Price Prediction: Determining Changes in Stock Pricing through Sentiment Analysis of Online Consumer Reviews

Daryl F. Boykin

University of Nevada, Las Vegas, dfboykin@gmail.com
PRICE PREDICTION: DETERMINING CHANGES IN STOCK PRICING THROUGH SENTIMENT ANALYSIS OF ONLINE CONSUMER REVIEWS

by

Daryl F. Boykin

Bachelor of Arts - Gaming Management
University of Nevada, Las Vegas
2012

Master of Science - Hotel Administration
University of Nevada, Las Vegas
2014

A dissertation submitted in partial fulfillment of the requirements for the
Doctor of Philosophy - Hospitality Administration

Hospitality Administration
William F. Harrah College of Hotel Administration
The Graduate College

University of Nevada, Las Vegas
May 2017
Dissertation Approval

The Graduate College
The University of Nevada, Las Vegas

April 4, 2017

This dissertation prepared by

Daryl F. Boykin

entitled

Price Prediction: Determining Changes in Stock Pricing through Sentiment Analysis of Online Consumer Reviews

is approved in partial fulfillment of the requirements for the degree of

Doctor of Philosophy - Hospitality Administration
Hospitality Administration

Ashok K. Singh, Ph.D.
Examination Committee Chair

Seyhmus Baloglu, Ph.D.
Examination Committee Member

Toni Repetti, Ph.D.
Examination Committee Member

Rohan J. Dalpatadu, Ph.D.
Graduate College Faculty Representative

Kathryn Hausbeck Korgan, Ph.D.
Graduate College Interim Dean
Abstract

The rapid growth of technology has changed the dynamics in which consumers socialize and make their purchasing decisions. The volume of online reviews has grown rapidly over the past decade, leading the peer groups of consumer to carry a disproportionate weight in the purchasing decision process. The sheer volume of reviews can be a daunting task for an operator to attempt to incorporate the reviews in their analysis. Sentiment analysis allows for large volumes of consumer reviews to be processed in a relatively easy, and time sensitive manner. The information contained in these reviews, the sentiment score, is the same feeling hospitality consumers are gathering from other consumers prior to making their purchasing decision. To demonstrate the importance of these reviews, this study will seek to model the directional change of a company’s stock price using the sentiment of the consumer’s reviews as the primary predictor. Support Vector Machines will help to classify a year’s worth of consumer reviews on nine distinct properties of a publicly traded Las Vegas gaming/hotel company. This is then modeled using ARIMA modelling techniques to forecast an out-of-time sample, and the accuracy will be assessed by showing that the results being due to random change are minimal. The model is able to accurately predict 28 out of 39 time periods in the out of time sample, which has less than a .0047 probability of being due to random chance.

Key Words: sentiment analysis, consumer socialization, ARIMA modelling, forecasting, revenue prediction, stock price prediction, consumer reviews, eWOM, online reviews
Acknowledgments

I would like to express my deepest gratitude and sincere appreciation to my committee chair Dr. Ashok Singh. His friendship and mentorship has guided me through this doctoral process. He has continued to convey a sense of wonder and exploration that drove this research, and instills a high level of academic standards. Without his guidance, this dissertation would not have been possible.

I would also like to thank my committee members, Dr. Rohan Dalpatadu and Dr. Sheymus Baloglu. Their commitment to excellence represents the best of the University of Nevada- Las Vegas, and helped me to fully develop the theories and ideas of this research.

In addition, I would like to thank Dr. Bo Bernhard and Dr. Tony Lucas. Dr. Bernhard’s teaching of “trying on different lenses” helped to light the fire of my interest in sentiment analysis, and continues to drive me to makes sure we are measuring the right things. Dr. Lucas’ introduction into time series as an undergrad, willing to explain a complicated statistical procedure and encouraging me to seek my advanced degrees is what fueled my interest in analytics.

Dr. James Busser, Dr. Mehmet Erdham, Dr. Sarah Tanford, Prof. William Werner, Dr. Billy Bai, Dr. Michael Dalbour, and Dr. Gail Sammons have all, in their own way, contributed to my enrichment as an academic and as a person. I would be reminiscent if I neglected to include them in my gratitude. Finally, I would like to thank ReviewTrackers for providing access to the data used in this research.
Dedication

This dissertation is dedicated to my loving and supporting family. My wife, Elizabeth Boykin, and son, Alexander Boykin, without their sacrifices, support, and understanding I would never have been able to complete this dissertation. They believed in me when I didn’t, had strength for me when I faltered, and rejoiced with my successes. Together in All Things.
# Table of Contents

Abstract ........................................................................................................................................ iii
Acknowledgments ........................................................................................................................... iv
Dedication ....................................................................................................................................... v
Table of Contents ........................................................................................................................ vi
List of Tables ................................................................................................................................... x
List of Figures ................................................................................................................................ xi
List of Equations .......................................................................................................................... xii
Chapter 1 ....................................................................................................................................... 1
  Introduction ................................................................................................................................... 1
  Problem Statement ....................................................................................................................... 3
  Purpose of the Study .................................................................................................................... 4
  Research Questions ...................................................................................................................... 4
  Delimitations ............................................................................................................................... 5
  Significance of the Study .............................................................................................................. 5
  Definitions ................................................................................................................................... 6
  Conclusion ................................................................................................................................... 7
Chapter 2 ....................................................................................................................................... 8
  Literature Review ......................................................................................................................... 8
  Consumer Socialization Theory ................................................................................................. 8
  Social Media and eWOM ............................................................................................................ 12
  Consumer Reviews .................................................................................................................... 13
  Consumer Reviews’ Link to Firm Performance ........................................................................ 15
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Granularity</td>
<td>41</td>
</tr>
<tr>
<td>Sentiment Identification</td>
<td>42</td>
</tr>
<tr>
<td>Data Processing</td>
<td>42</td>
</tr>
<tr>
<td>Preprocessing</td>
<td>43</td>
</tr>
<tr>
<td>Corpus evaluation</td>
<td>43</td>
</tr>
<tr>
<td>Sentiment scoring</td>
<td>44</td>
</tr>
<tr>
<td>ARIMA Analysis</td>
<td>45</td>
</tr>
<tr>
<td>Training the Model</td>
<td>45</td>
</tr>
<tr>
<td>Testing the Model- Prediction Accuracy and Validation</td>
<td>47</td>
</tr>
<tr>
<td>Limitations and Potential Errors</td>
<td>47</td>
</tr>
<tr>
<td>Summary</td>
<td>48</td>
</tr>
<tr>
<td>Chapter 4</td>
<td>49</td>
</tr>
<tr>
<td>Results</td>
<td>49</td>
</tr>
<tr>
<td>Consumer Reviews</td>
<td>49</td>
</tr>
<tr>
<td>Loading the Data</td>
<td>49</td>
</tr>
<tr>
<td>Cleaning and Scoring the Data</td>
<td>50</td>
</tr>
<tr>
<td>ARIMA Modeling</td>
<td>51</td>
</tr>
<tr>
<td>Defining the Model</td>
<td>52</td>
</tr>
<tr>
<td>Testing the Model</td>
<td>62</td>
</tr>
<tr>
<td>Conclusion</td>
<td>63</td>
</tr>
<tr>
<td>Chapter 5</td>
<td>64</td>
</tr>
<tr>
<td>Conclusion</td>
<td>64</td>
</tr>
<tr>
<td>Discussion of Findings</td>
<td>64</td>
</tr>
<tr>
<td>Section</td>
<td>Page</td>
</tr>
<tr>
<td>----------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>Consumer Reviews’ Influence on Stock price</td>
<td>65</td>
</tr>
<tr>
<td>Historical Trend</td>
<td>66</td>
</tr>
<tr>
<td>Managerial Implications</td>
<td>67</td>
</tr>
<tr>
<td>Response Time</td>
<td>68</td>
</tr>
<tr>
<td>Theoretical Implications</td>
<td>69</td>
</tr>
<tr>
<td>Limitations</td>
<td>70</td>
</tr>
<tr>
<td>Future Research</td>
<td>71</td>
</tr>
<tr>
<td>Summary</td>
<td>72</td>
</tr>
<tr>
<td>Appendix 1</td>
<td>73</td>
</tr>
<tr>
<td>Appendix 2</td>
<td>75</td>
</tr>
<tr>
<td>References</td>
<td>76</td>
</tr>
<tr>
<td>Curriculum Vitae</td>
<td>84</td>
</tr>
</tbody>
</table>
List of Tables

Table 1: Descriptive Statistics on Sentiment Scores................................................................. 51
Table 2: Lagged Correlation Values of Stock Price to Sentiment Score........................................ 53
Table 3: Variance Inflation Values of Predictor Variables.......................................................... 57
Table 4: AIC values for AR and MA terms from 0 to 9................................................................. 58
Table 5: Output from ARIMA Regression ....................................................................................... 61
Table 6: Confusion Matrix of Prediction ....................................................................................... 62
Table 7: Hypothesis Findings ....................................................................................................... 65
Table 8: Distribution of Reviews.................................................................................................. 75
List of Figures

Figure 1: Consumer socialization through social media .................................................. 10
Figure 2: Conceptual model of word-of-mouth ................................................................. 13
Figure 3: Hu, Ko, & Reddy's Model ................................................................................ 13
Figure 4: Process for sentiment analysis ........................................................................... 45
Figure 5: Distribution of reviews by day ............................................................................ 50
Figure 6: Plot of stock price over time .............................................................................. 54
Figure 7: Plot of percent price differential in stock price over time ................................ 55
Figure 8: Plot of entire percent differential of stock price over time .............................. 55
Figure 9: ACF graph of residuals from linear regression .................................................. 56
Figure 10: Partial ACF graph of residuals from linear regression ..................................... 56
Figure 11: ACF graph of residuals from ARIMA regression (3,0,8) .............................. 58
Figure 12: Partial ACF graph of residuals from ARIMA regression (3,0,8) ............... 59
Figure 13: p-Value of Box-Jenkins test ........................................................................... 60
Figure 14: Q-Q plot of residuals from ARIMA test ......................................................... 60
List of Equations

Equation 1: Bayes Formula ................................................................. 23
Equation 2: Bayes Formula (Derived) ...................................................... 23
Equation 3: Maximum Entropy Formula .................................................. 24
Equation 4: Support Vector Machines Formula ........................................ 24
Equation 5: Sentiment Ratio ................................................................. 33
Equation 6: Calculation of Dependent Variable ....................................... 40
Equation 7: Sentiment Score (Property Level) ......................................... 44
Equation 8: Daily Sentiment Score (Company Level) .............................. 44
Equation 9: Binomial Probability Equation .............................................. 62
Chapter 1

Introduction

As of 2014, mobile devices outnumber people on the entire planet (U.S. Travel Association, 2015). This rapid expansion of the digital environment has caused the internet to invade almost every aspect of daily life and caused it to become a significant influence in the consumer socialization of Generation X and Generation Y (Barber, 2013). This unfettered access and increasing reliance on the internet, have helped develop the plethora of review sites that have come to proliferate across the virtual landscape. These review sites have become an increasingly influential factor in the purchasing decisions of hospitality guests (Book, Tanford, Montgomery, & Love, 2015; Dhar & Chang, 2009; Ladhari & Michaud, 2015; Litvin, Goldsmith, & Pan, 2008). This new peer group of the consumer, sight unseen, can often become as influential as parents or close friends' personal experience, with 80% trusting online reviews almost as equally as a personal recommendation (BrightLocal, 2016). In addition, approximately 47% of these consumers are focused on the sentiment more than any other aspect of the reviews, and hospitality companies are the top ranked companies where these reviews matter the most (BrightLocal, 2016)

The proliferation of technology has created all new channels for the consumer to receive information. This information comes from marketers and peers and continues to evolve with each new iteration of technological innovation (Chong, Li, Ngai, Ch'ng, & Lee, 2016). As recent as 30 years ago, if a traveler wanted to know about a destination they had to rely on travel guides, travel agents, or speak with friends and family, whom they knew personally or encountered in their immediate geographic area and may not have been a recent experience. This close circle of peers and relatives while valuable in information was small, thus the
influence was limited but still important as this is how consumers learned to purchase and validated their purchases (Ward, 1974). With the invention of the internet, the information landscape changed dramatically. Consumers now seek information from other purchasers from around the world. They are no longer constrained by time or geographic location in the feedback from other consumers of the goods. The influence of the users’ peers became immense, to the point where online reviews are now a mainly deciding factor in purchasing intention, especially in hospitality (Litvin et al., 2008). Increased research demonstrates that the content, rather than the rating, of the review is a deciding factor for potential consumers (Hu, Ko, & Reddy, 2014). This focus on the content of the review rather than the rating of the review has culminated in the expansion of sentiment analysis. Researchers and operators can amass a vast quantity of these reviews thanks to the same technological innovations that spawned the review sites, the incremental growth of computer power that powers the internet and consumer electronics.

With this increased access to technology, there has also been an increased ability to rapidly perform analyses, which at one time would have been extraordinarily tedious and manually intensive, in a mere fraction of the time, and do it on datasets that astonishing in scope and size. The analytical field of sentiment analysis has become a growing focus in research being used to develop models to predict the success of Hollywood movies, online product sales, even the Dow Jones Industrial Average (Bollen, Mao, & Zeng, 2011; Chong et al., 2016; Doshi, Krauss, Nann, & Gloor, 2010; Joshi, Bharathi, & Rao, 2016; Lima et al., 2016).

While the prediction of the stock market is a lofty goal, the random walk nature of the news and market make the prediction difficult at best (Mlodinow, 2008; Zhang, Fuehres, & Gloor, 2011). If the market is the unpredictable, an arguable premise, what about a particular
company? While there are a nearly infinite number of factors that can influence a company's stock price, many of these are internally controllable by the firm. These can include labor management, proper financial management, or other decisions enacted by the company or particular property. Within hospitality, many of these factors carry with them a potential consumer impact. As competition increases between hospitality companies, and sometimes within a company, companies are no longer able to rely solely on managerial cost controls to maximize shareholder value (Chong et al., 2016). This has lead hospitality companies, especially gaming resorts, to increase expenditures of marketing dollars to attract more consumers to the properties, thus increasing revenue and eventually shareholder equity (Lucas & Bowen, 2002; Philander & Zhong, 2016).

