VALIDITY AND RELIABILITY OF WEB SEARCH BASED PREDICTIONS FOR CAR SALES.

Date: April 22, 2015
Study: Master Business Administration
Track: Innovation & Entrepreneurship

Student: M.C. (Mischa) Voortman
Student no.: s1020374
E-mail: mischa.voortman@gmail.com
Supervisors: Dr. A.B.J.M. (Fons) Wijnhoven
Dr. M.L. (Michel) Ehrenhard

TABLE OF CONTENTS

1 INTRODUCTION

Social media can provide opportunities or create risks for firms (Oehri & Teufel, 2012). With the social media, firms have the power to "influence consumers behaviour in the information search phase of their decision making process" (Agrawal & Yadav, 2012). As firms have the urge to control the voice and word-of-mouth about their brands and products, the need to analyze the word-of-mouth on social media developed. Moreover, marketing strategies via social media became a key element in the business models of firms. Subsequently, the development of social media mining-, opinion mining-, sentiment mining- or sentiment analysis tools emerged. Nowadays, there are many social media mining tools available on the internet. These tools work with algorithms which can filter social media posts and tweets about a brand or product, and classify posts and tweets as positive or negative. The reliability and validity of these social media mining tools are questionable, because one will encounter several problems associated with computerized sentiment classifiers. Pang and Lee (2008) discussed the problems related to sentiment mining and analysis (e.g. sentiment polarity, subjectivity, change of vocabulary, topic-sentiment interaction, order dependence), and Kim and Hovy (2004) also discussed the problems of sentiment word- and sentence classifications. Sentiment polarity refers to the classification of positive or negative sentiments to a sentence. However, a sentence can be recognized as positive, when in fact it is not intended as positive. For example, the following review of a perfume, "If you reading this because it is your darling fragrance, please wear it at home exclusively, and tape the windows shut." (Pang & Lee, 2008). No negative words occur, but the review of this perfume is not positive at all. Order dependence is related to the sentiment attached to a sentence is related to which order the words in a sentence have. For example, "A is better than B" is the exact opposite from "B is better than A", however, the same words are used (Pang & Lee, 2008). Change of vocabulary is related to the topic that the vocabulary of a population could change over time, and sentiment classifiers will be outdated eventually.

Although social media mining tools raise a number of questions and problems, the predictive power of social media is not left unrecognized. Moreover, the predictive power of social media and web search tools is widely discussed in literature. Since social media developed further, several research efforts have explored the potential of the predictive power of these media (Kalampokis, Tambouris, & Tarabanis, 2013). In their review paper, Kalampokis et al. (2013) state: "The majority of the empirical studies support SM (Social Media) predictive power, however more than one-third of these studies infer predictive power without employing predictive analytics. ... In addition, the use of sentiment-related variables

resulted often in controversial outcomes proving that SM data call for sophisticated sentiment analysis approaches." This highlights the fact that some researchers apparently have recognized the predictive power of social media and call for sentiment analysis approaches. In contrast, Couper (2013) refers to social media data as raw and unstructured "organic data", which is not ready for processing. Sentiment analysis is applicable to tweets and posts, because it comprehends sentences that can be analyzed. Kalampokis et al. (2013) show that predictions based on web searches (i.e. Google Trends, Yahoo Search Query Logs) can also be accurate. However, sentiment analysis for web search activities of individuals is likely to be impossible, since no statements are expressed in an individual's web search. Therefore, the question remains if it is possible that an individual's web search activity represents an intention to buy, since sales are being predicted on the basis of web searches.

In the next chapter the term prediction will be discussed, and four perspectives on prediction will be highlighted. Subsequently, in the third chapter previous studies and literature on predictions with use of social media-, social networking- and web search tools will be discussed. In the fourth chapter the research design for this thesis will be presented, with the associated hypotheses. The following section covers the data collection and methods of analysis. In chapter five, the analysis and results of the collected data will be elaborated and determine whether the hypotheses are supported or not. The last chapter describes the key findings and contains a brief discussion about previous studies. Subsequently, limitations are and recommendations for future research and predictions models are addressed.

2 PREDICTIONS

2.1 The concept of prediction and its variants

The term prediction is used often and sometimes inadequately. This immediately leads to the first question: what is a prediction? Shmueli (2010) distinguishes the definitions of explanations and predictions. Shmueli (2010) elaborates on the differences made by different authors, between explanations and predictions. She answers the question: "Why should there be a difference between explaining and predicting?". According to Shmueli (2010), the answer to this question is that measurable data are not entirely accurate reflections of the underlying constructs. "The operationalization of theories and constructs into statistical models and measurable data creates a disparity between the ability to explain phenomena at the conceptual level and the ability to generate predictions at the measureable level" (Shmueli, 2010). This means that explanations are merely very abstract based on evidence. Contrary, predictions operate on a measurable level. These predictions are conclusive when the extent to which the evidence supports a prediction is better than its alternatives. She defines explaining as causal explanation and explanatory modelling as the use of statistical models for testing causal explanations. Moreover, she defines predictive modelling as applying a statistical model or data mining algorithm to data for the purpose of predicting new or future observations. Slightly different than Gregor (2006), Shmueli (2010) considers predictive accuracy and explanatory power as two axes on a two-dimensional plot. Researchers should consider and report both the explanatory and predictive qualities of their models. She also states: "Explanatory power and predictive accuracy are different qualities; a model will possess some level of each." This sentence indicates that both explanation and prediction have to be taken into account. This aligns with the EP-theory of Gregor (2006), where both some explanation and prediction is reported. However, (Gregor, 2006) is in favour of inferring causality underlying a certain prediction model.

