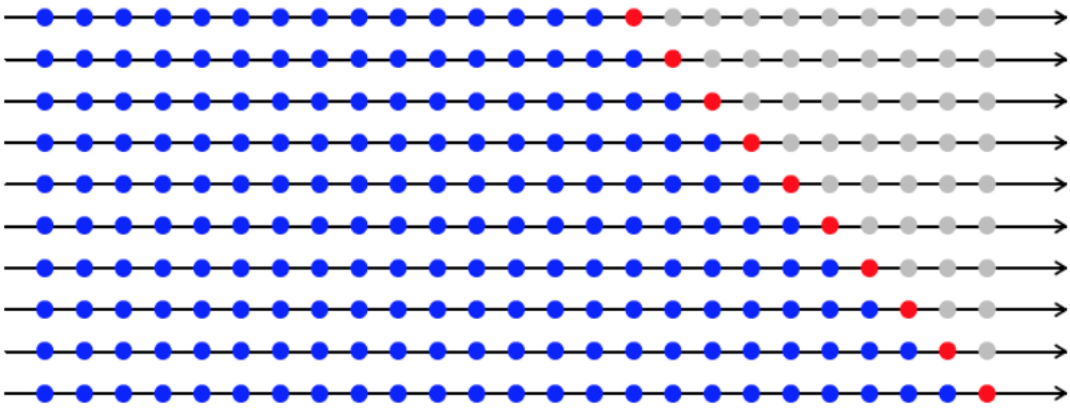


Figure 6.1 One-step Rolling Forecast (Hyndman & Athanaspoulos 2014)

6.2 Assessment Metric

Predictive model performance is computed according to the RMSE (Root-Mean-Square Error). RMSE compares the difference between actual values and predicted values, and penalizes any significant difference between the real value and the predicted value. The final value generated from RMSE has the same units with the forecast values. The equation of RMSE is shown below where N is the total number of observations in the testing set, i represents one observation in the testing set, Y is the actual value and $\hat{Y}$ is the predicted value with an estimated model.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(Y_i - \hat{Y}_i)^2} \tag{6. 1}$$

6.3 VAR and LSTM Implementation

We use three data sets to test forecasting and compare LSTM with VAR. Three data sets are named as Ford, Honda Motor Company (HMC) and Toyota to predict the abnormal stock return of three automobile manufacturers: Ford Motor Company, HMC, and Toyota Motor Corporation respectively. Each data set contains company stock abnormal returns, company product recall volume, sentiment measurements in product complaints, and economic variables: S&P index return, inflation rate, and treasury rate. Table 1 lists the names of the data sets and predictive variables.

Table 6.1 Three Data sets and Predictive Variables

Data Set	Variables in the Data Set
Ford	Ford Stock Abnormal Return, Ford Recall Volume, Ford Negative Score, Ford Compound Score, S&P Index Return, Inflation Rate, Treasury Rate.
HMC	HMC Stock Abnormal Return, HMC Recall Volume, HMC Negative Score, HMC Compound Score, S&P Index Return, Inflation Rate, Treasury Rate.
Toyota	Toyota Stock Abnormal Return, Toyota Recall Volume, Toyota Negative Score, Toyota Compound Score, S&P Index Return, Inflation Rate, Treasury Rate.

For each data set, we estimate VAR-based models and LSTM-based models with four input data that have four different variable combinations. Specifically, input data 1 only contains product recall volume and other economic variables. By including one consumer sentiment measurement: negative score of complaints in input data 1, we name the new data as input data 2. Input data 3 is generated by adding another different consumer sentiment measurement: compound score of complaints in input data 1. We put two sentiment measurements, product recall volume as well as other economic variables together and have the full data as input data 4. The objectives of the different input data are to find which model has the best forecasting performance and to investigate if including sentiment related measurements can improve the forecasting performance. Figure 6.2 shows details of variable combinations in four input data. First input only contains company stock abnormal return and economics variables. Second, third and fourth data sets contain company stock abnormal return, economics variables and different combinations of sentiment variables.