**Problem Statement**

While the relationship between the sentiment of consumer reviews, and sentiment of the digital Twitter population, has been demonstrated to have a relationship to the financial performance of industries, either through share price or financial performance, there has been limited studies of this very real impact in the area of hospitality (Anderson, 2012; Yu, Duan, & Cao, 2013). While we can understand why reviews are influential, quantifying these effects in a meaningful manner to operators or the company shareholders is only anecdotal or assumed due to the relationship of purchasing intention. Consumer socialization theory states that, even at an early age, consumers are socialized through their peer groups, and this is now a worldwide peer group for the current and future generations of consumers (Ward, 1974). Since the ease for consumers to review hospitality industries can only increase in time, creating a quantifiable impact for the shareholders and operators to operationalize what consumers are saying is necessary and invaluable to the hospitality industry and furthering academic hospitality research.
Purpose of the Study

The purpose of this study is to create a model that is predictive in nature of a publicly traded hospitality company that is a tourist destination in nature using sentiment analysis to identify a lead time that operators and investors can use to help determine any directional shift of the share price. The reviews are collected from a wide range of review sites, and a dictionary corpus is used to train the sentiment classification algorithm to classify the sentiment of the reviews properly. ARIMA, Auto-Regressive Integrated Moving Average, modeling is then trained using the sentiment as the primary lagged covariate on a year's worth of daily stock price data. The model is then validated on sixty out-of-time days following the model training period.

Research Questions

This dissertation uses a reliable and externally valid methodology to assign sentiment scores to a hospitality company's consumer reviews, which are then utilized in a valid and reliable modeling technique to determine the relationship between them and the share price (Bowerman, O’Connell, & Kohler, 2004). The company selected is in a popular tourist destination market with many consumer reviews. The sentiment scores are also compared to the historical price of the company and tested to confirm that any correct predictions are not due to chance.

This dissertation addresses the following research questions:

1. How does the consumer sentiment of online reviews for a gaming company impact the stock price?
2. How does the standard historical price compare in forecasting the future with the sentiment of the consumers’ reviews?
Delimitations

As with all studies several limitations exist. First is the generalizability of the findings. While this study sets forth a methodology that can be replicated in research and practice, the direct findings are not generalizable to the population as a single company was selected and used to train the model for a given period. Since it is only focused on a single company, the results may not be generalizable to other companies, or industries. Additionally, the model was built using only 2015 reviews. Given a different time period of reviews to train the model with, different results could occur. Third, while the model does account for many of the associated time variables and seasonal effects, it does not consider any macro indicators or cost control efforts by the company to increase shareholder equity. Finally, there exist several methods of sentiment classification. While they are improving, they are not infallible and have certain limitations within the natural language processing. Any improvement in the classification methodologies could improve the results found in this dissertation.

Significance of the Study

This dissertation can provide insight into the relationship between the hospitality business and consumer reviews in a meaningful and quantifiable way. If significant it could further the development of consumer socialization theory, encourage academic study into consumer sentiment for hospitality, and encourage the usage of more advanced machine learning algorithms for the usage of complex studies.

Of course, there are managerial and operational implications additionally. Since this is an applied study, operators could replicate the methods used in this dissertation to gain insight and understanding of their customer base through the sentiment analysis. Investors could use this tool to reduce the uncertainty that is natural in investment in the stock price, or compare the
company’s public statements to the consumer’s experience. Further refinement of the model put forth could also have implications on revenue management and volume forecasting.

Definitions

Sentiment Analysis: Sentiment analysis is a Natural Language Processing that analyses opinion, emotion, attitudes, and sentiment from a written source (Liu, 2012).

Corpus: A large volume of text that is used as a reference for word identification by sentiment analysis software (Philander & Zhong, 2016).

Consumer socialization: The theory that consumers learn their habits through external factors, such as parents and peers, and this learned behavior influences the way spending is performed as adults (Ward, 1974).

ARIMA Modelling: A statistical technique related to multiple linear regression. Designed to incorporate serial correlation in the dependent variables (Auto-Regressive), differencing to create a stationary dependent variables (Integrated), and to adjust for serial correlation of the error terms (Moving Average) (Bowerman, O'Connell, & Koehler, 2004).

Consumer Review: The unsolicited feedback purchasers of a good or service leave via a review site. This can be a wide variety of topics, the focus of this research in on the field of hospitality.

Stock (share) price: The public traded value of a company’s common stock, as valued in US dollars at the close of a trading day on the New York Stock Exchange.

Holiday: Major US holidays that traditionally accompany time periods of travel, with respect to Las Vegas. This could be a single day instance, or influence an entire weekend. New Year’s Eve is an example of a holiday influencing multiple days as Las Vegas hotels require multiple night stay to be able to book a room for New Year’s Eve.
**Seasonality**: Any predictable pattern that repeats regularly with respect to changes in time. This can occur with-in a week (i.e. Weekends are busier than Mid-week), within a month, or with-in a year (i.e. summer is slower than fall).

**Conclusion**

This chapter provided an introduction to consumer reviews, sentiment analysis, and the predictive ability of understanding what a consumer says about a product. The purpose and research questions guiding this dissertation is laid out, and limitations of this study are identified. Based on these research questions, the remaining chapters will be as follows. Chapter 2 will review the relevant literature of consumer socialization theory, electronic-word-of-mouth, consumer reviews, and sentiment analysis. Chapter 3 will describe in detail the methodology of this dissertation, with the results and analysis of the model in Chapter 4. Chapter 5 will explore the findings, the implications, both theoretical and practical, and guidance for future researchers.
Chapter 2

Literature Review

The idea of consumer reviews influencing purchasing intention is one that exists at length in the extant literature. This notion can be derived from the consumer socialization theory, as well as the theoretical underpinnings of the information acceptance model (Erkan & Evans, 2016; Hu, Koh, & Reddy, 2014; Ward, 1974). This literature review will examine the literature to develop further the ideas of consumer socialization theory, the information acceptance model, the impact of electronic word of mouth, and the usage of sentiment analysis.

Consumer Socialization Theory

Consumer socialization theory was first put forth by Ward in 1974 as research focused on the impact of marketing to children (Ward, 1974). While the theory focuses on the development of young adults, it does carry through to adult purchasing behavior (Moschis & Churchill Jr, 1978). This theory, developed in the 1970's, predates the invention of the internet, much less social media as we know it today. However, the concept that the greater the communication a consumer has with their peers, the greater their awareness of the goods in the marketplace would still be applicable today (Moschis & Churchill Jr, 1978). The proliferation of social media has expanded a consumer's peer group by an exponential amount. As the first generation of this social media age enters adulthood, Millennials have had essentially the largest peer group of any generation before them (Sparks & Browning, 2011). Through Twitter, Facebook, Instagram, and YouTube, this generation's peer group is no longer confined to school or neighborhood friends, but rather the number of users that visit these sites (Sparks & Browning, 2011). Millennials also have the tendency to utilize social media and the review sites as a source of information prior to making their purchasing decision (Moore, 2012). While peers have been considered great
influencers into consumer socialization, the impact of the social media is one that can carry an even more profound impact as the size of the peer group continues to influence even beyond the traditional development stage (Wang, Yu, & Wei, 2012). This influence is shown to have impacts on the adult consumers in the sense that the peer behavior establishes the social consumer norms, to the extent that the consumer prefers recommendations of these perceived "peers" when in comparison to the established rating systems by critics (Dhar & Chang, 2009). Wang, et al (2012) found in their study that the peer involvement through social media is supported as an addition into the overall consumer socialization framework. This impact on the consumer purchasing intention was both significant and positive in their study. Though they did not look at the sentiment of the communication, the mere focus of the research was the influence of the overall communication on the overall purchasing intentions of the consumers (Wang et al., 2012). This theoretical framework (Figure 1) has been evaluated in other studies to determine the influence of the sentiment on the actual purchasing of items.
The impact of the internet, and specifically social media, as an influential factor in the development of both informative and normative socialization in the newer generations is a factor that, as technology advances, is unable to be discarded (Barber, 2013). Barber (2013) explored this impact in his study and determined that both Generation X and Generation Y social networks dominated the usage of internet in hours per week. This impact carried through in the analysis, and it was determined that the internet provided informative influences on both generations at a statistically significant level (Barber, 2013). This new informative information source exceeded the traditional "peer" level per classical socialization theory (Barber, 2013). The impact of this is that both Generation X and Generation Y are age groups that have grown with the internet, with Generation Y never knowing a world without the internet (Lissitsa & Kol, 2016). The supplantation of more traditional socializers by the internet has also potentially created the opportunity for marketers to reach these consumers directly, by encouraging positive reviews, with less skepticism than more traditional means (Moscardelli & Liston-Heyes, 2005).
Further supporting this notion of the technological peer's influence on the consumer socialization process is the theoretical underpinning from the Information Acceptance Model (Erkan & Evans, 2016). Erkan and Evans (2016) expanded upon the Information Adoption Model to include various aspects of the Theory of Reasoned Action to produce a new model, the Information Acceptance Model. This causal factor behind this was that the Information Adoption Model does not weigh the necessary critique consumers must do when faced with online reviews and the information provided in them (Erkan & Evans, 2016). This information must be more than simply "consumed," it must carry a substantial impact on the consumer purchasing decision. Of the determinates of the model, they found that credibility to be the lowest weighed factor for information usefulness, which is highly influential in data adoption and thus purchasing intention, but the needs of the information carried the greatest impact on the measurement of information usefulness (Erkan & Evans, 2016). Additionally, this theory further addresses the technological aspects missing in the consumer socialization theory. Providing the need for consumer feedback by other potential consumers is highly influential in the new consumer's future purchasing intention. With this in consideration, the intangibility of products from hospitality, including casino resorts, establishes a need for consumers to review a source of information relevant to their purchasing intention that contains greater perceived validity than the established “experts.” (Dhar & Chang, 2009; Erkan & Evans, 2016; Wang et al., 2012). The sources for this information, given the modern technological landscape, are social media or consumer reviews often referred to as electronic Word of Mouth (eWOM).
Social Media and eWOM

The study of electronic word of mouth (eWOM) as a marketing initiative is extensive, with much of the body of literature focusing on the influential factors that eWOM can carry on a good and the sales thereof (Cabiddu, Carlo, & Piccoli, 2014; Dhar & Chang, 2009; Erkan & Evans, 2016; Ladhari & Michaud, 2015; Leung, Law, Hoof, & Buhalis, 2013; Litvin, Goldsmith, & Pan., 2008; Park & Nicolau, 2015; Sparks & Browning, 2011a; Sparks, Perkins, & Buckley, 2013; Wang et al., 2012). Litven, et al. (2008) go as far as to say that for the hospitality industry word of mouth is the most influential information source for consumers to make their decisions. In their research they consolidated the body of literature involving word of mouth into a conceptual model that identifies the sources of word of mouth, mediating variables, and the outcomes (Figure 2) (Litvin et al., 2008). This paper focuses on the "Consumption Experience" path of the communication. This path represents the consumer review and how their eWOM would be spread to the consumer.

eWOM is important, especially to hospitality industry, because hospitality businesses are strongly susceptible to the ebbs and flows of the online reviews of their consumers (Ladhari & Michaud, 2015). The influence that these consumer reviews have on the consumer purchasing intention is generally agreed upon in the extant body of literature (O’Connor, 2008). This is further expanded upon to say that the content of the review is what matters, not necessarily the star rating (N. Hu et al., 2014). Hu, Ko, and Reddy’s (2014) model demonstrate that the sentiment of the review can matter more than the overall star rating of the property (Figure 3).
The next section will look at eWOM from the standpoint of consumer reviews, and cover the relevant literature to expand upon Hu et al’s (2014) concept that what is said in the review matters and creates a financial impact to the hospitality industries.

**Consumer Reviews**

With technology growing at a substantial rate, the barriers of communication between consumers breaks down allowing the flow of information between consumers to potentially
override the communicated message from the seller regarding the product (Moe, Trusov, & Smith, 2011). This information disparity creates leverage for eWOM, and the vehicle for this sharing of information is often in the form of consumer reviews. Consumer reviews, especially online reviews, as stated by Ladhhari and Michard (2015) can be one of the most influential forces in the determination of a choice by a consumer looking for a service in the hospitality industry.

In a study of hotel properties in London and Paris, the traditional "expert" star ratings did not set the prices of the properties, rather the consumer ratings helped to dictate the prices (Öğüt & Onur Taş, 2012). In fact, the higher the "expert" star rating, the greater the sensitivity to the online review sentiment. Additionally, price may no longer be the deciding factor when selecting hotels or vacation destinations (Book, Tanford, Montgomery, & Love, 2015). In their study, Book, Tanford, Montgomery and Love (2015) found that price rated below both the valence of the reviews and their uniformity. In a study on airlines, negative word-of-mouth has been found to negatively impact not only sales, but the share price of the company as well (Luo, 2009). This impact was measurable and timely, in that the negative consequences did not peak for several months, and then lasted months later. For an industry that is similar to hospitality in many aspects, this is a significant finding that supports the hypothesis that there is a time \((t)\) that is relevant to the operation in which the sentiment of online reviews can predict the directionality of the firm's stock price.

Additionally, since the reviews are, generally, not compensated in any way, they can be typically considered unbiased and can be viewed as honest results (Moe et al., 2011). This trust in the reviews could cause a social desirability bias, and does in the short term, but the long term
effects of any bias are negligible and therefore a large sample of reviews can be viewed as free from this bias (Moe et al., 2011).

As the impact of consumer reviews have been demonstrated in the literature, impacting not only price but selection of the hotels, the hospitality industry must gauge the quantitative impact these reviews can have on revenue, and ultimately their share price (Book et al., 2015; Ladhari & Michaud, 2015). Looking at the simple star rating, or other metric provided for the top-level rating, may not give the whole picture though as Hu et al (2014) demonstrated, what is said can matter. The more positive the statements made regarding a well-managed hospitality firm, and the more positive the reviews left in response, then the greater the firm's revenues. Consumer, and investor, confidence can also sway and therefore potentially influence the stock price of the firm (Joshi et al., 2016).