Gregor (2006) examined the structural nature of theory in Information System. She addressed issues as causality, explanation and predictions. Five interrelated types of theory are distinguished by Gregor (2006): (1) theory of analyzing, (2) theory for explaining, (3) theory for predicting, (4) theory for explaining and predicting, and (5) theory for design and action. Gregor (2006) distinguishes a prediction from an explanation. An explanation has an underlying mode of reasoning about causality. However, "it is possible to achieve precise predictions without necessarily having understanding of the reasons why outcomes occur" (Gregor, 2006). This means that predictions are not necessarily originated by a causal relationship between variables. The difference between explanation and prediction has to be

clear. "An explanation theory provides an explanation of how, why and when things happened, relying on varying views of causality and methods for argumentation. A prediction theory, states what will happen in the future if certain preconditions hold. The degree of certainty in the prediction is expected to be only approximate or probabilistic in IS... Prediction goes hand in hand with testing (Gregor, 2006)". However, it is possible to provide predictions and have both testable propositions and causal explanations (explanation and prediction; EP-theory). Summarizing, there are three types of theories that can be distinguished: explanation theory, prediction theory, and EP-theory (see table 2.1).

Explanation	Says what is, how, why, when and where. The theory provides explanations but does not aim to predict with any precision. There are no testable propositions.
Prediction	Says what is and what will be. The theory provides predictions and has testable propositions but does not have well-developed justificatory causal explanations.
Explanation and prediction (EP)	Says what is, how, why, when, where, and what will be. Provides predictions and has both testable propositions and causal explanations.

Table 2.1: Explanation- and prediction theory by Gregor (2006).

2.2 Four types of prediction

There are four types of prediction that can be distinguished: (1) Pascalian, (2) Baconian, (3) Action Logic, and (4) the Self-Fulfilling Prophecy. Each type of prediction is outlined below. An example is used to illustrate each prediction type.

Firstly, the Pascalian prediction is based on laws of probability calculation and makes statements about relationships between variables in certain populations (Cohen, 1979). For example, assume that the population in the Netherlands is 60% women and 40% men. Throughout history it became clear that about 50% of the population in the Netherlands would get the flu during a flu epidemic. Now predict the number of men and women that will get the flu during an epidemic in the Netherlands, assuming that the population exists of 16 million people. This is a classical, simplistic example of a Pascalian prediction, where the laws of probability are used to make predictions about a certain population. This is a critical and rationalistic approach to predictions. The Pascalian prediction is enumerative, because the number of events that confirm a certain prediction increase the support for a hypothesis (Weinstock, Goodenough, & Klein, 2013). Does this type of prediction explain the phenomena why only 50% of the population will get the influenza disease or why the population is 60% men and 40% women? No, but accurate predictions are possible.

Secondly, the Baconian prediction comprehends causality and draws conclusions inductively on a number of observations. These observations are followed by a certain event. Baconian prediction assigns a particular cause to these events. Eventually, it will create a logical chain of cause and effect. This means when several cases and observations are collected, subsequently a regularity is derived from these cases. When these regularities are identified, one can predict an event based on these regularities. For example there is an increasing number of sales of pregnancy tests in a certain period. Nine to ten months later, there is an increased sale of diapers. When this happens more often and a correlation can be found between the increased sale of pregnancy tests in one month and an increased sale of diapers nine to 10 months later, this could provide enough evidence for a prediction. The next month that there is an increased sale of pregnancy tests, one could predict the number of sales of diapers for over ten months. On the other hand, women who purchase pregnancy tests are not necessarily pregnant. Therefore, an increase in pregnancy test sales could be the wrong representation of women who are actually pregnant. The latter implies that the explanation (causal relationship) is weak, however, the predictions are very accurate. Cohen (1979) article regarding the psychology of predictions elaborates on the Baconian perspective on four key ideas: "(1) The traditionally distinct methods of agreement and difference are generalised into a single 'method of relevant variables' for grading the inductive reliability of generalisations about natural phenomena in any domain that is assumed to obey causal laws. (2) The (Baconian) probability of an A's being a B is identified with the inductive reliability of the generalisation that all A's are B's. (3) Judgements of Baconian probability are seen to constrain one another in accordance with principles that are derivable within a certain modal-logical axiom-system but not within the classical calculus of chance. (4) Baconian probability functions are seen to deserve a place alongside Pascalian ones in any comprehensive theory of non-demonstrative inference, since Pascalian functions grade probabilification on *the assumption that* all relevant facts are specified in the evidence, while Baconian ones grade it *by the extent to which* all relevant facts are specified in the evidence." Moreover, when a Baconian prediction is favourable, it increases with the weight of evidence. The more samples that confirm a certain relationship, the greater the evidence. Cohen (1979) emphasizes that in modern science the use of the Baconian structure has become standard, regarding predictions. However, he also emphasizes that in some way the Baconian and Pascalian probabilities complement each other and some biases in causality are acknowledged. This aligns with the statement of Couper (2013), that in modern science (e.g. prediction based on social media data) the data is biased. On the other hand, predictions that are based on Baconian structure,