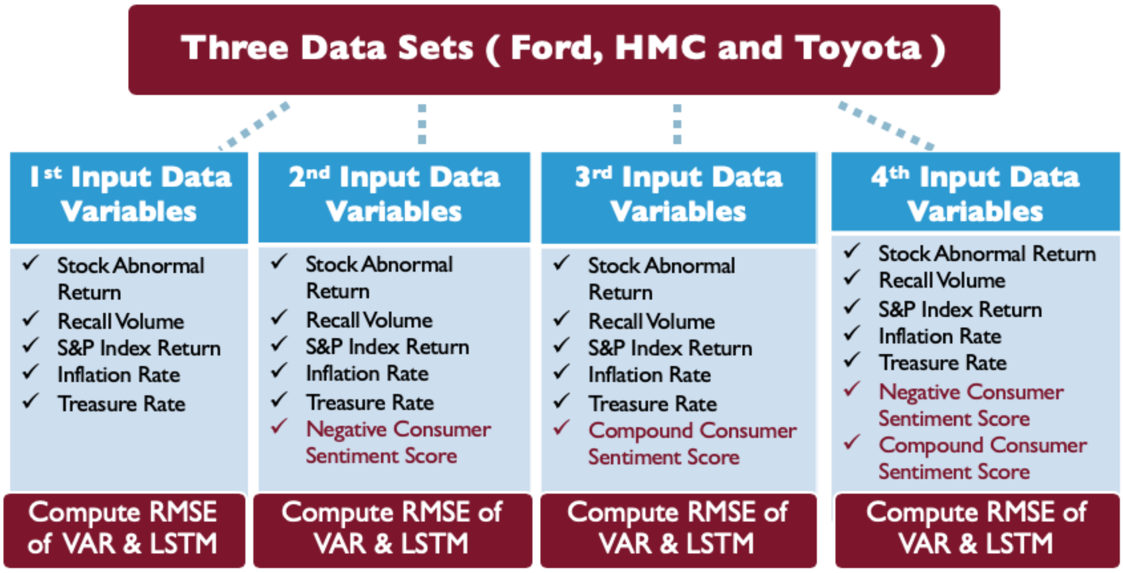


Figure 6.2 Four Input Data with Different Variable Combinations

6.3.1 VAR Forecasting Process:

(1) Conduct exploratory data analysis and visualize each time series with the line chart, Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots.

(2) Test random walk for each time series with an Augmented Dickey-Fuller test (ADF) and find that abnormal company stock return and S&P index are stationary time series, but product recall volume, negative sentiment score,

compound sentiment score, treasury rate and inflation rate are all non-stationary time series.

(3) Since VAR requires each time series to satisfy the stationary assumption, some time series need to be transformed and tested with ADF. We transform product recall volume with min and max scaling method, calculate percentage change of negative sentiment score and compound sentiment score with percentage change function in Python, and take the first difference of both inflation rate and treasury rate. After transformation, all variables in data sets are stationary.

(4) Prepare three data sets for forecasting monthly abnormal stock return of three different automobile manufacturers. In each data set, split 80% of first consecutive observations into the training set and set up the remaining 20% observations as the testing set.

(5) Use the statsmodels library with Python and fit input data with the VAR model from this library. Select order of VAR based on Akaike information criterion AIC.

(6) Perform the rolling forecast approach with each VAR model and compute the overall RMSE by comparing one-step forecasting values with actual values in the testing set.

6.3.2 LSTM Forecasting Process:

(1) Conduct exploratory data analysis with pandas, numpy and matplotlib libraries in Python. Scale all the features or variables with Min & Max scaler in sklearn library and set up random seed as 0 with numpy library in Python.

(2) Since deep learning LSTM is designed to implement supervised learning problems in machine learning, we need to reframe three different data set. For example, we have seven variables in figure 6.2: var1, var2 … and var7 in one data set. Since we would like to conduct one-step forecast for var1, a new column is generated by shifting var1 back 1 timestamp. In each row, there are var1(t-1), var2(t-1) … and var7(t-1) at time T-1, and var1(t) at time T. The new

column works as the target variable, and we use historical information to predict the future value of abnormal return.

	var1(t-1)	var2(t-1)	var3(t-1)	var4(t-1)	var5(t-1)	var6(t-1)
1	0.136348	0.477856	0.690141	0.874459	0.032820	0.980515
2	0.163175	0.473447	0.740845	0.896104	0.019225	0.995229
3	0.153160	0.437033	0.660094	0.865801	0.000000	1.000000
4	0.196921	0.457825	0.622535	0.774892	0.035156	0.983458
5	0.307238	0.489581	0.611268	0.757576	0.091927	0.892189

	var7(t-1)	var1(t)
1	0.163870	0.163175
2	0.000016	0.153160
3	0.000000	0.196921
4	0.036889	0.307238
5	0.071883	0.248767

Figure 6.3 Reframing Time Series Data for Supervised Learning

(3) Similar to splitting three data sets in VAR forecasting, we split 80% of first consecutive observations in one data set as training set and the remaining 20% observations as the testing set. The training set is used to estimate parameters and testing set works for LSTM forecasting validation. In machine learning, the training set is saved into two pandas dataframes: training-X and training-Y. Training-X contains the data of all features or predictive variables, while training-Y is the data of the target variable(s). Also, Testing set likewise include testing-X and testing-Y.

(4) Based on Ford, HMC and Toyota data sets, we conduct one-step forecast for abnormal return with four different types of variable combinations to study if including sentiment related data can improve predictive accuracy.

(5) Python is used to implement the LSTM algorithm along with Keras, an open source and TenserFlow, an open source machine learning framework. We construct neural networks with Keras and TensorFlow. One LSTM layer with 256 nodes is included. The loss function is “mean squared error” and optimizer is “adam”.

(6) Reshape all the training sets and testing sets so that data sets can be fit and analyzed on Keras neural networks. We fit training-X and training-Y into the neural networks and set up epochs equal to 50.

(7) Conduct rolling foresting to estimate one-step predictive values of abnormal return, compare them with corresponding actual values in testing-Y, and compute overall predictive error rate based on RMSE formula.