**Consumer Reviews’ Link to Firm Performance**

The extant literature exploring the link between online consumer reviews, or eWOM, and the financial performance of a firm exists in and outside of hospitality. Anderson (2012) explores the relationship between online reviews and the direct financial impact, in regards to revenue, they can have on firms. In his analysis, he found that consumers were visiting review sites at in greater number and at a greater frequency in 2010 than in the prior years. Additionally, over a quarter of the potential hotel guests visiting the review sites a mere five days prior to the vacation and booking (Anderson, 2012). This could provide the baseline for a lead time for when a review is left, and the potential financial impact upon the firm. He then used a logistic regression to determine the actual impact on the odds ratio of booking at a particular hotel given the star rating, price, and number of reviews. He found that an increase of a single star in the overall rating carried a 14.2% increase in the odds of the hotel being selected, given
that all other factors are equal. The volume of reviews also favorably increases the odds of the hotel being selected as the number of reviews increases, though to a lesser degree than the star rating. The author proceeded to carry the analysis through to ADR, Occupancy, and RevPAR. Using a metric that takes and scores properties based on their online reviews as collected from multiple sites, the author found that a one percent increase in this combined metric, which the author states is a proprietary algorithm using only the quantitative scores, leads to increased ADR, Occupancy, and RevPAR (Anderson, 2012). These are all commonly used metrics of a hotel’s performance and financial well-being.

In another study, the authors sought to directly address the question of the impact of reviews on a hotel’s performance (Xie, Zhang, & Zhang, 2014). Their research supports the notion that increased positivity in online reviews is beneficial to the financial performance of the hotel. Looking at reviews for a range of classification of properties, the authors used reviews from only one review site to gather not only the overall review rating, but the more detail attribute rating for the property. This more detailed rating is analogous to the information that could be left by the reviewer in the text of the review. Their study was conducted on hotel properties only in the state of Texas, using RevPAR numbers obtained from the Texas Comptrollers database. Using quarterly data, the authors found that for every point of increase in the overall rating, the RevPAR of the hotel can be expected to increase by approximately $66.70 (Xie et al., 2014). The unweighted coefficients of the regression model also demonstrated that the overall review strength outweighed the review variation and volume as a predictor in the success of the property. Their finding supports Anderson’s (2012) notion that the overall review left by consumers is related to the financial performance of a hotel company, as much as RevPAR is usable as a performance indicator.
Taking this line of thought to the focus of this research, the share price of a firm, a study by Yu, Duan, and Cao (2013) state “social media can provide the timely evaluation of firms’ performance, which allow the investors not only to follow consumers’ sentiment but also to predict their future business value.” They used stock price in their study due to the inaccessibility of daily sales data, which is not often released to the public, unlike the stock price (Yu, Duan, & Cao, 2013). The research focused on a broad range of social media, including Twitter, to determine the potential influence on the share price from such eWOM outlets may carry. Additionally, the authors included hotels into their list of firms for the study. The authors utilized sentiment analysis to extract the meaning of the social media reviews, and quantify it in a way to fit into an econometric risk/return model. Using these techniques at a one-day lag, the authors found significance in all the social media review outlets. The authors found that the lengthier review methods carried a greater impact than the relatively shorter Twitter. While the authors did not separate out hotels specifically in this study, the overall impact lends credence to the argument that what is said online can impact a firm’s share price. The impact can be even greater, according to the authors, than conventional media coverage (Yu et al., 2013).

Sentiment Analysis

Sentiment Analysis is a subset of text data mining, and utilizes many of the same base features in the initial setup of the data (Miner, 2012; Philander & Zhong, 2016). Miner et al. (2012) trace text classification and data mining back to the Library of Oxford with the library catalog of Thomas Hyde. While this might be the true historical roots of text classification, it was not until the development of modern computers that the process of natural language processing and sophisticated computer algorithms allowed for the classification of large volumes of text in short periods of time (Miner, 2012). The culmination of this computational ability to
parse natural language and extract intent and meaning was Watson, an analytical computer that beat the best “Jeopardy!” champions in a three-day test (Miner, 2012). While this dissertation will not develop a “Jeopardy!” champion, the usage of similar techniques to extract the sentiment of consumer reviews is still possible as technology has continued to develop even beyond Watson.

Traditionally, the process to determine the sentiment of consumers involved the pooling of a sample of qualitative data and manually processing the content, coding it to reflect the intent of the writer (Gibbs, 2007). Words that have a positive sentiment are generally assigned a value of positive one (1), while words that portray a negative sentiment are assigned a value of negative one (-1), and all other words are given no value or zero (0) (Miner, 2012). The overall score or sentiment of the text is the summation of each individual value (Miner, 2012; Philander & Zhong, 2016). This process is both manually intensive requiring multiple reviewers to establish inter-reviewer ratings, and time intensive to parse large data sets. As the volume of the data grows, this becomes both impractical and resource prohibitive to evaluate very large data sets (Philander & Zhong, 2016). To accomplish this with the growing sources of online data, computer algorithms have been developed to accomplish this analysis on the large amounts of data that can be extracted from modern sources (Pang & Lee, 2008). This involves two major processes. The first is to determine the subjective elements from the objective elements, then the second process gives weights to these “opinion words” (Chiu, Chiu, Sung, & Hsieh, 2015). There are multiple methods to extract the sentiment of the data, as technology and research continues to focus on the development of new techniques that increase reliability and accuracy in the assignment of the sentiment (Bai, 2011; Chiu et al., 2015; L.Lima et al., 2016; Philander &
Zhong, 2016). Ultimately, this comes down to two main methods, the lexicon based and the corpus based (Chiu et al., 2015).

Prior to choosing the method of sentiment analysis, it should be discussed the different levels of analysis that are able to be conducted. The main focus has, predominately, been in three potential levels of specificity, or "granularity," of the text data (Liu, 2012). Liu (2012) classified the three levels as follows:

**Document Level**: This granularity is to determine the sentiment, either positive or negative, of the document as a whole. The assumption is that the document expresses opinions on a single good, or product.

**Sentence Level**: This granularity evaluates the sentiment of each sentence within a document. Sentiment classification at this level includes a neutrality aspect in addition to the positive and negative aspects of the document level. This can typically involve separating the sentences according to subjectivity to separate factual contexts from opinion based statements.

**Entity Aspect Level**: This, the finest grain of analysis, looks at the sentiment as compared to the target entity. Comparing the relationship between the emotions expressed and the focus of the emotion provides the greatest detail, and potentially the greatest insight, into the sentiment being voiced by the consumer.

Just as the levels of granularity increase in focus and detail, so does the complexity of the overall analysis, with the entity level being the highest level of complexity (Liu, 2012). Liu (2012) goes on to the state that in addition to the granularity issue of sentiment analysis, there is also a topical issue. This is to mean that consumers leaving a review may use a comparison of one product, or entity, to another (Liu, 2012). An example of this is two reviews one saying "The Bellagio is a great hotel" vs "The Bellagio is much better than the Wynn hotel." The usage
of the word "better" in the second sentence can create a false positive sentiment if the assumed subject is the Wynn Hotel, however the actual intended target is the Bellagio Hotel (Liu, 2012). Once the appropriate level of granularity has been determined prior to analyzing the data, the appropriate sentiment determining methodology can be used.

The incremental usage of online behavior, the increasing technological capability of computers to even the most basic of consumers, and the advances of sophisticated software to drive the research has led to a greater focus on the usage of sentiment analysis for a variety of research purposes (Bai, 2011; Philander & Zhong, 2016). This machine driven methodology has also been found to not only reduce the human resources needed to complete an analysis, as modern computers are able to process large volumes of text data relatively quickly, but has been found to be of greater consistency and reliability as human dependent methodologies (Capriello, Mason, Davis, & Crotts, 2013). Unlike human methodologies, where human readers can read and interpret the meaning out of documents, machine based sentiment analysis needs a list of words to compare the document to for evaluation (Liu, 2012). This list can be built by one of two potential methods: Domain specific lexicon based, or general corpus based (Liu, 2012; Philander & Zhong, 2016).

**Lexicon Based Analysis**

Lexicon based methods use a dictionary of words that have values pre-assigned to them to determine the valence orientation (Chiu et al., 2015). This process involves manual creation of terminology to be identified, and can be considered a supervised approach and is built with the domain in mind (Taboada, Brooke, Tofiloski, Voll, & Stede, 2011). A major drawback to this approach is that the efficacy of the lexicon diminishes rapidly when used in a different domain, as terminology shifts and words that could represent a positive sentiment are now considered
negative (Taboada et al., 2011). An example of this is a review stating "The massage was slow and took hours" can be generally considered a positive sentiment as the purpose of the massage is to encourage relaxation and to maximize the time spend in that state. Conversely, "The dinner service was slow and took hours" conveys a different sentiment as the expectation of a restaurant's service is to be timely. Even though both of the examples are in the hospitality industry, and could potentially even occur in the same resort, the inclusion of the word "slow" into the semantic lexicon as a negative sentiment could have an adverse effect if the analysis is being done at a document, or even sentence level, analysis. The usage of a lexicon based methodology does have advantages however. The ability to include specific "slang" or local dialect uniqueness can be considered more carefully (Taboada et al., 2011). This includes specific attributes to the item being described, and can include items that, in separate contexts, could constitute gibberish or otherwise irrelevancy (Taboada et al., 2011).

**Corpus Based Analysis**

The corpus based methodology uses probability of association with a set of positive or negative words bases (Philander & Zhong, 2016). The corpus based methodology is best used when there is an available corpus of known words for a given domain or a dictionary is used containing both the word and any possible synonyms and antonyms (Liu, 2012). This too suffers the same potentially dangerous approach as the lexicon based, as word meaning can shift from one context to the next (Liu, 2012). While is it possible to incorporate both methodologies, that is to take an existing corpus and adapt certain words based on the given domain, dictionary usage could eliminate this need, as dictionaries are comprehensive in their inclusion of the words (Liu, 2012). This approach can be used in specific circumstances when the given corpus or dictionary
is not, such as when dealing with non-English languages (Abdulla, Ahmed, Shehab, & Al-Ayyoub, 2013).

While the hospitality industry is a unique industry, the usage of a lexicon based methodology is not essential to produce accurate results (Philander & Zhong, 2016). A dictionary, or corpus based, methodology can successfully be utilized, but the specific data might be reviewed prior to enhance accuracy of classification (Chiu et al., 2015). Once the appropriate granularity of analysis is determined, and either the lexicon or corpus methodology selected, then the next step of sentiment analysis is the classification.

**Document Classification**

Traditionally, at the document level, only two levels of classification are used, sentiments are either positive or negative (Liu, 2012). This broad classification of the document, especially in the case of consumer reviews, can be easily identified through the general rating assigned to the product (Liu, 2012). When examining if the best process for this is human intuition or machine learning, Pang, Lee, and Vaithyanathan (2002) had human subjects classify movie reviews, according to intuitive lists of words, into either positive, negative, or a "tie." The tie category was when the review was equally rated as positive and negative. The human intuition method resulted in moderate accuracy (58-64%), but very high tie rates (39-75%) (Pang et al., 2002). By examining the data through a person using a corpus built from the highest frequency words used in the data and classifying using this statistical build list, the accuracy rate exceed the intuitive lists in accuracy (69%) and reduced the overall tie rate (16%) (Pang et al., 2002). Using these human classified scoring as a champion to the challenger of machine based algorithms, they then tested three different machine based methodologies, Naïve Bayes, maximum entropy classification, and support vector machines (Pang et al., 2002). The framework for their testing
involved the classification of a set of predefined n-grams where each document \(d\) is a vector of \(n_i(d)\) \(<\text{as } i \text{ goes from 1 to } m\>\), which is the number of times \(f_i\) occurs in the document (Pang et al., 2002).

**Naïve Bayes.**

Naïve Bayes is the simplest of these three classification methods, as it is based upon Bayes Rule (Equation 1):

\[
P(c|d) = \frac{P(c)P(d|c)}{P(d)}
\]  

(1)

Assuming the \(f_i\)’s are conditionally independent of a given \(d\)’s class the following equation (Equation 2) can be derived (Pang et al., 2002):

\[
P_{NB}(c|d) := \frac{P(c)(\prod_{i=1}^{m} P(f_i|c)^{n_i(d)})}{P(d)}
\]  

(2)

This methodology, in view of equation 1, seems simple and the assumption of conditional independence may not survive to extend to actual situations, as a bi-gram could violate this assumption easily, performs very well and has been shown to be ideal for “certain problem classes with highly dependent features” (Pang et al., 2002).

**Maximum entropy classification.**

The next classification method used was maximum entropy classification. This methodology has proven to, on occasion, outperform Naïve Bayes methodology (Nigam, Lafferty, & McCallum, 1999). This methodology estimates \(P(c|d)\) in an exponential form (Equation 3) using the normalization function \(Z(d)\) (Pang et al., 2002):

\[
P_{ME}(c|d) := \frac{1}{Z(d)} \exp\left(\sum_i \lambda_i c F_i(c, d)\right)
\]  

(3)
The assumptions made in Bayes is not made here as the function $F_{i,c}$ only occurs if a bi-
gram is identified for the appropriate classification (Pang et al., 2002). Pang et al. (2002) used
an iterative training data set for ten iterations with a Gaussian prior to prevent the over fit of the
estimate $\lambda_{i,c}$, which allows to make the least number of assumptions of the document while
maintain the consistency of the document.

**Support vector machines.**

The third methodology used was support vector machines, which can also outperform
Naïve Bayes on a more consistent basis (Pang et al., 2002). This method is a “large margin”
methodology, as opposed to one using the probability of a correct classification like Naïve Bayes
and maximum entropy (Pang et al., 2002). This method is a dual optimization problem, which
the goal is to maximize the distance from a designated hyperplane through the optimization of
the following (Equation 4) (Pang et al., 2002):

$$\bar{\omega} := \sum_j \alpha_j c_j \tilde{d}_j, \ \alpha_j \geq 0 \quad (4)$$

If $\alpha_j$ is positive then the document vector is considered to be a *support vector* since it
contributes to the solution, which side of the hyperplane the document falls on, and thus
determining the sentiment classification (Pang et al., 2002).