avoid the paradox of the lottery. For example, when there is a lottery and one ticket of thousand tickets is the winning ticket, there is a very little chance of winning. Therefore, it is rational to believe that the first ticket will not win, and the second ticket neither, and so on. However, it is one hundred percent certain that at least one ticket should win. This uncertainty within the Pascalian structure does not exist in the Baconian structure. Cohen (1979) concludes with: "Above all 'the normative theory of prediction' must be taken to include Baconian as well as Pascalian modes of reasoning... It is undeniably reasonable to use the degree of likeness of the cause as one kind of criterion for the probability of the effect". The Baconian probability prediction is eliminative, when there are a variety of alternative explanations for a certain event, then excluding or eliminating the alternatives will increase the support of the hypothesis (Weinstock et al., 2013).

Thirdly, the theory of planned behaviour (Ajzen, 1991), which is an extension of the theory of reasoned action (Ajzen & Fishbein, 1980). The theory of planned behaviour tackles the limitations of the original model in dealing with behaviours over which people have incomplete volitional control (Ajzen, 1991). For accurate predictions, there are three considerations: (1) the measures of intention and of perceived behavioural control must correspond to or be compatible with the behaviour that is to predicted, (2) intentions and perceived behavioural control must remain stable in the interval between their assessment and observation of the behaviour, (3) the predictive behaviour from perceived behavioural control should improve to the extent that perceptions of behavioural control realistically reflect actual control. This theory has some similarities towards the non-deterministic view of probabilities. The non-deterministic view accounts for the "free will" of individuals. This means that an individual or event could trigger a certain reaction. Although, the reaction of an individual is not fixed per se. A reaction can be triggered, but this reaction depends on the free will of the other. This paradox is referred to as follows by Lyon (2011): "an event is determined to occur, but some probability is assigned to it not occurring." Lyon (2011) refers to this as the paradox of deterministic probabilities, which makes this type of prediction less reliable. For example, when playing chess. "When I send my horse to B2, my opponent will probably send his tower to D5". This is a classic example of a deterministic probability. It is clear the player is relying on the free will of the opponent and action-reaction responses. Another example is when during the current conflict in the Middle East, if US President Obama decides to send in ground forces to Syria and Iraq. This could lead to a reaction from the President of Russia, Vladimir Putin, to also send in ground forces to the Middle East.

The fourth type of prediction is aimed on self-fulfilling prophecy theory by Merton (1948). For example, when one can predict a certain event such as "lower mobile phone sales in the next quartile", and research has shown that an increase in "number of blog mentions" is positively correlated with sales. Then a firm can influence the prediction by increasing the blog mentions on the web. Eventually, this could lead to an increase in mobile phone sales in the next quartile. The self-fulfilling prophecy is defined by Merton (1948) as follows and outlined by an example: "The parable tells us that public definitions of a situation (prophecies or predictions) become an integral part of the situation and thus affect subsequent developments. This is peculiar to human affairs. It is not found in the world of nature. Predictions of the return of Halley's comet do not influence its orbit. But the rumoured insolvency of Millingville's bank did affect the actual outcome. The prophecy of collapse led to its own fulfilment. ... Consider the case of the examination neurosis. Convinced that he is destined to fail, the anxious student devotes more time to worry than to study and then turns in a poor examination. The initial fallacious anxiety is transformed into an entirely justified fear." Another example is that a self-fulfilling prophecy could emerge by users of a social media channel. An individual with many followers and who normally gets many retweets (generally known as an "influencer") on Twitter, can create a chain reaction only by being persuasive or dissuasive about a certain product. People retweet this and the tweet will get attention and trigger new behaviours. This eventually could lead to people following the influencer's opinion. For example, "the new iPhone 6 is very bad, and its battery life is not even half a day, no one should buy it". People notice it, retweet it, and eventually decide to react and perhaps make the decision not to purchase an iPhone 6. This could lead to a possible decrease in iPhone 6 sales, and the statement "no one should buy it" becomes a self-fulfilling prophecy. Merton (1948) defined the self-fulfilling prophecy as "a false definition of the situation evoking a new behaviour which makes the originally false conception come true" (Biggs, 2009).