7. Results

In this section, the Mean Root-Mean-Square Error (RMSE) values of all VAR and LSTM models trained with different input data in three data sets are presented. We compare and discuss the forecasting performance of VAR and LSTM models trained with different data by comparing RMSE values. Lower RMSE value derived from smaller difference between actual values and forecast values and means a better predictive performance of related model. RMSE reduction rates are also calculated by comparing the percentage change of one RMSE value of predictive model with another RMSE of another predictive model. A higher RMSE reduction rate shows a larger improvement in terms of the forecasting performance.

Table 7.1 Mean Root-Mean-Square Error of VAR and LSTM Models

	Input Data1 Mean RMSE	Input Data2 Mean RMSE	Input Data3 Mean RMSE	Input Data4 Mean RMSE
Ford-VAR	0.241	0.145	0.153	0.153
Ford-LSTM	0.063	0.063	0.064	0.059
HMC-VAR	0.072	0.073	0.068	0.075
HMC-LSTM	0.042	0.046	0.041	0.045
Toyota-VAR	0.229	0.239	0.132	0.105
Toyota-LSTM	0.037	0.039	0.039	0.040

Table 7.2 VAR and LSTM Predictive Performance

	Lowest VAR Mean RMSE	Lowest LSTM Mean RMSE	LSTM Reduction Rate

Ford Data Set	0.145	0.059	59.31%
HMC Data Set	0.068	0.041	39.70%
Toyota Data set	0.105	0.037	64.76%

Table 7.3 LSTM-VAR RMSE Reduction Rate for Input Data

	Input Data1 RMSE Reduction Rate	Input Data2 RMSE Reduction Rate	Input Data3 RMSE Reduction Rate	Input Data4 RMSE Reduction Rate
Ford Dataset LSTM-VAR	73.86%	56.55%	58.17%	61.44%
HMC Dataset LSTM-VAR	41.67%	36.99%	39.71%	40.00%
Toyota Dataset LSTM-VAR	83.84%	83.68%	70.45%	61.90%

The forecasting results and comparison of predictive performance are reported in Table 7.1, Table 7.2 and Table 7.3. The initial training set has 53 monthly records. We conduct one-step rolling forecast for future 13 monthly abnormal returns. We first compare RMSE scores generated from both LSTM and VAR with same input data from same data set. For example, we compare the first pair of RMSE scores in the circle shown in Table 7.1. The RMSE score generated by fitting first input data from Ford data set into LSTM is 0.063 which is 73.86% lower than the RMSE of the VAR model estimated with same input data. The result is shown in Table 7.3 suggests that LSTM consistently outperforms VAR with different input data across three data sets. For each input data in different data set, LSTM-based models achieve a reduction rate in RMSE between 36.99% - 83.84%, by comparing with VAR-based models.

To evaluate the overall forecasting performance of LSTM and VAR in three data sets, we compare the lowest RMSE scores of LSTM and VAR estimated with same data set. In Ford data set, the lowest average Root-Mean-Square Error (RMSE) using rolling VAR and rolling LSTM are 0.145 and 0.059 respectively, on average yielding a 59.31% reduction in predictive error rates achieved by LSTM. As for HMC data set, LSTM also makes an average of 39.70% reduction in predictive error rates, and the lowest mean RMSE values generated from VAR and LSTM are 0.068 and 0.041, respectively. When comparing VAR and LSTM in the Toyota data set, the lowest mean RMSE values of VAR and LSTM methods are 0.105 and 0.037, respectively; therefore, the LSTM-based model could achieve 64.76% reduction in RMSE. On average, LSTM-based models on average have 59.02% predictive error rate reductions in this research.

Table 7.4 VAR & LSTM RMSE Reduction after Adding Sentiment Scores

	VAR RMSE Reduction Value	VAR RMSE Reduction Rate	LSTM RMSE Reduction Value	LSTM RMSE Reduction Rate
Ford Data Set	0.096	39.83%	0.004	6.35%
HMC Data Set	0.004	5.56%	0.001	2.38%
Toyota Data set	0.124	54.15%	-0.002	-5.41%

Next, we investigate the change of RMSE values after adding sentiment scores in the same predictive models. For example, we use the RMSE value of VAR model estimated by the first input data of Ford data set to make comparison with the RMSE values of VAR models that are fitted with the 2nd, 3rd and 4th input data of Ford data set respectively. Here the 2nd, 3rd and 4th are the input data that contains different combinations of sentiment variables. As Table 7.1 shown, fitting 2nd input of Ford data set into VAR produces a lowest RMSE error score (0.145), therefore the RMSE reduction rate of VAR-Ford is 38.93% that is calculated by 0.241 and 0.145. Following this procedure, we calculate the predictive improvements or the reductions of RMSE error of all VAR and LSTM estimated with different input data in three data sets. VAR in Ford data set achieves 39.83% reduction of mean RMSE after adding negative sentiment score, and LSTM reduces 6.35% in mean RMSE by including both negative and compound sentiment scores. VAR make 5.56% reduction of mean RMSE in HMC data set, and LSTM acquires 2.38% reduction in the value of mean RMSE. In the Toyota data set, LSTM makes no progress in predictive performance with sentiment variables, while VAR models reduce mean RMSE value by 54.15%. Overall, VAR-based models make more improvement in predictive performance after adding consumer sentiment. There is no clear improvement made by LSTM with sentiment data, but LSTM-based models still have much lower RMSE values compared with RMES values of VAR, which is more than 50% reduction of the RMSE values of VAR-based models.