Using unigrams and bi-grams helped to achieve the best results for Pang et al’s (2002)
analysis of the movie reviews. Even Naïve Bayes, in which the additions of the bi-grams
theoretically violate the assumption of conditional independence, performed better than with
unigram frequency alone, though .4 % worse than unigram presence (Pang et al., 2002). All
three methods far exceeded the human classification benchmarks set in the control methods, and
the three methods were close in their accuracy percentages, with support vector machines having
a slight edge over Naïve Bayes (Pang et al., 2002). Pang et al (2002) did note that the sentiment identification was not as accurate as the topical classifications that used the same methodology, and theorized that this may be due to the nature of language used in the reviews and the contradictory meaning that can be built in context. As an example, a reviewer said that the movie in question was “a really awful movie…the ninth floor of hell. The plot is such a mess that it’s terrible. But I loved it” (Pang et al., 2002). This type of review could easily be determined as positive by a human classifier, but the machine based algorithms would identify the large volume of negatives and potentially incorrectly classify it. To this end, the authors recommend identifying if the sentences are focused on the correct topic as “the whole is not necessarily the sum of the parts” (Pang et al., 2002).

The methodology of document level review is adequate for providing an overall idea of the sentiment of a product, but the restrictions on it are that it is unable to provide details regarding the aspects of the product, nor is it easily applied to discussion forums where the topic might be implied or shifting (Liu, 2012). While Liu (2012) states that “online reviews do not need sentiment classification because almost all reviews already have use-assigned star ratings”, subsequent research has shown that the content of the review might be more influential than initially though as the devil may be in the details (N. Hu et al., 2014).

**Sentence Based Sentiment Analysis**

Sentence based sentiment analysis is similar in many aspects to document level analysis (Liu, 2012). This is due to the nature of sentences, which are complete thoughts, and essentially just short documents from an analytical perspective (Liu, 2012). The process for sentence sentiment analysis is similar to that of the document analysis, but is focuses more on the differencing of objective versus subjective statements (Liu, 2012).
Generally speaking, hospitality reviews are simple sentences that form multiple individual documents. However, as noted by Liu (2012) and contrasted by Hu et al. (2014), reviews have a known overall sentiment that may not reflect the specific sentiment contained within the review. Sentence sentiment analysis could potentially conflict with the overall rating left by the reviewer, or individual parts of the review could. While Philander and Zhong (2016) used a document based classification method, their domain of interest was Twitter. Twitter restricts any “tweet” to a maximum of 140 characters. This restriction limits the number of sentences that can be formed, as just these three past sentences would have exceeded the 140-character limit.

Online review sites, in contrast, do not place such a short limit on the length of the review (Sparks & Browning, 2011). Some sites do not place any restrictions on the length of the review, leading to the reviews online to be varied in both depth of sentiment and length (Sparks & Browning, 2011). For example, the following review was taken from TripAdvisor and the reviewer left five stars for the property. This review is verbatim and provided as an example of the writing of an exceptionally lengthy review. It also demonstrates the challenges of sentiment analysis, in regarding to improper sentences and spelling errors that can be common in consumer reviews.

We stay at the Bellagio every year and although there are things that irritate us, we can't imagine staying anywhere else. Check-in was the usual scrum with long lines that never seen to move and of course, I always pick the shortest queue with the neediest guests in front of me who ask the most stupid questions and can talk for ever - hey ho, such is life!! On getting to the front, the agent was pleasant enough but she did not go that extra mile to assist me. We were booked into a fountain view room with a specific
request for non-smoking but none were available and all booked up - even at 10am!!! I gave her the "I suffer from bronchitis" line (which I don't!) but she couldn't have cared less and no amount of pleading would help! I am an M-life platinum holder and was staying at the Hotel for 14 nights so I thought she'd at least try but she was stubborn and awkward in her manner as most of the agents seem to be - hardly surprising really as some of the guests they have to deal with would drive you to drink - but don't treat us all the same! As a result, I went over to see my host - Mr. Michael Black, and he sorted me out a non-smoking fountain view room in seconds! He even updated my card to VIP status as a result of the inconvenience which was a very kind gesture and much appreciated. We arrived at our room - 28004 which on first impression was lovely located near to the lifts and with an excellent view of the fountains. We unpacked our belongings and then noticed that there was a very loud buzzing noise which was really irritating. As a result, we rang the front desk and they sent an engineer up. On arrival back at our room, he had left a note apologising stating that the buzzing noise was generators on the floor above and that we would have to change rooms. This was annoying but we were pleased that the engineer attended quickly and dealt with our concerns. It is quite obvious that this is a long standing issue which cannot be rectified - avoid 28004!!! On the following day, I went to see our host, and he arranged for a new room on the 28th floor quite close by which I requested as it was much easier to transfer our belongings. He did ask me if I was sure that the 28th floor was the best option with the previous issue which I foolishly said yes!! I was told to check back at the VIP desk at 4pm as the previous guest had organised a late check out. I duly went back at 3.30pm to be told that the room had been allocated to another guest by an agent on the main front
I was fuming with no hint of an apology! I was offered rooms on other floors but as 4 sets of lifts service various levels, it was a case of moving our stuff on hangers etc. from the 28th floor to the ground floor, traipsing across the lobby and then getting another elevator to the new floor!! I told the agent this was not acceptable and I was told there was no other fountain view rooms on the 28th floor. I ditched the English "non-complaining" approach and adopted the American approach and demanded that she contact my host so I could speak with him. She said that he wouldn't be able to do anything but I insisted that she contact him straight away. She called up his number and shoved me on the phone without having the courtesy to speak with him first to advise him of the situation. Whilst I was speaking with him, she looked at me with intent hatred and scowled throughout. Of course, my host sorted it out within minutes and a room was arranged on the 28th floor - albeit it quite a distance away but that did not matter. He also arranged for us to enjoy a complimentary meal at PRIME for that evening which was absolutely fantastic and made up for everything. The agent....... if looks could kill!!! She slammed the new room key on the counter to which I said "Have a nice day!!" The transfer to room 28-040 took no time at all as we were helped by a very kind and considerate Housekeeper who told us that the room we were in was renowned for the buzzing problem. We checked the room out before we moved our belongings and it seemed even nicer than the previous room.... or so we thought!! We were awoken at 7am by the loudest rumbling/plumbing noises that I have ever heard - it was like an earthquake!! It is quite clear that when anyone nearby emptied a bath or flushed the toilet the whole drain system gurgles, bangs and makes loads of noise. We decided we were not moving again so did not complain although it was quite clear that there were problems as
the engineer was in surrounding rooms daily trying to sort it out without success. One evening we came back and the room absolutely stank of stagnant water/drains and of course our toilet blocked twice so the engineer sorted this out straight away. Avoid the 28th floor....... You may wonder why the 5 star review? Well apart from these issues, everything else was perfect. The Hotel is magnificent, all the restaurants are first class, the staff are wonderful, the Housekeeping was excellent. The casino itself is great too although as with all Vegas hotels, very smoky! We will always stay at the Bellagio on our annual trips to Vegas as it is a top-class Hotel with an excellent location and facilities. Our host, Michael Black, is an excellent asset too and I am sure if he wasn't there to iron out any issues and look after us like he always does, then we would have to consider alternatives as some of the check-in agents couldn't care less whether you stay there or not which is such a shame and let's the remaining staff down who are dedicated to your happiness (Review Trackers, 2016).

While the reviewer left five stars on TripAdvisor, and even commented about the contradictory evaluation of the posting, evaluating this review at a document level could potentially lead to a false negative. However, analyzing this review at a sentence level and comparing to the document level and neighboring sentences would produce a result of greater accuracy (McDonald, Hannan, Neylon, Wells, & Reynar, 2007). In their research, they proposed that by evaluating sentences in the context of the document evaluation could produce more accurate results. This exceeded the accuracy of not only the document level classification, but they showed that evaluating each sentence in a vacuum, not contingent upon the previous sentence sentiment, produced the least accurate results (McDonald et al., 2007). The increase in accuracy of the analysis is accomplished through the usage of nesting the classification and
creating a constraint of the sentence classification to that of the overall document classification (McDonald et al., 2007). That is to say, that each sentence is evaluated from a position of the polarity of the overall document, and additionally in comparison to each sentence. While Liu (2012) recommends the careful evaluation of the sentences to identify the objective non-sentence based statements from those that are subjective and sentiment based or objective sentiment based, McDonald et al.'s (2007) evaluates the object statements as neutral and therefore no bearing on the overall classification.

**Aspect Level Classification**

The final methodology, aspect classification, is to identify and associate individual aspects with the sentiment (Liu, 2012). This is highly useful when the individual aspects of the product, or entity, are the ones within the scope of the research question (Liu, 2012). For example, in the above review the aspects of the Front Desk and Rooms might carry a negative sentiment, while the aspect of the Host, casino, and restaurants, specifically Prime, would carry a positive sentiment. Typically, this involves calculating the sentiment scores for the subjective sentiment words, then taking the multiplicative inverse of the score by the distance between the subjective sentiment and the aspect (Liu, 2012). The other facets of this method are similar to the sentence and document methodology. The additional step of identifying the aspect or the target of the subjective sentiment complicates the analysis as it would require the use of either an expanded lexicon or additional algorithms to assign likelihoods to possible aspects (Liu, 2012).

Understanding the mechanics, and the intricacies of sentiment analysis is essential to determine the appropriate methodology. The level of detail required in the research question (i.e. does the individual aspect of the entity impact the question? Does the review length allow for conflicting statements? Or does the polarity of the total sentiment matter only?) dictates the
granularity of the analysis. The idiosyncrasies of the entity reviewed determine the corpus style to use, does the nature of the entity create conflicting sentiment words or have unique phrasing that would be required, or does it follow the general normality of the natural language. The literature shows that what is said matters, and that these reviews are creating a significant impact on the consumer behavior (Book et al., 2015; Hu et al., 2014). This is because of the social influence that a consumer's peer group creates on their socialization, and the willingness to accept the perceived trustworthy unbiased input from consumers of an intangible product (Moe et al., 2011; Wang et al., 2012). The next section of this dissertation will review the literature that uses sentiment analysis as a predictor. While the literature on this in the pure hospitality industry is limited, the methodology is used to predict stock price, film revenue, and sales.

**Sentiment Analysis as Predictor**

As the development of sentiment analysis expanded, so did the ways researchers look to utilize and operationalize it? As the relationship between reviews and sales has become evident in the literature, there becomes an almost natural progression to attempt to predict sales as a function of the sentiment expressed in the review.

**Early Prediction Attempts**

One of the earliest efforts, was to predict the success of movies per the Hollywood Stock Exchange (Doshi, Krauss, Nann, & Gloor, 2010). Doshi et al (2010) gathered reviews from Internet Movie Database (IMDB) and Rotten Tomatoes as a balanced approach to weigh consumer reviews, from IMDB, and "expert" reviews, from Rotten Tomatoes, to create a "balanced" review set. While they used "expert" reviews, this may not carry the influence that would be needed for a prediction as "peer" consumer reviews are often valued more than the critic or "expert" reviews (Öğüt & Onur Taş, 2012). The authors used IMDB with the theory
that as the polarity of the reviews build in either direction, the number of consumers to see the film, and thus increase the box office value of the film, would build (Doshi et al., 2010).

Doshi et al (2010) used a modified lexicon approach by taking a dictionary and modifying it according to the appropriate movie's genre. This was accomplished by looking at the term document matrix of the reviews, and adjusting the corpus to the specific unigrams and bigrams that had high frequency in the reviews. This was then applied via the document level methodology, as there was no mention of the scoring of individual sentences, rather the entire review was scored as a single document (Liu, 2012). The sentiment score was scored traditionally, with positive words given a value of positive one and negative reviews given a value of negative one, with the overall sentiment of the review being the net sum of the scores (Liu, 2012). This value was then used in their regression algorithm to predict the Hollywood Stock Exchange price (Doshi et al., 2010). In their regression, the authors used a time lag by manually lagging the data compared to the dependent variable (Doshi et al., 2010). The mean square error, being used as the goodness of fit testing, ranged between 10.387 and 12.853 providing a reasonable estimate of the dependent variable for each day (Lehmann, 1991). While Doshi et al (2010) did not use an ARIMA model for the time series data prediction, which is generally recommended, their results are still significant and shows that the sentiment of the reviews are able to reasonably estimate the performance of the film (Lehmann, 1991). Since the Hollywood Stock Exchange is a futures exchange and therefore subjective to investor whims, and the implementation of the budget of a film is a complex calculation that, while related to the ticket sales at the box office, is still dependent on the spending and budget that went into the production of the film (Doshi et al., 2010); this is a potentially analogous to the hospitality
industry, where the revenue is merely one factor in the final profitability of the firm and the stock price is subjected to the whims of the investors.

**Advancing Prediction Techniques**

The usage of sentiment analysis to accurately predict box office revenue has been successfully repeated in different cultures and using different regions and film types as well, such as China (Liu, Ding, Chen, Chen, & Guo, 2016). In their work, Liu et al (2016) used the "sentiment ratio" of the document level twitter statement. The “sentiment ratio” is simply the ratio of the positive sentiments less the negative sentiments as can be seen in equation 5 for the sentiment ratio for time $t$.

$$Sentiment\ Ratio_t = \frac{P_t - N_t}{T_t}$$ (5)

Where $P_t$ is the number of overall positive reviews at time $t$, $N_t$ is the number of overall negative reviews at time $t$, and $T_t$ is the total number of reviews for time $t$. Some simple algebra shows that the range of these scores are continuous and bounded between positive one (completely positive reviews) and negative one (completely negative reviews). Bollen, Mao, and Zeng (2011) tied this ratio to the stock market, via the Dow Jones Industrial Average, by measuring the sentiment of the public mood via Twitter. Their paper resulted in a significant prediction of the market index, with up to five days’ lead time (Bollen et al., 2011). This was accomplished by adding another dimension to the typically bivariate value of the sentiment score, a range of depth of emotion rather than just positive or negative (Bollen et al., 2011). This was done by providing a co-occurrence weighting to each word in the lexicon, so that when the word was related back to the positive or negative association the value was no longer binary (Bollen et al., 2011). This weighing of the individual sentiment scores was then summed up for
each document, and the sentiment ratio then became a range of value instead of a single ratio. This outperformed the traditional sentiment ratio considerably, providing not only a mean average percentage error of 1.83%, but an accurate directional classification of 86.7% (Bollen et al., 2011). Bollen et al (2011) then proceeded to evaluate this significance of this by checking each day's success and failure rate, then determining the likelihood of getting the number of successful results from the time period and calculating the probability from a binomial distribution, and then repeated it for all 20 of their time periods. The ending result was that there was only three percent probability that the volume of success could have occurred by chance (Bollen et al., 2011). Bollen et al's (2011) methodology has since been repeated using sentiment from the news and investors (Schneider & Gupta, 2016; Zhang et al., 2011). This significant result, using large scale public data on an indexed stock price establishes the theoretical framework for this dissertation.