The theories described above are distinguished by the three different types of theory provided by Gregor (2006). A matrix is conducted with the three types of theory listed in the rows and the methods of reasoning in the columns. Some well-know theories are in the matrix in addition to the previous described prediction theories, for example the evolution theory (Darwin, 1859). This matrix describes the differences between different kinds of theories. However, the term 'prediction' is not totally clear in academic research. Moreover, there is not a specific 'theory of prediction'.

	DEDUCTION	**INDUCTION**	**ACTION LOGIC**
EXPLANATION (Gregor, 2006; Shmueli, 2010)		*Darwinian evolution theory*; which cannot be tested for predictive accuracy, but gives an explanation of the evolution (Shmueli, 2010). *String theory*; currently producing untestable predictions (Shmueli, 2010).	
PREDICTION (Gregor, 2006; Shmueli, 2010)	*Pascalian probability* (Cohen, 1979); does not necessarily explain why phenomena occur, but deductively reasons from several statements or general rules to reach a logical conclusion.		*Self-fulfilling Prophecy* (Merton, 1948); a false definition of the situation evoking a new behaviour which makes the originally false conception come true.
EP-THEORY (Gregor, 2006)		*Baconian prediction* (Cohen, 1979); comprehends causality and draws conclusions inductively on a number of observations that confirm a phenomenon.	*The theory of planned behaviour* (Ajzen, 1991); model that predicts consumer's behavioural intentions and allows for explanation of these intentions. *Non-deterministic prediction*: an event is determined to occur, but some probability is assigned to it not occurring (Lyon, 2011).

Table 2.2: Matrix of different theories for predicting, explaining or EP.

2.3 Predictions in social media research

The prediction model in this thesis consists of social media events or web searches followed by another event, for example an increase or decrease in sales. Considering this construct of the predictions being carried out, it is primarily based on the Baconian prediction model. Which means it will be statistically tested, and the initiated relationship has an underlying causality, based on a theoretical construct. The main problems of Baconian predictions are the reliability and validity issues. Moreover, Baconian predictions bring a lot of biases with it (Couper, 2013). This leads to four major challenges for analyzing social media predictions. The first challenge is the variables to be concluded, and the validity of its measurements. Tweets, Facebook posts or web searches, what variables are used for social

media predictions? Moreover, what is the validity of its measurements? Secondly, when no time lag is included we would be "predicting the present". Therefore, time lag is an essential component for making predictions. The third challenge applies to the data, what data can be collected, and is suitable for predictions, and where to retrieve this data? The fourth issue concerns the reliability of the collected data, and if the data reflect the possible causal relationship. Basic methodology literature in addition to the analysis of previous studies on social media or web search based predictions and the methodology of those predictions should be sufficient to elaborate on these challenges. In the next chapter previous studies will be addressed and several subjects will be brought to the attention. To examine a proper prediction model for this thesis it is key to explore the variables and the validity, the applied time lags, and the reliability of the data. These will be elaborated in the next chapter. The main questions are: What is the best method to develop a prediction model, based on the Baconian prediction perspective? And how have other researchers conducted such a prediction model, taken into account the validity and reliability problems discussed earlier?

3 LITERATURE REVIEW

3.1 Literature review strategy

Several research efforts have explored the predictive power of social media(Kalampokis et al., 2013). An overview of the literature is provided by Kalampokis et al. (2013). Other literature was found by searching several journals (MISQuarterly, Information System Research and the Journal of MIS). Google Scholar and Web of Science were consulted for more literature. The primary search keywords were "social media mining", "sentiment mining" and "social media predictions". After some articles were selected and analyzed, keywords like "Web Search/Google Trends predictions" and "Twitter predictions" were selected, because these social media platforms were used very often for predicting certain outcomes. When articles were found, the reference lists of the articles were checked to find more contributing articles. This resulted in a selection of 42 articles (see appendix A). The four challenges, mentioned in the previous chapter are the guideline for the next paragraphs. Which variables are used and how is the validity of the variables justified? What are the time lags between the dependent and independent variables? How and where is the data collected? What is the reliability of the data? Does the data justify an underlying causal relationship?

3.2 Variables and the validity of its measurements

3.2.1 Variables

Asur and Huberman (2010), constructed a linear regression model for predicting box-office revenues for movies in advance of their release, by analyzing tweets about movies. The results were stunning; they outperformed in accuracy those of the Hollywood Stock Exchange. They have shown that there is a strong correlation (R=.90, Adjusted R-square=.80) between the amount of attention (in this case average tweet rate per hour) a given movie gets and its ranking in the future. They emphasize the application of this method to a large panoply of topics(Asur & Huberman, 2010). They have focussed on movies, because of two reasons. Firstly, the topic 'movies' is of considerable interest among social media users. Secondly, the revenues or as they call it "real-world outcomes" can easily be derived from box-office revenues for movies. Twitter was used as the social media platform for providing the input data. The independent variable was the average tweet-rate (tweets per hour on a specific movie). They also used the time-series of the 7 days prior to a movie's release of the average tweet rate which resulted in a stronger relationship between the variables. The relationship between the box office gross and average tweet rate had a positive correlation of .90, which