8. Discussion and Implications

8.1 Research Implications

Prior studies suggest that the product recall volume can hurt investors' confidence and have an impact on the financial value of a company (Chen, Ganesan, & Liu 2009; Liu, Shankar, & Yun 2017). Researchers also find the level of consumer complaint influences the company market value in a direct linear way (Luo, 2007; Luo 2009) or a non-linear way (Claro et al. 2014). In this study, we look into product-harm events in the automobile industry and specifically focus on predicting the impacts of both product recall volume and consumer sentiment in negative online complaints. Product recall has been studied by many researchers (Kalaignanam, Kushwaha & Eilert 2013; Liu, Shankar, & Yun 2017; Swartz et al. 2017), but a close study on predictive modeling the effect of the product recall on firm value is very scarce. Different from prior studies, this research compares two multivariate time series forecasting methods: VAR and LSTM, and evaluate their predictive performances in a context of product-harm events. Consumer sentiment from product-harm complaints is further studied and included in predictive models. As the results suggest, large improvements of forecasting error reduction have been achieved in VAR-based models with consumer sentiments, therefore this study help lay a foundation for future research of product recall and consumer sentiment, and develop more theories of consumer sentiment in product-harm events. This research also makes two contributions to product-harm events literature in marketing research. First, we implement and compare two typical time series forecasting methods from the statistical modeling and machine learning domains. Advantages and disadvantages of applying VAR and LSTM in product-harm events are discussed. Second, we use lexicon and rule-based sentiment analysis tool --- VADER to extract consumer sentiments in product-harm complaints. By including sentiment variables in predictive models, we aim to examine if including sentiment related data will improve the forecasting performance of VAR-based models and LSTM-based models. Our study could encourage more researchers leveraging efficient sentiment tools and extracting consumer sentiment from online word of mouth.

8.2 Methodological Comparison: VAR VS LSTM

With the one-step rolling forecast approach, we predict the next-month abnormal return of company stock. The forecasting results show LSTM-based models always outperforms VAR-based models in terms of predictive accuracy across all three data sets with different input data. Based on corresponding RMSE values, LSTM-based models have RMSE reduction rates that are between 36.99% and 83.84% compared with VAR-based models. On average, LSTM-based models can achieve 59.02% predictive error rate reductions. Combining the knowledge from a prior study (Yau 2017) with our implementations of VAR and LSTM, we summarize several differences between VAR and LSTM time series forecasting methods. First, VAR-based models are applied to stationary time series only, while LSTM doesn't have this stationarity assumption requirement. When time series is not stationary, researchers need to transform nonstationary series into stationary series and then fit the data into VAR. Second, VAR-based models are constructed based on a linear system of equations, but LSTM-based models have a much more complex structure that involves many non-linear transformations. Statistical modeling has a long history of explaining coefficients of linear structural models, but interpreting deep learning neural network is a new research subject and still needs more well-articulated theories to help researchers or managers understand the black box in Machine Learning algorithms. Third, compared with LSTM-based model, VAR-based model implementation is relatively straightforward, and its parameters are relatively easy to set up. When implementing LSTM forecasting with Keras, however, data preprocessing such as min & max scaling is required. In addition, we need to set up many parameters in LSTM forecasting for finalizing the neural network architecture design. Parameters such as the number of layers, the number of nodes, batch size, loss function, activation function, learning rate, optimizer algorithm, etc., could be tuned to optimize forecasting performance of one neural network, but finding the optimal parameters is difficult. Also, since LSTM contains a more complex structure and more parameters to estimate, training one VAR-based model usually take less time than training one LSTM-based model. When facing real-world business problems, it's important for managers to understand the trade-off between interpretability and predictability of predictive models, and also evaluate both efficiency and accuracy of implementing modeling.

By including consumer sentiment data in predictive models, we also calculate RMSE reduction rates achieved by VAR-based models and LSTM-based models. The result shows that VAR could greatly improve predictive performance with consumer sentiment data, while there is no great progress shown in LSTM. VAR-based models and LSTM-based models on average decrease RMSE values by 33.18% and 1.11%, respectively. Even though VAR made more significant progress after adding consumer sentiment in forecasting, the RMSE values generated by LSTM-based models are still far lower than the RMSE values from corresponding VAR-based models, on average achieving more than 50% reduction in RMSE values of VAR forecasting. One explanation is that adding consumer sentiment in the simple linear structure of VAR-based models could help VAR capture more information and make more accurate predictions for future abnormal stock return. As for LSTM-based models, we notice the complex non-linear structure of LSTM already performs very good forecasting without adding consumer sentiment and achieves relatively low RMSE values. After adding consumer sentiment, some of the LSTM-based models slightly improve predictive performance, but there is no significant improvement. In automobile product-harm events, we thus conclude that considering consumer sentiment from negative complaint can help VAR improve forecasting accuracy, but LSTM can perform much better forecasting even without adding any consumer sentiment data.