**Predicting in Hospitality**

While there is limited research using sentiment analysis in the literature, it is not completely devoid of any research. Hospitality researchers have recognized the importance of online consumer reviews, and have started to evaluate the information that these reviews can provide. This is often limited in scope to the specific aspects of guest satisfaction, as in Duan, Yu, Cao, and Levy's (2015) study, or demonstrating the ability to extract the information from vast quantity of online information (Philander & Zhong, 2016). While the understanding of the consumer for hospitality has increased through this literature, and some operating practices are recommended, leveraging this vast quantity of data into an operational format and providing tools to only provide valuable business intelligence have not been fully explored. Understanding what is being said in the consumer peer space via online reviews, would provide marketers
avenues to alter or reinforce their best practices, and understanding that this carries a significant impact to the share price allows the business to potentially quantify this aspect in a meaningful way to all parties (Yu, Duan, & Cao, 2013). This is important in the casino resort, where marketing has followed a more traditional method of intuitive understanding, believing that the marketing promotions increase foot traffic and thus revenue; this is in addition to knowing the amenity's impact may have on the sentiment of the reviews, and thus the share price, which can sometimes be questionable (Lucas & Bowen, 2002; Lucas & Tanford, 2010). Operators could also know, since consumer reviews are highly influential in the purchasing decision of consumers, any potential fallout or dramatic deviation from expected visitor volume, as the sentiment can inform of what the future customer's peers are telling them regarding the property (Sparks & Browning, 2011; Sparks, Perkins, & Buckley, 2013). As researchers, the continuing impact of utilizing the vast amounts of data being shared openly, and willingly, by the public-at-large can provide a deeper understanding of how the changing technological landscape and demographics of society can either reinforce, or reinvent, the theories that much of the academic literature is predicated upon.

**Conceptual Framework and Hypothesis**

The share price of a publicly traded company is reliant upon more than just the revenue generated at the property (Nguyen, Shirai, & Velcin, 2015). Just as the purchasing intention of the hospitality consumer can be predicated on by the online review, so can the investment intention of the investor or broker. At the very center of any of these purchasing intentions, being stock purchasing or visiting the casino resort, is what is being said in the consumer reviews. As this level of information continues to grow, and becomes incrementally easy to access, investors are no longer beholden to the quarterly earnings reports of the company, but
can gauge for themselves what is being said about the different properties. Modern hospitality companies are multi-branched businesses, with properties all over the world servicing many varying levels of consumers. Is the stock price related to the majority sentiment of the business or is it beholden to the individual property's consumer's mood?

Las Vegas resort companies could additionally see additional challenges as one company may open several properties in close proximity to each other. This inherent competition between same company properties could potentially mitigate the overall negative impact from a single property experiencing a period of negatively aligned consumer reviews. While these properties may operate largely independently of each other, with their own general managers dictating the expenditure of the properties resources (service employees, marketing dollars, etc…), they still represent a part of the whole parent company and therefore experience the same pain if the stock price of the company were to experience an adverse impact. Can the consumers of one property dictate a swing in the stock price of a current hospitality company?

From these comments and reviews it is possible to predict with a reasonable amount of accuracy the directionality of an index fund, and potentially even individual non-hospitality companies that are influenced, but not as dependent upon the sentiment of consumer for their sale of goods (Bollen et al., 2011; Chong et al., 2016; L.Lima et al., 2016; Nguyen et al., 2015; Schneider & Gupta, 2016; Sparks & Browning, 2011). Would the same apply to a hospitality company, and would the lead time be sufficient for the corporate officers or investors to respond to the upcoming shift? The impact of strong negative reviews can be persistent for a time period, but the impact of them does tend to smooth out after a period of time, which leads to the question of if there is a relationship between sentiment and hospitality company stock price, is the relationship no only significant, but meaningful?
Research Questions

Stock price forecasting can either be considered a fool's errand or the Holy Grail (Mlodinow, 2008). This is because of the "noise" associated with the many influences involved with stock price. According to Mlodinow (2008), the common disclaimer for all stock traders and investment advisors is "Past performance is no guarantee of future behavior." Yet a common prediction method for stock price is to use historical data as a predictor (Nguyen et al., 2015). Since this is often the standard that new predictors can compare to, can the sentiment of hospitality consumer reviews have more of an influence on the share price?

Ultimately these become the research questions of this dissertation:

1. How does the consumer sentiment of online reviews for a gaming company impact the stock price?
2. How does the standard historical price compare in forecasting the future with the sentiment of the consumers’ reviews?

Hypotheses

The following hypotheses can then be derived from the research questions:

H1: The Company’s overall consumer sentiment from online hospitality reviews will be significant in predicting the future directionality of the company’s share price.

H2: The predictive power of the sentiment of the hospitality company’s consumer reviews will exceed the predictive power of the historical trend

Summary

In this chapter, the theoretical framework for this study was explored. The theoretical foundation in the consumer socialization theory to understanding why eWOM, and consumer reviews in particular, are relevant to the hospitality industry. The concept of sentiment analysis
was explored and explained to understand the decisions that are made regarding the criterion and methodological selection. The current body of literature was explored in how sentiment analysis is successfully used both inside and outside of hospitality, and using these results the remaining gaps and questions in the literature was identified resulting in seven testable hypotheses. The next section will discuss the methodology of this study to test these hypotheses.
Chapter 3

Methodology

This chapter describes the research design, data, semantic analytic process, prediction methodology, and validation analysis used in this study. These are intended to test the hypothesis enumerated at the end of Chapter 2. This chapter is designed as follows: the first section will describe the nature of the data collection, including collection sources and the entity focus point of the reviews, the next section will describe the process of the sentiment analysis conducted including the preprocessing methodology, the third section will describe the ARIMA model to be used to test the hypotheses, the final section will discuss the process of validation and prediction accuracy.

Data Collection

Since this research is dependent upon the consumer reviews, data was collected from multiple reviews sites to allow for the broadest scope of reviews. While the majority of the reviews are from sites similar to TripAdvisor, Expedia, and Booking.com, data from other sites were collected additionally, including Yelp and Hotels.com. This diversity of the information gathered should help to reduce the social desirability bias that could be present in online consumer reviews (Moe, Trusov, & Smith, 2011). The time period selected for sampling is December 2014 through February 2016. This sampling of the data will allow for the usage of a testing set for the sentiment analysis, December 2014, the model building time period, calendar year 2015, and then an out-of-time sample to validate the model of January and February 2016.

Focal Company

The company of focus of this research is MGM International, a publicly traded gaming hospitality company with multiple properties around the world. While this could potentially
cause issues with the stock-price, as the company's non-Vegas property profile is not as numerous as the Las Vegas properties, in 2015 there were ten Las Vegas gaming resorts compared to eight non-Las Vegas properties. Additionally, the Las Vegas properties comprised over 60% of MGM International’s adjusted EBITDA (MGM International, 2015). This would lend credence that the well-being of these properties is still of vital importance to the company as a whole. Therefore, the reviews to be considered will be reviews on the Las Vegas properties of the company. Additionally, due to limitations on the researchers, only reviews in English will be regarded as valid in the sample. Given the number of properties, and the time period, there are expected to be approximately one-hundred reviews per day. Over the course of the year, this would result in a raw data sample of approximately 36,500 reviews for the model build, exclusive of the out-of-time sample and test sample for the sentiment analysis. These reviews can range from a few to no words to several paragraphs, as demonstrated in Chapter 2. Therefore, the complexity and conformity of the data will be evaluated to validate that the review samples are similar in contextual nature. The historical stock price for this company is taken from the Yahoo! Finance page for MGM International. The closing price for the day will be differenced from the closing price of the previous trading day to calculate the dependent variable for the ARIMA regression (CP\textsubscript{t}-CP\textsubscript{t-1}) (Equation 6).

\[
DV = CP_t - CP_{t-1}
\]

(Equation 6)

**Review Collection**

The reviews are collected and provided by ReviewTrackers, a company providing online reputation management (ReviewTrackers.2016). According to the company website:

ReviewTrackers is the award-winning software that elevates the voice of the customer and enables brands to innovate based on customer feedback. The platform
empowers businesses by unlocking actionable customer intelligence that helps them manage online reviews, improve brand reputation, and make data-driven decisions that result in increased profitability. Trusted by over 30,000 businesses, ReviewTrackers is the premier customer feedback solution for enterprise businesses (ReviewTrackers.2016). Specifically, the entities within the MGM International Company that the reviews will be focused on are the Bellagio, Luxor, MGM Grand, Aria Casino Resort (shared equity with Infinity World Development), Monte Carlo, Mandalay Bay, New York New York, Circus Circus, The Mirage, and Excalibur. All the reviews received are retrieved from the individual properties' links on their respective review site and include the consumer rating, where applicable, the date of retrieval, the time of the review, and the text of the review itself. Any ratings with no review left will be discarded for the purpose of this analysis, as the star rating is not the focus of this research, rather the information left by the reviewer per Hu et al (2014).

**Data Processing and Sentiment Analysis**

The data is raw data when initially retrieved from the review site, therefore it requires some initial preprocessing prior to performing the sentiment analysis. This is in the best practices as laid forth by Miner (2012). Prior to this step, however the granularity of the analysis must be considered.

**Granularity**

If this research were to focus on Twitter Tweets, then document granularity would suffice to adequately estimate the interpreted sentiment by the reader. However, given that this research is focused on the sometimes-lengthy reviews provided on consumer reviews sites, a document and sentence dependent granularity is the preferred methodology. By using the methodology discussed in McDonald, Hannan, Neylon, Wells, and Reynar (2007) the document level rating
can inform the sentence level rating, and vice versa. This is done through their MIRA algorithm and constraining the sentence inference by the known rating left by the reviewer. If the review site does not offer the option to leave a rating, or it is not provided, then the constraint will be the number of labeling that are scored the highest in the old weight vector (McDonald et al., 2007). By modifying this methodology with the support vector machine methodology discussed by Pang et al. (2002), the true sentiment of each review would be weighed by the corresponding sentiment of the sentences, and each sentence will be weighed by the surrounding context as well.

**Sentiment Identification**

For the corpus, a dictionary based method is used to analyze the consumer reviews from the reviews sites. There is nothing inherent in hospitality reviews that would lend credence that certain words need to be adjusted from their specific dictionary. The dictionary corpus for this research will come from the WordNet database. This sentiment lexicon that is continually updated as it builds of the synonyms of adjective words, thus it determines polarity of words with similar meanings, accepts the new word then continues to grow and look for more synonyms (M. Hu & Liu, 2004; Miller & Fellbaum, 1998). This corpus was used by Philander and Zhong (2016) with significant effect, and there is no actual reason to contradict this finding.

**Data Processing**

The data will be processed in the R-Studio software package. This is an open sourced statistical software package that is open sourced and adaptable for research purposes, having been used in hospitality and other fields of research (R Core Team, 2015). While there are no known packages for the software program that will explicitly perform the MIRA algorithm, the R software allows for the researchers to code and adapt the performance to mirror that of the
algorithm. The library e1701 provides the support vector machine functions for R to be able to process the data through this methodology (Meyer, Dimitriadou, Hornik, Weingessel, & Leisch, 2015).

**Preprocessing.**

After the data is checked and loaded into the statistical software, the data must be preprocessed before it is analyzed for the sentiment contained. The process of preprocessing text data is laid out in Miner (2012), the process for this research is identified in Figure 4. The first step is to validate that the entire data was loaded into the software correctly and accurately. This is done by checking the length of the information, and the mean overall rating. Any missing data that is not removed due to no text review or due to being in a foreign language needs to be uploaded correctly. The data is then prepared to be processed. The first step will be to split each review into individual sentences. This is typically done by splitting the data at each period, but will be examined to validate that new sentences were not created. This is necessary for this research since each sentence will inform the document rating and the document rating will inform the sentence rating. After the data is split, the case of all text will be adjusted to lower case. This is to prevent the improper assignment of words due to case sensitivity. All stop words, punctuation, and whitespace will then be removed from each sentence. Stop words are the high frequency words in the English language that carry no meaning, or value to sentiment analysis (i.e. "A", "to", "in", "the", etc…) each sentence is then separated into separate tokens.

**Corpus evaluation.**

Once the data is has undergone the preprocessing, a term document matrix is created. This is a matrix of the high frequency words used in the reviews. This is examined to make any specific adjustments to the corpus, and to check for word misspellings. If word misspellings are
found, then they are corrected and another term document matrix is created. This process is
continued until there are no further identified misspellings.

**Sentiment scoring.**

Once this is all complete the training data for the sentiment analysis algorithm will be
processed and validated. By training the algorithms on a smaller data set, this can be classified
by the researcher and the correct classification rate can be calculated. Once the training of the
algorithm is completed, the analysis is then performed on the testing data. The sentiment score
for each property will be a modified version from Liu et al (2016). The daily sentiment \( S_{ip} \)
where \( i \) is the date, and \( p \) is the property) score for the property is the summation of the positive
weighted scores less the negative weighted score \( S_{POS}, S_{NEG} \) divided by the total number of
sentences (Equation 7). Meanwhile, the daily sentiment score \( S_{ic} \) for the company is the
summation of each property score divided by the total properties scored for the day (Equation 8).

\[
S_{ip} = \frac{\sum_{n=1}^{m} (s_{POS} - s_{NEG})}{n}, \text{for all } m \text{ sentences}(n) \text{ for property } p \text{ at time } i \tag{7}
\]

\[
S_{ic} = \frac{\sum_{p=1}^{10} (S_{ip})}{n_{ip}}, \text{for all } p \text{ at time } i \text{ where } n_{ip} \text{ is the number of properties at time } i \tag{8}
\]
ARIMA Analysis

After the sentiment scores have been calculated for each review for each property, the next step is to build the predictive model. Much like the sentiment analysis, this is a two-step process. The first phase involved training the model on the years' worth of data, followed by an out-of-time sample to validate the model.