means that there is a strong linear relationship between the two variables. For the sentiment analysis they dealt with the classification problem of text being labelled as positive, negative or neutral. They used "thousands of workers from Amazon Mechanical Turk (https://www.mturk.com/) to assign sentiments to a large random sample of tweets". Moreover, they assured that "each tweet was labelled by three different people". All the samples were pre-processed by elimination of stop words, special characters and urls or user-ids. In short, they have shown that a successful predictor for box-office gross can be the average tweet-rate (per hour), 7 days prior to the movie's release. They have also shown that sentiment analysis on a specific movie provides some improvement, but not as much as the average tweet-rate of a movie. They eventually provided a generalized model for predicting the revenue of a product using social media (see table 3.1). The adjusted R-square as a measure of the predictive power of the relationship was .973 for the variables Tweet-rate time series including the theatre count. Asur and Huberman (2010) justify this model by referring to a "collective wisdom" of users of social media, which led to the decision of investigating its power at predicting real-world outcomes. They did not refer to specific theories or models which could explain a causal relationship underlying the prediction model as Gregor (2006) proposed in her EP-theory.

$$y = \beta a * A + \beta p * P + \beta d * D + \varepsilon$$

Parameters:	Example Asur&Huberman2010 (movies)
A = rate of attention seeking	(average tweet rate)
P = polarity of sentiments and reviews	(PNratio = tweets with positive sentiment divided by tweets with negative sentiment)
D = distribution parameter	(number of theaters where movies are released)
y = revenue to be predicted	(box office revenues)
β = regression coefficients	x
ε = error	x

Table 3.1: Prediction model designed by Asur and Huberman (2010).

In the study of Franch (2013), the independent variables were the number of YouTube views, number of mentions on Twitter, and number of mentions on Google Blogs. To extract the Twitter mentions, the Twitter application "Topsy Pro" was used. Additionally, the sentiment rating from Twitter Sentiment[1] (Sentiment 140) was used as an independent variable. The dependent variable was the outcome of the British Election in 2010. The results

[1] http://twittersentiment.appspot.com/ (Sentiment 140)

showed that their model could predict the outcomes with an accuracy of one percent point difference with the real outcomes. They theoretically justify this model by referring to "the wisdom of the crowds" by Surowiecki (2005), "...that illustrates the predicting power of common people when their forecasts of uncertain quantities or future events are aggregated. According to this main idea, 'boundedly rational individual(s)' are capable of making, all together, a near-to-optimal decision, often outperforming every individual's intelligence, meaning that the crowd, taken as an intelligent entity, is smarter than most of any human counterparts taken singularly." They validate their model by measuring "the approval of the future Prime Minister" or "popularity of each candidate" with the before mentioned independent variables.

Goel, Hofman, Lahaie, Pennock, and Watts (2010), used web search query logs of Yahoo! Web Search to predict Box Office Revenues, the rank on a Top 100 list of a song and Video Game Sales for a specific game. In this research the number of web searches was used for predicting different outcomes, namely a rank on a top 100, box office revenue (in $) and video game sales (in units)[2]. Results showed that there was a strong relationship between search-based predictions and real outcomes (movies R=.94, music R=.70, and video games R=.80). They did not use the R-square as an indicator for the predictive power, however, they used the correlation coefficient to indicate a strong relationship between predicted outcomes and real outcomes. They justify the model by referring to earlier work from Asur and Huberman (2010) and Gruhl, Guha, Kumar, Novak, and Tomkins (2005), "As people increasingly turn to the Internet for news, information and research purposes, it is tempting to view online activity at any moment in time as a snapshot of the collective consciousness, reflecting the instantaneous interests, concerns and intentions of the global population." However, the prediction studies they refer to did not comprehend any explanation of a possible causal relationship between the variables.

Lassen, Madsen, and Vatrapu (2014) investigated if they could predict iPhone sales based on the number of tweets with the keyword "iphone". One of their main findings is the strength of Twitter as a social data source for predicting smartphone sales. They calculated a weighted average for the tweets every Quarter from 2010 until 2013. It is stated that the principles for monthly weighting, would follow – more or less – the same principles if monthly sales data is available. The results show that there is a strong predictive power between tweets and iPhone sales with the R-square coefficient of .95 and .96 for multiple

[2] http://www.vgchartz.com/ ; http://www.billboard.com/charts/hot-100 ; http://www.imdb.com/

regression, with sentiment (retrieved from Topsy Pro of the same Quarter) as the second variable. The average error of the prediction model was 5-10% for the iPhone sales. The prediction model of Lassen et al. (2014) is in table 3.2. They have extended on the research of Asur and Huberman (2010), by measuring the relationship between twitter data and quarterly sales of iPhones. Therefore, they "investigate a new domain (smartphone sales), and theoretically grounding their analysis in relevant domain theory". The underlying theory is the AIDA model, which comprehends the stages in a sales process: Awareness/Attention, Interest, Desire, and Action (Li & Leckenby, 2007). Moreover, they state that tweets are treated as a proxy for a user's attention towards the object of analysis (the iPhone). This means that they have tried to explain the relationship between "social media data" and "real-world outcomes".