8.3 Managerial Implications

Understanding how to accurately predict the impacts of product recall and negative eWOM on company abnormal stock return is essential to managers since they need to make strategies, policies or remedies based on the predictive feedback from the market especially in the context of product-harm crisis. Our empirical study supports the view that deep learning LSTM outperforms traditional VAR in terms of the predictive ability of company market value and encourages managers to adopts more machine learning techniques in the big data era. Successfully utilizing machine learning in product-harm crisis could help managers efficiently understand the severity of potential issues in advance and make some efforts to reduce the influences of negative events such as product recalls and consumer complaints.

This research also explores lexicon and rule-based sentiment analysis tool named VADER that is attuned explicitly to sentiments expressed in microblog-like contexts and outperforms individual human rater in terms of accessing emotions (Hutto & Gilbert 2014). Using this sentiment analysis tool, marketers are able to efficiently monitor and quantify consumer sentiment in eWOM without human intervention in real time. The results of our study indicate that leveraging consumer sentiment measurement in predictive modeling such as VAR will improve forecasting performance. It's important for managers to leverage consumer sentiment from word of mouth in order to foresee the impacts of negative events and thus implement remedial actions. The better forecasting performance the predictive model can provide with consumer sentiment, the more likely marketers are able to identify potential opportunities, manage consumer satisfaction level and reduce the dissemination of negative viral information (Godes & Mayzlin 2004).

9. Limitation and Future Research

This study has three limitations that offer indications for future research. First, we mainly focus on the U.S. automobile industry because of its availability of consumer complaints in product-harm events and the high frequency of product recalls. The generalizability of our results needs to be further verified with different data sets from other sectors or other countries. Second, this study aggregates monthly consumer sentiment and product recall data from Jan-2005 to July-2010 on NHTSA for predicting the monthly abnormal stock return of one company. More extensive data sets, for example, daily observations or data collected from more than 10-year period of time could be analyzed. Third, consumer sentiment is only examined from one source in this study: consumers 'complaints posted on NHTSA.

As for further research, the first interesting direction is to study more extensive data by collecting daily records within a longer period from different industries that involve product-harm events. With more granular observations, we could set up a time window of each product recall events, collect the data of relevant variables only within this time window, and exclude noise from other important events. Future research could also investigative the impacts of product recall and consumer sentiment of one company on different companies such as competitors in the same industry and the suppliers from other industries. In order to make an intuitive comparison with VAR, we construct one specific LSTM-based neural network that only contains one hidden layer with 256 nodes in this study. The loss function and optimizer of the neural network are set up as “mean squared error” and “adam”, respectively. Another future research could extend to forecasting optimization by investigating other neural network architectures with more hidden layers and a more complex internal structure. At the same time, understanding the estimated results of complex internal structure or the black box within deep learning LSTM neural networks is challenging compared with VAR models. To have a more profound impact on managerial decision making, researchers and practitioners could explore more interpretation methods for machine learning algorithm models and compare the results with the interpretation generated from traditional statistical models. Besides, it would also be interesting to look into data from other social networks such as online automobile forums, related Facebook groups and tweets in Twitter based on keyword search. Trying different consumer sentiment extraction methods on the text from these online platforms. One exciting direction is predicting the sentiment of consumer online complaints with machine learning supervised learning algorithms. Having human raters to label some complaints as a training data set, researchers could then build up a predictive model to predict the sentiment score of unlabeled consumer complaint by applying machine learning algorithms that are designed for mining text data. Last but not least, it is important to take ethical issues into considerations when analyzing a large number of consumer’s opinions, reviews or complaints in future research. As the internet of things (IOT) is becoming the backbone of consumer value, consumers’ identifiable information is much easier to collect than before, therefore protecting consumers' privacy rights in the context of big data analysis turns into one essential subject in marketing research.

REFERENCES

Adebiyi, A.A., Adewumi, A.O., Ayo, C.K. (2014). Stock Price Prediction Using the ARIMA Model. UKSim-AMSS 16th International Conference on Computer Modelling and Simulation.

Ahluwalia, R. (2000). Examination of psychological processes underlying resistance to persuasion. *Journal of Consumer Research*, 27, 217–232.

Ahluwalia, R. (2002). How prevalent is the negativity effect in consumer environments? *Journal of Consumer Research*, 29, 270–279.

Alonso, A.M., Garc´ıa-Martos, C. (2012). Time Series Analysis - Forecasting with ARIMA models. Universidad Carlos III de Madrid, Universidad Polit´ecnica de Madrid.

Arndt, J. (1976) Role of Product-Related Conversations in the Diffusion of a New Product. *Journal of Marketing Research,* 1, 4: 291–295.