Training the Model

ARIMA, or Auto-Regressive Integrated Moving Average model, is a regression technique that is designed to handle time-series data (Bowerman, O'Connell, & Koehler, 2004). In looking at the time trend of hospitality data, there are natural seasonal fluctuations that can
occur, which could cause instability in normal linear regression models or multiple regression models.

Since the stock price doesn't change on weekends, but there may be reviews with lagged dates for weekends, the scores for weekend reviews will be added to the previous trading day’s values. Since the sentiment of the company is a function of the sentiment of the individual properties, it is reasonable to assume that these two variables have a high collinearity which would invalidate the models, thus only the company score will be used. For hypothesis 2 the coefficient for the sentiment score must be an operational value as well as the standard error being of greater value than the historical time lagged trend from the ARIMA model.

Each model must be evaluated for the time series stationarity, no significant serial correlation through the acfs, pacfs and Box-Jenkins testing, normality of the residuals, and no evidence of heteroscedasticity. A year is used to train the models to account for all holidays and any other year seasonal trends that could occur. Holidays will be coded into the model as binary indicator variables. An appropriate trend variable is added to make the dependent variable stationary over time, and day of week variables with Friday being selected as the base time period. Friday was selected as the base time as it is the end of the trading week. Additionally, the incorporation of a universal “index” was selected in the form of the NASDAQ stock index. This will be used to measure the influence of investor and consumer confidence as it is generally accepted as a proxy for these constructs. The NASDAQ Index was selected over the Dow Jones Industrial Average to prevent the incorporation of possible multicolinearity into the modeling process as MGM International is traded on the New York Stock Exchange, and the NASDAQ Index is a composite index of stock only traded on the NASDAQ stock exchange.
Testing the Model- Prediction Accuracy and Validation

Once the final models have been trained on the testing data, the ARIMA models are then used to predict the directionality of the stock price on the out-of-time validation samples. This validation of two months represents at 39 out-of-time time periods. While this is an extended time period to forecast, this would prove of sufficient size to be able to calculate the model validations. For a time series model, the \( R^2 \) is less useful than the AIC/BIC values which identify if the model is over fit to the data, additionally, since this research is focused on prediction the values that are used for validation are the MSE and MAPE. The correct classification rate will also be calculated. If the MSE and MAPE are of an acceptable error rate, then the classification rate will be used to determine the likelihood of the results due to pure chance. This will be done in a similar manner as Bollen, Mao, and Zheng (2011). Since there are only two options on the results of the prediction, success or failure, a binomial probability test should yield the likelihood of successfully predicting the correct number of days. A probability of less than 5% probability of having the same amount of success over the 60 trials would be the threshold for this research.

Limitations and Potential Errors

The first limitation is that this data is evaluated only in the English language. While taking into consideration China, and its impact on MGM, this limitation could affect the final results on the analysis. Additionally, though the extant literature has identified a relationship between the general English speaking population's sentiment and stock price, and between investor's sentiment and stock price, that is not to say that there is direct causation. The scope of this research is not to identify the causal relationship, or the latent variables, that may cause both to move in the same direction. Rather this research is to show the applied usage of sentiment
analysis as a predictive tool, to encourage industry to invest in the sentiment analysis tools, and researchers to continue to improve upon the methodology regarding sentiment analysis and other natural language processing. Improper classification of the sentiment score could pose a threat to the internal validity, but would be subject to extensive review to validate in addition to the training data set. The reliability of this research is tested out to thirty-nine time periods, and therefore it is believed that this study should be reliable if replicated within the hospitality industry. However, since this study only considers one publicly traded gaming property, and focuses on the company's Las Vegas properties, it is not generalizable. It is the researcher's hope that this study will be replicated and applied to other populations, languages, and industries beyond hospitality.

**Summary**

This section discussed the detailed methodology for this study. The considerations for data collection, preprocessing of the data, performing the sentiment analysis, creation of the ARIMA model, and finally the testing and validation of the model's predictive ability. This was all discussed to address the specific hypotheses and research questions of the study. The results of the analysis and prediction modeling are discussed in the next chapter.
Chapter 4

Results

This chapter will discuss the details of the analysis for this study. The general data on the reviews is discussed first, followed by the sentiment analysis on them. The building of the ARIMA model on the in-time sample, including the selection of lag and model selection will be discussed second. Finally the out of time sample, and the model predictions will be discussed and evaluated to test the following hypothesis:

H1: The Company’s overall consumer sentiment from online hospitality reviews will be significant in predicting the future directionality of the company’s share price.

H2: The predictive power of the sentiment of the hospitality company’s consumer reviews will exceed the predictive power of the historical trend.

Consumer Reviews

The handling of the consumer review data was broken down into three steps. The first step was the gathering and cleaning of the reviews. This removes any non-informational reviews that would not be pursuant to this research. The data is then loaded into the R software package, and further cleaned. This step includes updated the corpus for relevant terms. Finally, the training set was scored and the training algorithm was applied to the remaining dataset.

Loading the Data

The reviews were collected from Review Trackers, an online reputation management company (ReviewTrackers). For the sentiment testing time period, the in-time sample, and the out of time validation there were a total of 74,955 different reviews supplied. Not all the reviews contained text information, some were simply star ratings, and some reviews were not in English. The package textcat in the R software package was used to assign the initial assignment
of language to each review. Any review with all symbols, or with no text in the review field was discarded. Appendix 2 contains the table with the volume of reviews by source and language.

Any non-English review was manually checked to verify and removed from the dataset. This left 59,861 reviews over the 456 days in the dataset. Figure 5 shows the daily distribution of the reviews. The largest number of reviews came around New Year’s Eve and the Christmas Holidays, which is one of the busiest times in Las Vegas. With the data properly prepared, the next step of the sentiment analysis was to load the final group of reviews into the software.

![Volume of Reviews](image)

**Figure 5.** Distribution of reviews by day.

**Cleaning and Scoring the Data**

Using a base corpus, it was run through WordNet to increase the volumes of positive words and the volume of negative words. The all reviews from December 2014 were then loaded into the scoring algorithm which decomposed each review into individual sentences, removed all English stop words, excess punctuation, and remaining white space. Each sentence
was then tokenized to the root word and grouping into corresponding uni-grams, bi-grams, and tri-grams. These words were then loaded into the Document Term Matrix and then scored by matching the corpus words with the words in the matrix. The sum of the scoring developed the score for the sentence. As each review was then compiled and the total score recorded for the review according to Equation 7. Table 1 contains the descriptive statistics on the scored reviews. The 4,344 scored reviews were then condensed into daily scores through the use of Equation 8. These were then used as the training set for the SVM algorithm to score the remaining reviews.

Table 1

Descriptive Statistics on Sentiment Scores

<table>
<thead>
<tr>
<th>Score</th>
<th>Mean</th>
<th>Standard Error</th>
<th>Median</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sentiment Score</td>
<td>0.449</td>
<td>0.0095</td>
<td>0.3810</td>
<td>-4.0000</td>
<td>7.0000</td>
<td>4344</td>
</tr>
<tr>
<td>(Property)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sentiment Score</td>
<td>6.368</td>
<td>0.3213</td>
<td>5.5020</td>
<td>3.4528</td>
<td>11.4947</td>
<td>31</td>
</tr>
<tr>
<td>(Daily)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sentiment Score</td>
<td>5.974</td>
<td>0.1213</td>
<td>5.8202</td>
<td>1.0769</td>
<td>31.9156</td>
<td>425</td>
</tr>
<tr>
<td>(2015 &amp; 2016)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

After the scoring is complete the processing is done for the predictor variable for the ARIMA process.

**ARIMA Modeling**

The ARIMA modeling required the identification of the appropriate lag for the best correlation between the predictor variable (score) and the dependent variable (stock price). This lag would represent the potential time differential between the posting of the reviews and any subsequent impact on the stock price. The closing price of MGM International’s stock price was obtained and an algorithm was written to process the correlation of the score to the stock price.
over 30 future lags. This is to say that the score at time $t_0$ would potentially have a correlation with the stock price at time $t_{\text{lag}}$ where $\text{lag}$ is potentially any value in the algorithm between 1 and thirty (Table 2). It was discovered in this process that the non-trading days of the weekends and holidays were confounding the correlation. As a result, the decision was made to remove the weekends and non-trading holidays from consideration. Any scores for these days were placed on the previous trading day. This removed both Saturdays and Sundays and many holidays from possible inclusion into the ARIMA model as there was no value possible for these. The requirement for the ARIMA to have a continuous time period is satisfied through the understanding that it is a continuous time period in respect to trading days as opposed to calendar days.

**Defining the Model**

Prior to building the complete model on the in time sample, 2015, the dependent variable must be examined to satisfy the requirements for an ARIMA, or more specifically ARIMAX since this research is testing non-time predictors, regression. The assumptions of ARIMA are additional to the assumptions of multiple linear regression as ARIMA is a variation on the traditional multiple linear regression. Additional models were evaluated using dummy variables for the quarterly earnings releases and using a differentiated transformation on the NASDAQ Index (Index) variable. These variables were selected to help explain potential variance due to the quarterly earnings statements and through the differentiated shifts of the NASDAQ. However, since these variables and transformations provided no model improvement, and demonstrated no significance, it was determined that these variables and transformations were not impactful to this research. Additionally, the Dow Jones Gambling index (DJUSCA) contains
MGM Internal stock which would place the dependent variable as a predictor and therefore was also excluded from consideration for the model.

Table 2

<table>
<thead>
<tr>
<th>Lag</th>
<th>Pearson's Correlation coefficient</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-0.13</td>
<td>0.04</td>
</tr>
<tr>
<td>2</td>
<td>0.09</td>
<td>0.14</td>
</tr>
<tr>
<td>3</td>
<td>0.01</td>
<td>0.82</td>
</tr>
<tr>
<td>4</td>
<td>-0.04</td>
<td>0.52</td>
</tr>
<tr>
<td>5</td>
<td>-0.07</td>
<td>0.26</td>
</tr>
<tr>
<td>6</td>
<td>0.00</td>
<td>0.96</td>
</tr>
<tr>
<td>7</td>
<td>0.05</td>
<td>0.42</td>
</tr>
<tr>
<td>8</td>
<td>-0.10</td>
<td>0.13</td>
</tr>
<tr>
<td>9</td>
<td>0.01</td>
<td>0.91</td>
</tr>
<tr>
<td>10</td>
<td>0.00</td>
<td>0.96</td>
</tr>
<tr>
<td>11</td>
<td>-0.07</td>
<td>0.28</td>
</tr>
<tr>
<td>12</td>
<td>0.03</td>
<td>0.68</td>
</tr>
<tr>
<td>13</td>
<td>-0.01</td>
<td>0.88</td>
</tr>
<tr>
<td>14</td>
<td>0.05</td>
<td>0.45</td>
</tr>
<tr>
<td>15</td>
<td>-0.10</td>
<td>0.13</td>
</tr>
<tr>
<td>16</td>
<td>-0.03</td>
<td>0.68</td>
</tr>
<tr>
<td>17</td>
<td>0.04</td>
<td>0.49</td>
</tr>
<tr>
<td>18</td>
<td>-0.02</td>
<td>0.72</td>
</tr>
<tr>
<td>19</td>
<td><strong>0.15</strong></td>
<td><strong>0.02</strong></td>
</tr>
<tr>
<td>20</td>
<td>-0.07</td>
<td>0.27</td>
</tr>
<tr>
<td>21</td>
<td>-0.08</td>
<td>0.23</td>
</tr>
<tr>
<td>22</td>
<td>0.07</td>
<td>0.27</td>
</tr>
<tr>
<td>23</td>
<td>0.09</td>
<td>0.18</td>
</tr>
<tr>
<td>24</td>
<td>-0.04</td>
<td>0.57</td>
</tr>
<tr>
<td>25</td>
<td>-0.13</td>
<td>0.05</td>
</tr>
<tr>
<td>26</td>
<td>-0.01</td>
<td>0.86</td>
</tr>
<tr>
<td>27</td>
<td>0.08</td>
<td>0.24</td>
</tr>
<tr>
<td>28</td>
<td>-0.10</td>
<td>0.10</td>
</tr>
<tr>
<td>29</td>
<td>0.10</td>
<td>0.10</td>
</tr>
<tr>
<td>30</td>
<td>0.06</td>
<td>0.38</td>
</tr>
</tbody>
</table>
The first, and an often defining, assumption is the stationarity of the mean of the dependent variable. This was checked through looking at the stock price (StockPrice) plotted out over the time interval (Figure 6). The data does not have a stationary mean, and therefore unit roots would have to be identified to identify the level of integration needed to create stationarity. This, however, would only happen if the regression is focused on the point accuracy of the dependent variable. The focus of this research is the directional shift therefore the percent difference was calculated. The values of the percent difference over time (Figure 7) shows that the mean of the percent difference of the stock price is stationary. Since the dependent variable needs to be lagged in respect to the sentiment score, a greater time period of closing stock price was obtained (Figure 8). This allowed for the usage of all sentiment scores, and helped to not lose data to the calculation of the percent difference change as well.

Figure 6. Plot of stock price over time
**Figure 7.** Plot of percent price differential in stock price over time.

**Figure 8.** Plot of entire percent differential of stock price over time.
The lagged difference was then evaluated through a multiple linear regression model first to determine the appropriateness of the variables selected for collinearity. All the variables evaluated were below an acceptable threshold of a VIF 10 (Table 3) (Repetti, 2013). The ACF and PACF graphs were evaluated and spikes were noticed that could represent the need for auto-regressive and/or moving average terms to be incorporated into the model (Figure 9 and Figure 10).