$y = \beta a * Atw + \beta p * Ptw + \alpha + \varepsilon$	
Parameters:	
Atw = Time lagged and season weighted Twitter data	Slightly different from Asur & Huberman (2010) method.
Ptw = Sentiment of Atw	
α = alpha	
y = iPhone sales in Units	
β = regression coefficients	
ε = error	

Table 3.2: Prediction model of Lassen, Madsen & Vatrapu (2014).

3.2.2 Validity

As stated in the previous chapter, Baconian predictions come with a lot of issues, including validity. Validity refers to the extent to which an empirical measure adequately reflects the real meaning of the concept under consideration (Babbie, 2012). There are four types of validity. Face validity means that the concepts make it seem a reasonable measure for a certain variable (Babbie, 2012). For example, the frequency of attending classes, getting good grades for assignments and asking questions could be a good indicator for the level of study activity. This means it has good face validity. The *face validity* of the variables seems adequate for the study of Lassen et al. (2014), the amount of attention a product comprehends is represented by the amount of chatter on Twitter. Moreover, the eventual amount of purchases to be predicted is represented by factual (quarterly) sales numbers of the same products. Criterion-related validity, sometimes called *predictive validity*, is based on an external criterion. For example, the validity of College Board exams is shown in their ability to predict students' success in college (Babbie, 2012). This type of validity is very strong with some of the prediction studies discussed in the previous paragraph. Researchers show that

their model can predict future outcomes very accurate (Asur & Huberman, 2010; Goel et al., 2010; Lassen et al., 2014). However, not all of these prediction models are explained by a construct of theories, and therefore do not all comprehend causal relationships. This is called the construct validity, which is based on the logical relationships among variables. Moreover, the degree to which a measure relates to other variables as expected within the system of theoretical relationships (Babbie, 2012). "For example, studying the sources and consequences of marital satisfaction. As part of the research a measure for marital satisfaction is developed. To test its validity certain theoretical expectations of the relation of marital satisfaction to other variables will be developed. Than you might reasonably conclude that satisfied husbands and wives will less likely to cheat than dissatisfied ones. This would constitute evidence of the measure's construct validity" (Babbie, 2012). The discrepancy between the theoretical explanations subjacent to a prediction model was elaborated in the second chapter of this thesis. An explanation for a causal relationship between the variables is not elaborated in a large proportion of previous prediction studies. Merely, they are interested in the independent variables which are linked to social media, and their potential predictive power for (future) real-world outcomes. For example, the relationship of an individual's tweet towards an intention to buy, and therefore the eventual purchase is not validated by a theoretical construct in prediction studies. Finally, content validity refers to the degree to which a measure covers the range of meanings included within a concept. For example, when testing the degree that a person has mastered the English language, testing the vocabulary is not enough. It should comprehend all aspects, for example, grammar, spelling, adjectives and prepositions. In the field of web-search predictions, the intention to buy of an individual is merely captured by number of searches on the web. However, the intention to buy of an individual could comprehend more information search resources, such as 'talking to a friend' or 'visiting stores' to learn more about the product (Kotler, 2000).

3.3 Time lag

Time lag reflects the period of time until predictions take effect in reality. For example, social media data can be gathered in one week, where a specific tipping point has been identified. However, the actual tipping point of sales in reality can be a day, week or month later. The time lag could differ per research topic or per product. Lassen et al. (2014) show the strongest relationship between predicted iPhone sales and actual iPhone sales when using a time lag of 20 days. This seems a good estimate to use for products like smartphones. The underlying AIDA model in the study discussed in the previous paragraph shows that the

time lag is the time between (t1) attention (tweet) and (t2) action (purchase). In the study of Asur and Huberman (2010), it is debated that predictions on box office revenues are estimated one week prior to their release. Ghose and Ipeirotis (2011) studied product and sales data based on reviewer characteristics, but add 'reviewer history', 'review readability' and 'review subjectivity' as variables. They analyzed each review independently of the other existing reviews. They conclude that when a review is more subjective about a specific product it also shows an increase in sales for that product. Moreover, a higher 'readability' score is also related with higher sales. In this research they used four different time intervals (1 day, 3 days, 7 days and 14 days). The reason for this 'time lag' is that the researchers wanted to observe how far in time relevant and adequate predictions can be conducted and still get reasonable results. The result was that when the time lag increases, the accuracy also increases slightly. This means that reviews do not have an immediate effect particularly, but are most accurate with a time lag of 14 days. Gruhl et al. (2005) investigated if blog mentions correlated with spikes in book sales ranks on amazon.com. They concluded that when a book is mentioned more than 200 times within a specific period of time, the time lag decreases to 8.2 days. Contrary, a book that was mentioned less than 50 times, perceived a time lag of 17.2 days. However, they used a data set of 50 books and found that the time lag could differ from a couple of days to several weeks. Furthermore, the sales rank of only 10 of the 50 books were highly correlated with spikes in blog mentions. It is clear that the time lag is different for different (types of) products. Moreover, the time lag can be adjusted during the training set, so the best correlation can be found for number of tweets and number of sales for example. The function cross-correlation can be used to find the best time lag (Gruhl et al., 2005).