Banerjee, A., Marcellino, M., Masten, I. (2005). Forecasting Macroeconomic Variables for the New Member States of the European Union. European Central Bank. Working paper Series 482.

Barberis, N., Schliefer, A., & Vishny, R. (1998). A Model of Investor Sentiment. *Journal of Financial Economics*, 49 (3), 307–43.

Bentivegna, F. J. (2002). Fatores de impacto no sucesso do marketing boca a boca online. *Revista de Administração de Empresas*, 42(1), 79-87.

Berman, B. (1999). Planning for the Inevitable Product Recall. *Business Horizons*, 42 (2), 69–78.

Borah, A. & Tellis, G. J. (2016). Halo (Spillover) Effects in Social Media: Do Product Recalls of One Brand Hurt or Help Rival Brands? *Journal of Marketing Research*, April 2016, Vol. 53, No. 2, pp. 143-160.

Breiman, L. (2001), "*Statistical Modeling: The Two Cultures, Statistical Science*", Vol.

16, No. 3, pg.199-215

Brownlee, J., 2016. Time Series Prediction with LSTM Recurrent Neural Networks in Python with Keras.
https://machinelearningmastery.com/time-series-prediction-lstm-recurrent-neural-networks-python-keras/

Brown, J. J., & Reingen, P. (1987). Social ties and word-of-mouth referral behavior. *Journal of Consumer Research*, 14(3), 350-362.

Bromiley, P., & Marcus, A. (1989). The Deterrent to Dubious Corporate Behavior: Profitability, Probability, and Safety Recalls. *Strategic Management Journal*, 10 (3), 233–50.

Brown, S. W., Webster, F. E., Jr., Steenkamp, J.-B. E. M., Wilkie, W. L., Sheth, J. N., Sisodia, R., S., Kerin, R. A., MacInnis, D., McAlister, L., Raju, J. S., et al. (2005). Marketing renaissance: opportunities and imperatives for improving marketing thought, practice, and infrastructure. *Journal of Marketing*, 69(4), 1-25. doi: 10.1509/jmkg.2005.69.4.1

Chan, W.S. (2003). Stock Price Reaction to News and No-News: Drift and Reversal After Headlines. *Journal of Financial Economics*, 70 (2), 223–60.

Chen, Y., Ganesan, S., &, Liu, Y. (2009). Does a Firm's Product-Recall Strategy Affect Its Financial Value? *Journal of Marketing,* Vol. 73 (November 2009), 214–226.

Chevalier, J. A. & Mayzlin, D. (2006). The Effect of Word of Mouth on Sales: Online Book Reviews. *Journal of Marketing Research*, 43 (August), 345–54.

Choudhury, T., Kumar, V., & Nigam, D. (2015). An Innovative and Automatic Lung and Oral Cancer Classification Using Soft Computing Techniques. *International Journal of Computer Science & Mobile Computing*, Vol.4 Issue.12, pg. 313-323.

Claro, D. P., Fragoso, A. F. G. R., Neto, S. A. L., & Claro, P. B. O. (2014). Consumer Complaints and Company Market Value. *BAR, Rio de Janeiro*, v. 11, n. 3, art. 1, pp. 248-263, July/Sept. 2014

Dawar, N. & Pillutla, M. M. (2000). Impact of Product-Harm Crises on Brand Equity: The Moderating Role of Consumer Expectations. *Journal of Marketing Research*, 37 (May), 215–26.

De Matos, C. A., & Rossi, C. A.V. (2008). Word-of-mouth communications in marketing: a meta-analytic review of the antecedents and moderators. *Journal of the Academy of Marketing Science*, 36, 578–596.

Del Negro, M. & Schorfheide, F. (2009). Bayesian Macroeconometrics, *Handbook of Bayesian Econometrics.*

Ding X., Zhang, Y., Liu, T., & Duan, J. (2015). Deep Learning for Event-Driven Stock Prediction. *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence* (IJCAI): 2327-2333.

Duhan, D. F., Johnson, S. D., Wilcox, J. B., & Harrell, G. D. (1997). Influences on consumer use of word-of-mouth recommendation sources. *Journal of the Academy of Marketing Science*, 25, 283–295.

Eilert, M., Jayachandran, S., Kalaignanam, K., & Swartz, T. A. (2017). Does It Pay to Recall Your Product Early? An Empirical Investigation in the Automobile Industry. *Journal of Marketing, Vol. 81 (May 2017), 111–129.*

Gamboa, J. (2017). Deep Learning for Time-Series Analysis. University of Kaiserslautern, Kaiserslautern, Germany.

Gers, F.A., Schmidhuber, J., & Cummins, F. (2000). Learning to Forget: Continual Prediction with LSTM. *Neural Computation* 12(10): 2451-2471.

Géron, A. (2017). Hands-On Machine Learning with Scikit-Learn and TensorFlow.

Godes, D., & Mayzlin, D. (2004). Using online conversations to study word-of-mouth communication. *Marketing Science*, 23(4), 545-560.