**Figure 9.** ACF graph of residuals from linear regression.

**Figure 10.** Partial ACF graph of residuals from linear regression.
As the AIC is the best indicator for parsimony and goodness of fit for ARIMA modeling, the potential lags were each evaluated for their AIC value. The lowest AIC value with no errors in the regression was selected. Since the percentage differential in stock price has a stationary mean, there was no need to incorporate an integrated value into the model. The ARIMA model (3,0,8) returned the lowest AIC of 1120.065. Two models with lower value returned (4,0,9 and 9,0,9) convergence issues in the optim algorithm used in the ARIMA processing, and therefore were forced to be discarded as the model is unstable at convergence (Table 4).
Table 4

AIC values for AR and MA terms from 0 to 9

<table>
<thead>
<tr>
<th>Auto-Regressive Values</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1174.436</td>
<td>1175.581</td>
<td>1174.385</td>
<td>1175.994</td>
<td>1177.495</td>
<td>1179.259</td>
<td>1161.732</td>
<td>1159.043</td>
<td>1160.507</td>
<td>1161.039</td>
</tr>
<tr>
<td>1</td>
<td>1175.774</td>
<td>1174.725</td>
<td>1175.724</td>
<td>1177.713</td>
<td>1156.293</td>
<td>1157.68</td>
<td>1158.039</td>
<td>1159.521</td>
<td>1161.521</td>
<td>1162.699</td>
</tr>
<tr>
<td>2</td>
<td>1174.991</td>
<td>1175.761</td>
<td>1169.481</td>
<td>1177.55</td>
<td>1177.229</td>
<td>1160.203</td>
<td>1127.177</td>
<td>1161.154</td>
<td>1163.019</td>
<td>1164.703</td>
</tr>
<tr>
<td>3</td>
<td>1176.51</td>
<td>1177.74</td>
<td>1171.407</td>
<td>1129.683</td>
<td>1126.175</td>
<td>1127.711</td>
<td>1129.264</td>
<td>1120.121</td>
<td>1120.065</td>
<td>1120.323</td>
</tr>
<tr>
<td>4</td>
<td>1177.931</td>
<td>1179.607</td>
<td>1125.408</td>
<td>1152.422</td>
<td>1122.572</td>
<td>1123.596</td>
<td>1128.929</td>
<td>1131.81</td>
<td>1124.419</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1179.899</td>
<td>1181.594</td>
<td>1127.406</td>
<td>1128.922</td>
<td>1132.62</td>
<td>1135.564</td>
<td>1131.06</td>
<td>1132.38</td>
<td>1132.885</td>
<td>1120.479</td>
</tr>
<tr>
<td>6</td>
<td>1179.756</td>
<td>1151.041</td>
<td>1178.85</td>
<td>1129.586</td>
<td>1128.258</td>
<td>1130.674</td>
<td>1132.238</td>
<td>1128.376</td>
<td>1135.968</td>
<td>1135.07</td>
</tr>
<tr>
<td>7</td>
<td>1180.663</td>
<td>1182.913</td>
<td>1154.896</td>
<td>1157.028</td>
<td>1130.123</td>
<td>1133.527</td>
<td>1127.25</td>
<td>1133.426</td>
<td>1134.483</td>
<td>1135.591</td>
</tr>
<tr>
<td>8</td>
<td>1182.133</td>
<td>1152.441</td>
<td>1130.825</td>
<td>1132.982</td>
<td>1131.965</td>
<td>1134.109</td>
<td>1128.877</td>
<td>1129.522</td>
<td>1131.31</td>
<td>1133.673</td>
</tr>
<tr>
<td>9</td>
<td>1182.78</td>
<td>1151.561</td>
<td>1132.595</td>
<td>1134.524</td>
<td>1127.267</td>
<td>1134.286</td>
<td>1131.181</td>
<td>1140.347</td>
<td>1134.017</td>
<td></td>
</tr>
</tbody>
</table>

The ACF and PACF of the model had an apparent lag at the tenth lag (Figure 11 and Figure 12).

![Figure 11. ACF graph of residuals from ARIMA regression (3,0,8).](image-url)
Figure 12. Partial ACF graph of residuals from ARIMA regression (3,0,8).

As a result a Box-Jenkins test was performed on up to thirty lags (Figure 13). No p-value suggested the presence of any remaining auto-correlation or partial correlation with the residuals. The residuals were also placed on a Q-Q plot and with a normal trend-line, and no significant deviation from normality was noticed (Figure 14).
Figure 13. p-Value of Box-Jenkins test.

Figure 14. Q-Q plot of residuals from ARIMA test.
With the model parameters defined, and the lowest AIC selected, the model was run on the training data set of 2015. The overall model had a training set measurement errors of a RMSE = 1.68628 and the Score variable has a p-value of 0.002 (Table 5). While a one point in score change only represents an eleven basis point change in the percent difference of the stock price, the fact that this is significant supports H1, with a significant p-value the null is able to be rejected in favor of H1, given the accuracy of the prediction.

Table 5

*Output from ARIMA Regression*

<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Coeff</th>
<th>S.E.</th>
<th>VIF</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Score</td>
<td>0.113</td>
<td>0.037</td>
<td>1.9</td>
<td>0.00***</td>
</tr>
<tr>
<td>Index</td>
<td>-0.002</td>
<td>0.002</td>
<td>2.9</td>
<td>0.34</td>
</tr>
<tr>
<td>Mon</td>
<td>0.912</td>
<td>0.603</td>
<td>2.5</td>
<td>0.13</td>
</tr>
<tr>
<td>Tues</td>
<td>0.905</td>
<td>0.584</td>
<td>2.4</td>
<td>0.12</td>
</tr>
<tr>
<td>Wed</td>
<td>1.399</td>
<td>0.600</td>
<td>2.5</td>
<td>0.02**</td>
</tr>
<tr>
<td>Thurs</td>
<td>0.928</td>
<td>0.569</td>
<td>2.3</td>
<td>0.10*</td>
</tr>
<tr>
<td>Jan</td>
<td>-0.555</td>
<td>0.986</td>
<td>3.4</td>
<td>0.57</td>
</tr>
<tr>
<td>Feb</td>
<td>-0.748</td>
<td>0.639</td>
<td>2.4</td>
<td>0.24</td>
</tr>
<tr>
<td>Mar</td>
<td>-0.598</td>
<td>0.368</td>
<td>2.2</td>
<td>0.10*</td>
</tr>
<tr>
<td>Apr</td>
<td>-0.601</td>
<td>0.390</td>
<td>2.0</td>
<td>0.12</td>
</tr>
<tr>
<td>May</td>
<td>-0.502</td>
<td>0.215</td>
<td>1.9</td>
<td>0.02**</td>
</tr>
<tr>
<td>Jun</td>
<td>-0.080</td>
<td>0.305</td>
<td>1.9</td>
<td>0.79</td>
</tr>
<tr>
<td>Jul</td>
<td>0.470</td>
<td>0.217</td>
<td>2.0</td>
<td>0.03**</td>
</tr>
<tr>
<td>Aug</td>
<td>-1.118</td>
<td>0.367</td>
<td>2.2</td>
<td>0.00***</td>
</tr>
<tr>
<td>Sep</td>
<td>-0.003</td>
<td>0.904</td>
<td>3.0</td>
<td>1.00</td>
</tr>
<tr>
<td>Oct</td>
<td>-0.282</td>
<td>0.514</td>
<td>2.4</td>
<td>0.58</td>
</tr>
<tr>
<td>Dec</td>
<td>-0.890</td>
<td>0.399</td>
<td>2.0</td>
<td>0.03**</td>
</tr>
<tr>
<td>NYE</td>
<td>-0.166</td>
<td>1.352</td>
<td>1.1</td>
<td>0.90</td>
</tr>
<tr>
<td>CNY</td>
<td>4.413</td>
<td>1.952</td>
<td>1.1</td>
<td>0.02**</td>
</tr>
<tr>
<td>Vets</td>
<td>0.073</td>
<td>2.459</td>
<td>1.1</td>
<td>0.98</td>
</tr>
<tr>
<td>AIC</td>
<td>1120.06</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RMSE</td>
<td>1.968628</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Order</td>
<td>3,0,8</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note. *p-value <.10. **p-value <.05. ***p-value <.01.*

61
Testing the Model

The next step involved to predict the next 39 time periods. The sign of each prediction was then used to determine the directional shift of the stock price. While the time $t_0$ was for January and February the prediction represent the future change in the stock price at a 19 business day lag. This was then compared to the actual changes in the stock price. The confusion matrix (Table 6) shows that 28 of the 39 time periods were correctly identified.

Table 6

Confusion Matrix of Prediction

<table>
<thead>
<tr>
<th>Prediction</th>
<th>Direction Change</th>
<th>Negative</th>
<th>Positive</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Negative</td>
<td>12</td>
<td>6</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>Positive</td>
<td>5</td>
<td>16</td>
<td>21</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>17</td>
<td>22</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The misclassification rate is only .28, with an accuracy rate of .718. The true positive and true negative rates are not consequential, as the accuracy of the directionality is the focus of this study. Since the uninformed probability of correctly guessing the directional shift of a stock price is .5, a binomial distribution calculation was used to determine the probability of being able to accurately guessing 28 or more time periods of 39 correct (Equation 9).

$$1 - \sum_{i=0}^{27} \binom{39}{i} \cdot 0.5^i \cdot 0.5^{39-i}$$

(9)
The probability of doing this was calculated to be .0047. With such a low probability, forty-seven basis points, the null was rejected that the correct predictions were due to chance. Given that the significance of the sentiment scoring and that the Index variable did not have a significant p-value in the final ARIMA model, there is reason to reject both the null hypotheses in favor of $H_1$ and $H_2$.

**Conclusion**

In this chapter, the results of the analysis were discussed. The decisions made to lag selection, the transformation of the stock price, and model selection were all justified using valid statistical techniques. The model was evaluated and found to predict the future directional change of the stock price beyond the likelihood due to random chance. The next section will discuss the potential impacts, limitations, and conclusions.
Chapter 5

Conclusion

This chapter presents the key findings of this research, including the major implications. The contributions, managerial and theoretical, will be discussed. This research provides insight into the application of sentiment analysis as a predictor of price, and including these findings into predictive modelling. Any limitations that may apply to this research will be detailed. Finally, ideas for future research will be addressed.

Discussion of Findings

When Ward (1974) put forth the consumer socialization theory, it was done so to understand the relationship between the influence the growing media of television and radio marketing has on children. As this theory carries into the twenty-first century, the growing media platform is more digital. In Ward’s time the consumer social group was geographically limited to friends in the same city or town. As society becomes digitally connected, this restriction is removed and the consumers’ peer group expands dramatically. Sparks and Browning (2011) identified this, as the current generation moving into the marketplace has the largest peer group of any generation prior. This makes the eWOM influence imminently of greater importance. The juggernaut like influence of social media reviews has impacted the consumer socialization theory, as Wang, Yu, and Wei (2002) theorized, and is making a larger impact in the purchasing decision of modern consumers. This influence on the purchasing decisions of eWOM is present in the literature as Book, Tanford, Montgomery, and Love (2015) demonstrated that the deciding factor for purchase was what other consumers say, even more so than price. Anderson (2012) explored this relationship even further, tying reviews to a company’s commonly used performance metrics.
This relationship between what consumers leave on review websites and the firm performance is the focal research question of this study, but associated questions were generated through further literature review. From these research questions, two hypotheses formed and were then tested. Table 7 shows the hypothesis and the resulting decision as inferred by the testing conducting in this dissertation.

Table 7

**Hypothesis Findings**

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Company's overall consumer sentiment from online hospitality reviews will be significant in predicting the future directionality of the company's share price</td>
<td>Supported</td>
</tr>
<tr>
<td>The predictive power of the sentiment of the hospitality company’s consumer reviews will exceed the predictive power of the historical trend</td>
<td>Supported</td>
</tr>
</tbody>
</table>

**Consumer Reviews’ Influence on Stock price**

The first result of this study finds that there is a statistically significant relationship between the sentiment value of consumer reviews and that of the share price. What, and how, people share their experience about hospitality impacts the company in many ways (Anderson, 2012). This influence then carries through to the share price of the stock at a later date. This is possibly because of the time lag between when consumers read the reviews, and the penultimate point of realizing the outcome of the purchasing decision (Anderson, 2012). While the finding of this survey do not fully support those of Xie, Zhang, and Zhang (2014), as the coefficient of
the regression for the sentiment score was only eleven basis points, depending on the company this could represent the weighted value of RevPAR for the purchasing time. While an eleven basis point coefficient may not seem very large, for stocks that move in small increments this could determine an increase or decrease for the day. Similarly, as consumer reviews are essentially “public” accessible data, investors following a hospitality company could be influenced by strongly negative reviews and decide that the future well-being of the company may be in jeopardy, as customer satisfaction is the ultimate product of a hospitality company (Yu, Duan, & Cao, 2013).

**Historical Trend**

In Mlodinow (2008), he states that the creed of all financial advisors is the past performance does not predict future returns. This research sought to test this widely held belief, which Nguyen, Shirai, and Velcin (2015) describe as being a common occurrence. Through the usage of the NASDAQ index, there is an ability to get a sense of the stock price at time \( t_0 \). The insignificance of this predictor in the final model (p-value =.34), would infer that the historical price has no bearing on the final change in values. While this supports the hypothesis, a multitude of other economic factors could impact the stock price instead of investor or consumer confidence, for which NASDAQ is often a suitable proxy. The failure of the index to be significant at a nineteen day lag in the final model would cast doubt of the validity of the predictions described by Nguyen et al (2015). The final model, however, did include three autoregressive terms. These terms represent the three previous time periods for the dependent variable. The inclusion of these terms demonstrates that the historical differentials in price does carry some influence on the final time period dependent variable. This would support the forecasts that Nguyen et al (2015) describe, but the power of these is limited to only three time
periods. This would be insufficient time for the operators to respond, additionally the error terms require a greater lead time to forecast the current time prediction, with eight moving average terms. Finally, the hypothesis for this research is fully supported as the model was evaluated without the predictor variable and the lowest AIC model was of the order (0,0,0), meaning no autoregressive or moving average terms. Thus, the historical trend only applies when in the context of the sentiment score of the company, supporting hypothesis two.