Choi and Varian (2012) predicted the 'present' by collecting web-search data for a certain month and predicting the sales for the same month. This cannot be seen as a real prediction, therefore, they suggest that future research should consider a time lag in the model for predicting future outcomes. This implies that time lag is a necessary condition for making actual predictions instead of 'predicting the present'. This means that the initial prediction model between social media and future outcomes, depends on the time lag. Therefore, time lag cannot be seen as a moderating variable, but rather a concept that should be part of the prediction model. The time lag of web-search based predictions could be explained by the theory of the consumer buyer process provided by Kotler (2000). He explains the 5 stages of a consumer buying process: (1) problem recognition, (2) information search, (3) evaluation of alternatives, (4) purchase decision, and (5) post-purchase behaviour. The time lag exists between two events over time: (t1) information search on the web (commercial information

search sources) and (t2) the actual purchase of a product. The length of this time lag is dependent on two factors price and perceived risk. Research has shown that the decision time of the customer increases with the height of the price (Somervuori & Ravaja, 2013). This means the time lag could be different for expensive product types and cheap product types.

3.4 Platforms and data reliability

3.4.1 Web-search engines and microblogs

There are several platforms that can be used for predicting future outcomes. In this section, web-search- (Google and Yahoo!) and microblog (Twitter and Facebook) platforms are evaluated based on several articles. Both of these platforms should be an easy accessible data source, through API or Topsy Pro, or the pre-processed data on Google Trends and Yahoo query logs. In some of the studies an older version of Google Trends "Google Insights for Search" (GIS) was used. GIS has been shut down since 27th of September 2012, and was merged to Google Trends. Before the shutdown of GIS, more in-depth information was publicly available[3]. The article of Lui, Metaxas, and Mustafaraj (2011), showed that GIS was not the best predictor for elections. There was almost no correlation between the GIS data and the actual election polls (r=.02). Contrary, the article of Vosen and Schmidt (2011) shows a different result. They investigated GIS as a predictor for private consumption. They concluded that GIS as a predictor provides better results in comparison to the generally used survey-based predictions. However, this article also dates from 2011 and used data that was available with GIS, which is not accessible anymore. Nowadays, Google Trends can provide us with a pre-processed data set of relative search volumes for a particular subject. How much of the Google Trends data differs from the GIS system is not clear. Choi and Varian (2012) studied the predictive power of Google Trends for several categories, whereas the category "Motor Vehicles and Parts" is one of the topics. Choi and Varian (2012) developed a regression model and could predict 80.8% (Adjusted R-Square: 0.808) of the variance in the dependent variable (motor vehicles and parts sales), using the independent variable (Google Trends categories of Motor Vehicles and Parts) as the predictor.

Twitter is often used as a platform to predict certain outcomes. It is used to predict election outcomes (Franch, 2013; Lui et al., 2011; Metaxas, Mustafaraj, & Gayo-Avello, 2011; Sang & Bos, 2012; Tumasjan, Sprenger, Sandner, & Welpe, 2010), disease outbreaks or influenza (Achrekar, Gandhe, Lazarus, Yu, & Liu, 2011; Culotta, 2010; Ritterman, Osborne,

[3] http://www.conductor.com/blog/2013/01/what-was-google-insights-for-search/

& Klein, 2009; Signorini, Segre, & Polgreen, 2011), but also sales or revenues (Asur & Huberman, 2010; Lassen et al., 2014). Web search and Google Trends are also often used in prediction models (Bordino et al., 2012; Choi & Varian, 2012; Ettredge, Gerdes, & Karuga, 2005; Ginsberg et al., 2008; Goel et al., 2010; Guzman, 2011; Lui et al., 2011; Polgreen, Chen, Pennock, Nelson, & Weinstein, 2008; Vosen & Schmidt, 2011; Wu & Brynjolfsson, 2013). Twitter is used often, because it is an open source platform. If someone has enough knowledge of API, tweets should be easy to extract. Moreover, Twitter has proven its predictive power with high R-square values (Asur & Huberman, 2010; Lassen et al., 2014).