Govindaraj, S., Jaggi, B., & Lin, B. (2004). Market Overreaction to Product Recall Revisited: The Case of Firestone Tires and the Ford Explorer. *Review of Quantitative Finance and Accounting*, 23 (1), 31–54.

Graves, A., Jaitly, N., & Mohamed, A. (2013). Hybrid Speech Recognition with Deep Bidirectional LSTM. IEEE Workshop on In Automatic Speech Recognition and Understanding (ASRU). pp. 273-278.

Graves A., Liwicki M., Fernandez S., Bertolami R., & Bunke H., Schmidhuber J. (2009). A Novel Connectionist System for Improved Unconstrained Handwriting Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 31(5).

Guttag, J. V. (2013), *Introduction to Computation and Programming Using Python*. The MIT Press 2013.

Harald, V. H., Helsen, K., & Dekimpe, M. G. (2007). The Impact of a Product-Harm Crisis on Marketing Effectiveness. *Marketing Science*, 26 (2), 230–45.

Hennig-Thurau, T., Gwinner, K. P., Walsh, G., & Gremler, D. D. (2004). Electronic word of mouth via consumer opinion platforms: What motivates consumers to articulate themselves on the Internet? *Journal of Interactive Marketing*, 18, 38e52.

Herr, P. M., Kardes, F. R., & Kim, J. (1991). Effects of word-of-mouth and product-attribute information on persuasion: An accessibility diagnosticity perspective. J*ournal of Consumer Research,* 17(4), 454–462.

Hochreiter, S. & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation* 9(8):1735-1780.

Huck, N. (2009). Pairs Selection and Outranking: An Application to the S&P 100 Index. *European Journal of Operational Research* 196(2): 819-825.

Hutto, C. J. & Gilbert, E. (2014). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. Association for the Advancement of Artificial Intelligence.

Hyndman, R.J. & Athanasopoulos, G. (2014). Forecasting: Principles and Practice. Retrieved from https://robjhyndman.com/uwafiles/fpp-notes.pdf

I. Melnyk, B. Matthews, A. Banerjee, & N. Oza (2016). Semi-Markov switching vector autoregressive model based anomaly detection in aviation systems. *Knowledge Discovery and Data Mining*.

Jordan, M. I. (2014). Statistics and machine learning. Reddit, Retrieved 2014-10-01. https://www.reddit.com/r/MachineLearning/comments/2fxi6v/ama_michael_i_jordan/

Joshi K., Bharathi, H.N., & Rao, J. (2013). Stock Trend Prediction using News Sentiment Analysis. https://arxiv.org/ftp/arxiv/papers/1607/1607.01958.pdf

Kalaignanam K., Kushwaha T., & Eilert M. (2013) The Impact of Product Recalls on Future Product Reliability and Future Accidents: Evidence from the Automobile Industry. *Journal of Marketing*: March 2013, Vol. 77, No. 2, pp. 41-57.

Khashei M. & Bijari M. (2011). A Novel Hybridization of Artificial Neural Networks and ARIMA Models for Time Series forecasting. *Applied Soft Computing,* 11(2): 2664-2675.

Kohzadi, N., Boyd, M. S., Kermanshahi, B., & Kaastra, I. (1996) A Comparison of Artificial Neural Network and Time Series Models for Forecasting Commodity Prices. *International Journal of Neuro-computing* 10: 169-181.

Kirmani, A., Sood, S., & Bridges, S. (1999). The ownership effect in responses to brand line stretches. *Journal of Marketing*, 63(1), 88–101.

Kumar, V., Choi, J.B., & Greene, M. (2016). Synergistic effects of social media and traditional marketing on brand sales: capturing the time varying effects. *Journal of the Academy of Marketing Science*.

Kübler, R. V., Wieringa, J. E., & Pauwels, K. H. (2017), *Advanced Methods for Modeling*, Chapter 19, Machine Learning and Big Data.

Kwak, H., Lee, C., Park, H., & Moon, S. (2010), "What Is Twitter, a Social Network or a News Media?" in Proceedings of the 19th International Conference on World Wide Web. *New York: Association for Computing Machinery,* 591–600.

Laufer, D. & Coombs, W. T. (2006). How Should a Company Respond to a Product Harm Crisis? The Role of Corporate Reputation and Consumer-Based Cues. *Business Horizons*, 49 (5), 379–85.

Lee, S. I. & Yoo, S. J. (2017). A Deep Efficient Frontier Method for Optimal Investments. Department of Computer Engineering, Sejong University, Seoul, 05006, Republic of Korea.

Liu Y. & Shankar V. (2015). The Dynamic Impact of Product-Harm Crises on Brand Preference and Advertising Effectiveness: An Empirical Analysis of the Automobile Industry. *Management Science*, Vol. 61, No. 10, October 2015, pp. 2514–2535.

Liu Y., Shankar V., & Yun W. (2017). Crisis Management Strategies and the Long-Term Effects of Product Recalls on Firm Value. *Journal of Marketing:* September 2017, Vol. 81, No. 5, pp. 30-48.