**Managerial Implications**

The importance of the findings of this research may be of great importance to the hospitality industry. Currently industry professionals are aware of the importance of eWOM, per Litven, Goldsmith, and Pan (2008), but the implications of negative reviews may not be fully realized. The fact that these reviews have a relationship with the directionality of the company’s public stock price, should give reason for most operators to pay closer attention to the reviews left by their customers. Additionally, as MGM International has multiple properties, each with a potentially different service level expectation, a closer examination of the diversification of the company can now be evaluated. For example, if negative reviews from one aspect of the company, perhaps one that services a more budget minded traveler, are resulting in a lower average score, then that property could be looked at in greater detail so that the business decision of if it is profitable to continue to operate in that space. The operator’s first priority is to preserve shareholder equity, and regardless of the attachment to tradition or history, if a property is resulting in reduced shareholder equity a revision or removal should be evaluated. Further, detailing the service failures that caused the poor review can be enumerated and action plans to recover can be planned in advance of the predicted share turn.
Marketing managers should look at this research and see opportunity for what could essentially be determined as free advertising. This dissertation can provide support for the brand or marketing manager to ask for resources to further track, develop, and analyze these consumer reviews, and then recycle them into the marketing material. Using data mining techniques social networks can be mined and reviews from friends can be pushed into the marketing material. This could result in more efficient and impactful marketing materials for the company, given the positive reviews. The identification of the time-lag could allow managers to quantify the impact from marketing efforts, capital investments, or other sources that have traditionally been “unable to measure the impact.” Since the impact could potentially be realized through the consumer reviews.

**Response Time**

The index stock used in this study, NASDAQ, was acceptable for a time $t_0$ prediction, as it was significant in the simple non-lagged linear model. This however would give investors or operators no time to respond to any forecast. The NASDAQ and NYSE both close at the same time, and therefore it would be impossible to act upon the information gained from forecasting using the same day’s NASDAQ alone. One would need to forecast the NASDAQ with point accuracy as well as the individual company’s share price, and the model built upon a model would have cumulative errors. This would make any predicative modeling unstable at best, and deceptive at worst. To combat this, the lagging of the stock price in relation to the sentiment score allows the modeler to utilize information for a future time period. This supported the findings in Yu et al (2013), which identified a single day lag. While this study found that the single day lag carried a statistically significant correlation with percent change in the stock price ($r=-.13$, $p$-vale=.04). The 19-business day lag had a higher correlation and was still significant
While the single day lag supports the findings of Yu et al (2013), the nineteen-business day lag provides a greater opportunity for operators to respond to the model, and make strategic plans. This is not to say that Yu et al (2013) did not provide a deeper understanding, simply that the final audience for the research is different. While Yu et al (2013) sought to inform investors of the influence of consumer reviews, this research seeks to inform both investors and operators of the company.

This lag would also allow for the prediction of future impacts from a crisis management standpoint. In an unfortunate event that could result in massive negative consumer reviews, the company is able to prepare themselves and their shareholders for the event, or, even better, prevent such an occurrence due to the awareness of the relationship from consumer reviews.

**Theoretical Implications**

This research increased the body of knowledge regarding the methodology and application of natural language processing. The methods described and performed in this dissertation demonstrate the ability, and value, of exploring non-traditional sources of data and using big data analytical techniques, help to draw inferences modernizing many theories in hospitality and business. The inclusion of eWOM and social media into the consumer socialization theory, as advocated by Wang et al (2012), is supported through this research. As technology continues to advance, academic resources should be used to observe if the theories of the business and hospitality world continue to apply, or needs modification. By utilizing technological techniques and big data analytical tools, researchers are no longer limited in time to evaluate large quantities of qualitative data. This could advance academic research to use qualitative data in quantitative manners, and develop ideas for the construct that drive the consumer behavior.
Limitations

As with all research, there exists certain limitations to this study. The first, and foremost, is the generalizability of the findings of this research. This research was conducted only on one hospitality gaming company in Las Vegas, NV. While the methodology is applicable in virtually any and every industry, to which online reviews are accessible, the specific findings may not apply to all other companies inside or outside of Las Vegas.

Varying results could be due to the nature of the gaming industry, and the consequential aspect that the majority of players may lose money. Additionally, the Las Vegas market is strongly driven by tourism which might differ in other geographic locations. Findings for other hospitality gaming or non-gaming hospitality companies may vary.

Additionally, the findings are limited chronologically, as the model window did not include the development or addition of new properties to the MGM International portfolio. Likewise, the long term feasibility of the model development was not tested, and the model re-development time period not identified. During the model development time period, there were no large announcements made by MGM International, and so the inclusion of these press releases could not be incorporated into the model. Other macro-economic events, likewise, were attempted to be represented through the NASDAQ Index, but there was no indication of major economic shocks present. As an example, the recent presidential race dramatically stands to alter the economic environment, but this could not be predicted in the model developed.

Finally, on the topic of representativeness, this study only used English reviews. If this study was re-performed in another culture, or in another language, the result may vary greatly due to cultural norms, or the implications of many words. In the same concept of limitations, the
codex used to identify the sentiment scores may not have been the comprehensive word base to utilized, and the any improvement in the sentiment scoring process will improve the results.

**Future Research**

The remarkably thin volume of the extant literature on sentiment analysis, and the incorporation into the hospitality research stream, creates ample opportunity for future research to be developed. The first line of inquiry that could stem from this research is to relate the sentiment score to the gross revenue of a property. While the directional shift predictions provided in this dissertation are valuable and novel in nature, the ability to predict with point accuracy the future gross revenue of a company would improve the revenue management functions of a company dramatically.

Further enhancing the natural language processing algorithms used in research would additionally provide growth to this infancy of this research stream in hospitality. This could include developing a truly comprehensive lexicon of n-grams that may be unique to the industry so that any sentiment scoring can be done with even greater accuracy.

Deeper studies into the results presented here, at an aspect term level, should also be performed by future research. Challenging the full-service model of the modern gaming hotel and casino could be done by evaluating which, if any, areas are weighed in the sentiment score to determine the predictive ability. While doing this again the stock price would provide insight, performing this against the gross revenue would allow researchers to inform industry operators on areas that are true revenue generators from a customer perspective as opposed to the business perspective. In other words, what matters most to the consumer and how can the business capitalize on this preference.
Finally, utilizing the findings of this study, future studies into the relationship between the sentiment scores of the reviews and consumer purchasing intention can be elaborated upon. What other aspect are truly involved, and why is there the nineteen business day lag. What impact does having to “walk” a guest to a sister property have on the review and therefore on the relationship? Are comparison reviews, between competing firms, impactful to both firms?

Summary

As technology continues to advance and consumers become even further interconnected, the disparate impact this new peer group has on the socialization of the consumer cannot be ignored. While the issue of technology presents itself in changing viewpoints and behaviors of the consumer, it also presents itself as an opportunity to the researcher and operator of hospitality companies. eWOM is both rich in data, and meaning direct and unfiltered from the consumer. The next generation of consumers have already identified the value of the information contained in these reviews, and are ready to act upon the recommendation of those who have already had the experience.

This study ventured to quantify the value of these reviews by finding the relationship between them and the share price of the company. This chapter reviewed the findings of the study and identified important practical and academic implications for hospitality, and general business, research. The results suggest that there is a definitive relationship between the content, or sentiment, of consumer reviews and any directional shift of a company’s stock price. While previous literature examined how this information influenced the consumer, the direct impact to the company shareholder means that all levels of the business should familiarize themselves with the information. Similarly, it suggests an evolution of classical theories of consumer behavior that provide the insight into why this relationship exists.
Appendix 1

UNLV Social/Behavioral IRB - Administrative Review
Notice of Excluded Activity

DATE: April 13, 2017

TO: Ashok Singh, Ph. D
FROM: UNLV Social/Behavioral IRB

PROTOCOL TITLE: [1057002-1] Price Prediction: Determining changes in stock pricing through sentiment analysis

SUBMISSION TYPE: New Project

ACTION: EXCLUDED - NOT HUMAN SUBJECTS RESEARCH

REVIEW DATE: April 13, 2017

REVIEW TYPE: Administrative Review

Thank you for your submission of New Project materials for this protocol. This memorandum is notification that the protocol referenced above has been reviewed as indicated in Federal regulatory statutes 45CFR46.

The UNLV Social/Behavioral IRB has determined this protocol does not meet the definition of human subjects research under the purview of the IRB according to federal regulations. It is not in need of further review or approval by the IRB.

We will retain a copy of this correspondence with our records.

Any changes to the excluded activity may cause this protocol to require a different level of IRB review. Should any changes need to be made, please submit a Modification Form.
If you have questions, please contact the Office of Research Integrity - Human Subjects at IRB@unlv.edu or call 702-895-2794. Please include your protocol title and IRBNet ID in all correspondence.
## Appendix 2

### Table 8

**Distribution of Reviews**

<table>
<thead>
<tr>
<th>Language</th>
<th>Agoda</th>
<th>BBB</th>
<th>BestDay</th>
<th>Booking.com</th>
<th>Citysearch</th>
<th>Expedia</th>
<th>Facebook</th>
<th>GoGo Bot</th>
<th>Google</th>
<th>Google+ Local</th>
<th>HolidayCheck</th>
<th>HotelClub</th>
<th>Hotels.com</th>
<th>Merchantcircle</th>
<th>Orbitz</th>
<th>Priceline</th>
<th>TripAdvisor</th>
<th>TripAdvisor Q&amp;A</th>
<th>Venere</th>
<th>VirtualTourist</th>
<th>Yahoo Travel</th>
<th>Yellowpages</th>
<th>Yelp</th>
</tr>
</thead>
<tbody>
<tr>
<td>english</td>
<td>195</td>
<td>9</td>
<td>11</td>
<td>1765</td>
<td>5</td>
<td>2528d</td>
<td>540</td>
<td>37</td>
<td>1833</td>
<td>131</td>
<td>9</td>
<td>14</td>
<td>5904</td>
<td>2866</td>
<td>1</td>
<td>1863</td>
<td>480</td>
<td>1237</td>
<td>34</td>
<td>27</td>
<td>1802</td>
<td>1692</td>
<td>1</td>
</tr>
<tr>
<td>esperanto</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>estonian</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>swedish</td>
<td>203</td>
<td>2</td>
<td>1</td>
<td>2010</td>
<td>2</td>
<td>218</td>
<td>1</td>
<td>1</td>
<td>15</td>
<td>17</td>
<td>1</td>
<td>8</td>
<td>170</td>
<td>2123</td>
<td>1</td>
<td>27</td>
<td>16</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>278</td>
<td>179</td>
<td>1</td>
</tr>
<tr>
<td>danish</td>
<td>11</td>
<td>11</td>
<td>10</td>
<td>11</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>1</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>french</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>1</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
</tr>
<tr>
<td>irish</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>norwegian</td>
<td>49</td>
<td>49</td>
<td>49</td>
<td>49</td>
<td>49</td>
<td>49</td>
<td>49</td>
<td>49</td>
<td>49</td>
<td>49</td>
<td>49</td>
<td>49</td>
<td>49</td>
<td>49</td>
<td>1</td>
<td>49</td>
<td>49</td>
<td>49</td>
<td>49</td>
<td>49</td>
<td>49</td>
<td>49</td>
<td>49</td>
</tr>
<tr>
<td>polish</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>portuguese</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>russian_windows1251</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>1</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>scots</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>tagalog</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>welsh</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>
References


“I hope it is not as bad as I fear”. Procedia - Social and Behavioral Sciences, 26, 55-62.

doi://dx.doi.org/10.1016/j.sbspro.2011.10.562
Curriculum Vitae

Daryl Boykin, Ph.D.

8857 Salvestrin Point Ave
Las Vegas, NV 89148

(702)354-5436
Email: dfboykin@gmail.com

Education

University of Nevada-Las Vegas
September 2014-May 2017
Doctorate of Philosophy
  Hospitality Administration
Dissertation: Price Prediction: Determining changes in stock pricing through sentiment Analysis

University of Nevada-Las Vegas
September 2012-May 2014
Master of Science
  Hotel Administration

University of Nevada-Las Vegas
September 2009–May 2012
Bachelor of Science
  Gaming Management
Cum Laude

Teaching Experience

Casino Math- GAM470 (Individual classes)
Quantitative Analysis HOA430 (Individual Classes)

Teaching Interests

Quantitative Analysis
Casino Analytics
Hospitality Services Management
Human Resource Management
Revenue Management

Research Experience

Assistant Editor: Gaming Research and Review Journal

Research Interests

Gaming Analytics
Bayesian Analytics
Application of Bayesian Inference in Hospitality

Presentations
Poster: Daryl Boykin, Table Games Revenue Management: A Bayesian Approach, 15th International Conference on Gambling and Risk Taking, Las Vegas, NV, May 2013


Professional Experience

Risk Analyst III
Credit One Bank
September 2014-Present
Las Vegas, NV
- Oversee 3 analyst team for credit portfolio of over 7.5 million customers nationwide
- Work with Marketing department to develop, test, and measure initiatives to increase active accounts and account balances.
- Develop strategic plans with consideration for operations needs to meet company goals.

Poker Room Manager
The Suncoast Hotel Casino
June 2010-September 2014
Las Vegas, NV
- Manage 10 table card room with 30 person staff, including salaried supervisors
- Insure all gaming and internal controls are appropriately followed
- Analyze promotions, financials, and staff to determine maximum effectiveness

Poker Shift Manager
The Orleans Hotel Casino
August 2005-June 2010
Las Vegas, NV
- Manage staff 10-15 dealers and oversee operation of Poker Room shift
Bar Manager  
*October 1998-May 2005*

**SEGA Gameworks, LLC**  
*Las Vegas, NV*

- Increased sales working with vendors for special prices and marketing materials
- Increased Secret Shopper Scores through focus on service standards and quality
- Decreased bar costs by providing pout training and testing to all bar staff
- Manage staff of 40000 sq ft all age venue as Manager on duty

**Skills**

- Highly Skilled with Microsoft Office and all products within, including Excel and Powerpoint
- Skilled with multiple statistical analytic techniques, including MLR, Time Series, Principle Components, and Panel Regression.
- Familiar with SPSS and MiniTab
- Expert understanding of SAS and R
- Years of experience in Hospitality Industry, and familiar with all aspects of Casino Industry
- Skilled with cluster analysis, and applications to databases
- Deep understanding of Bayesian analytic techniques
- Refined critical thinking abilities
- Strong presentation skills, both verbal and non-verbal
- Experience working with very large datasets to parse needed information for research, or for managerial needs