Luo, X. (2007). Consumer negative voice and firm- idiosyncratic stock returns. *J*ournal of Marketing,71(3), 75–88. doi: 10.1509/jmkg.71.3.75.

Luo, X. (2009). Quantifying the long-term impact of negative word of mouth on cash flows and stock prices. *Marketing Science*, 28(1), 148-165. doi: 10.1287/mksc.1080.0389.

Mary, S. (1990). Measuring Image Spillovers in Umbrella- Branded Products. *Journal of Business*, 63 (3), 309–329.

Mintel (2015), Seven in 10 Americans seek out opinions before making purchases. http://www.mintel.com/press-centre/social-and-lifestyle/seven-in-10-americans-seek-out-opinions-before-making-purchases

Naylor, G. S. (2016). *Journal of Consumer Satisfaction Dissatisfaction & Complaining Behavior,* Volume 29, 2016 | 131.

Neural Network architectures. (2019).
Retrieved from URL (http://cs231n.github.io/neural-networks-1/#nn)

Rhee, M. & Haunschild, P. R. (2006). The Liability of Good Reputation. *Organization Science*, 17 (1), 101–117.

Richins, M. L. (1983). Negative word-of-mouth by dissatisfied consumers: a pilot study. *Journal of Marketing*, 47(1), 68-78.

Roehm, M. L. & Brady, M. K. (2007). Consumer Responses to Performance Failures by High-Equity Brands. *Journal of Consumer Research*, 34(December), 537–545.

Roondiwala, M., Patel, H., & Varma, S. (2017). Predicting Stock Prices Using LSTM. *International Journal of Science and Research (IJSR)* 6(4).

Sak, H., Senior, A., & Beaufays, F. (2014), Long Short-Term Memory Recurrent Neural Network Architectures for Large Scale Acoustic Modeling.

Santos, C. P. dos, & Fernandes, D. V. D. H. (2011). Perceptions of justice after recovery efforts in internet purchasing: the impact on consumer trust and loyalty toward retailing sites and online shopping in general. *Brazilian Administration Review*, 8(3), 225-246. Retrieved from http://www.scielo.br/pdf/bar/v8n3/a02v8n3.pdf.
doi: 10.1590/S1807-76922011000300002

Schmidhuber J. (2015). Deep learning in neural networks: An overview. *Neural Networks* 61: 85-117.

Siami-Namini, S. & Namin, A. S. (2018). Forecasting Economics and Financial Time Series: ARIMA vs. LSTM.

Singh, J. (1990). Voice, exit, and negative word-of-mouth behaviors: an investigation across three service categories. *Journal of the Academy of Marketing Science*, 18, 1–15.

Statista. Market share held by selected automobile manufacturers in the U.S. in 2017. https://www.statista.com/statistics/249375/us-market-share-of-selected-automobile-manufacturers/

Taylor, S. J. (2007). Modelling financial time series. *World Scientific Publishing*.

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

Van Laer, T. and De Ruyter, K. (2012). In stories we trust: how narrative apologies provide cover for competitive vulnerability after integrity-violating blog posts. *International Journal of Research in Marketing*, 27(2), 164–174. doi: 10.1016/j.ijresmar.2009.12.010

Verhagen, T., Nauta, A., & Feldberg, F. (2013). Negative online word-ofmouth: Behavioral indicator or emotional release? *Computers in Human Behavior*, 29, 1430–1440.

Wasserman, L. (2012). Statistics and Machine Learning, accessed online June, 5 2016 at https://normaldeviate.wordpress.com/2012/06/12/statistics-versus-machine-learning-52

Wilson, A. E., Giebelhausen, M. D., & Brad, M. K. (2017). Negative word of mouth can be a positive for consumers connected to the brand. *Academy of Marketing Science*, 45:534–547.

Winchester, M., Romaniuk, J., & Bogomolova, S. (2008). Positive and negative brand beliefs and brand defection/uptake. *European Journal of Marketing*, 42(5/6), 553-570. doi: 10.1108/03090560810862507.

Xiong R., Nichols, E.P., &, Shen, Y. (2015). Deep Learning Stock Volatility with Domestic Trends. https://arxiv.org/abs/1512.04916

Yau, J. (2017). Time Series Forecasting Using Recurrent Neural Network and Vector Autoregressive Model: When and How.
https://www.youtube.com/watch?v=_vQ0W_qXMxk

Yim J. (2002). A Comparison of Neural Networks with Time Series Models for Forecasting Returns on Stock Market.
Index. Springer-Verlag Berlin Heidelberg, IEA/AIE, LNAI 2358: 25-35.

Zaremba, W., Sutskever, I., &, Vinyals, O. (2015). Recurrent Neural Network Regularization. Under review as a conference paper at ICLR 2015

Zhang, P.G. & Qi, M. (2005). Neural Network Forecasting for Seasonal and Trend Time Series, *European Journal of Operational Research* 160: 501-514.

Zhu, X. (2015). Blog: Understand LSTM Neural Networks. https://www.jianshu.com/p/9dc9f41f0b29