A platform which is less often used is Facebook. Facebook is not an open source platform. This means that messages, comments and posts are not publicly available for research (Couper, 2013). Therefore, Facebook cannot provide easy to process and accessible input data. Couper (2013) states that about one-third of the Facebook community has no demographic data available and not all users are active users. For example, it was estimated that almost 9% of the Facebook accounts is fake, duplicate, undesirable or misclassified (Couper, 2013). However, some social media monitoring sites (i.e. Social Mention, How Sociable and Brandwatch) claim they can obtain and filter messages from Facebook, for brand monitoring purposes. However, the algorithms of these monitoring tools are a black-box and we have no insight in how the messages are extracted or if it is consistent. This also acknowledged by other researchers. Chan, Pitt, and Nel (2014) just assumed that a tool like Social Mention is accurate and reliable for a measure of social media discourse. However, this is a limitation of their study. Moreover, they advocate for an independent confirmation of the trustworthiness and reliability of the data providers such as Social Mention. To gain a better understanding of the methodologies, they proposed to work directly with these services. Botha, Farshid, and Pitt (2011) support his view: "First, it would be wise to find ways of confirming the reliability and validity of data gathered by services as How Sociable. This might be done by consulting and working directly with these service providers in an effort to gain a better understanding of their methodologies and results.

3.4.2 Data reliability

The main problems with Baconian predictions also concern reliability. We need to elaborate on the definition reliability. Babbie (2012) states that reliability is a matter of whether a particular technique, applied repeatedly to the same object, yields the same result each time. For example, you want to know your exact weight. You are going to stand on a scale and it gives a certain weight. Stand on it another time and it gives the same weight.

Concluding the measurement is reliable, note that when weighing more times increases the reliability. Reliability decreases if there is only one observer, because he is probably subjective. That means that there has to be more than a single source of data to increase the reliability, and to apply the test-retest method (Babbie, 2012). However, many studies use only one source of data to base their predictions on. For example, Twitter (Asur & Huberman, 2010; Lassen et al., 2014), Google Trends (Choi & Varian, 2012), and Yahoo! Query logs (Goel et al., 2010).

Couper (2013) discussed several reliability issues of social media data. Table 3.3 shows some of the differences and similarities between social media data and survey data discussed by Couper (2013).

SOCIAL MEDIA	**SURVEYS**
No demographic data available	Demographic questions in survey
Limited type of data	Type of data is only limited to type of questions
Biased (selection bias & measurement bias)	Selection biases may be negligible. Measurement biases are less.
Short-term trends	Long-term trends
Privacy issues	Confidentially and anonymous response possible
Not all social media is easy accessible for research purposes	Public access to the data – conditional on confidentiality restrictions and disclosure limitations.
Possibility of data manipulation	Manipulation is almost impossible
Large sample sizes, but not always accurate	Smaller sample sizes, but more accurate
File drawer effect	File drawer effect

Table 3.3: Comparison of Social Media data and Survey based data.

It shows that social media data has a number of issues that cannot be resolved, whereas surveys can elude these issues. Selection and measurement biases are two of these issues. Selection biases refer to the fact that only a small sample of the entire population uses certain social media, often the elite (e.g. 13% of the US population has a Twitter account). It is not taken into account how many users are active. Secondly, the measurement bias focuses on the question "To what extent do people's posts represent their 'true' values, beliefs, behaviours etc.?"(Couper, 2013). This question has not been answered so far. Do people behave differently behind a personal computer, as in real life? Manipulation is also one of the main issues concerning social media data. For example, companies might create specific interest in a certain topic by using a certain code to generate content automatically (Couper, 2013) and thus create a self fulfilling prophecy towards higher sales. This influences the entire market and the results of social media (monitoring) websites. The problem of the file drawer effect is

an issue that concerns both methods, social media and surveys. “The file drawer effect refers to the problem that journals are filled with 5% of the studies that show Type I errors, while the file drawers are filled with the 95% of the studies that show nonsignificant results” (Rosenthal, 1979). This means that only studies that support hypotheses that are in favour of big data are reported in journals. The Type I error refers to studies that reject the null-hypothesis, because they get significant results that the null hypothesis (e.g. against the predictive power of social media) is false. However, the null-hypothesis could actually be true. “There are many other published papers using internet searches or Twitter analyses to “predict” a variety of things... While these papers trumpet the success of the method (by showing high correlations between the organic data and benchmark measures), we do not know how many efforts to find such relationships have failed (Couper, 2013).”

3.5 Implications for this study

There are some research gaps in the field of prediction studies. Choi and Varian (2012) predicted sales of “motor vehicles and parts”. These sales numbers were based on surveys disseminated to car dealers about current sales numbers. The Google Trends’ category “motor vehicles and parts” was the independent variable. They did not investigate predicting factual sales (i.e. weekly or monthly) of specific car models with Google Trends. Moreover, they predicted the ‘present’ and encouraged that Google Trends data might predict the future, which implies that the introduction of a time lag in the model is necessary. Recent research efforts have focussed on movies (Asur & Huberman, 2010; Goel et al., 2010), video games (Goel et al., 2010), books (Gruhl et al., 2005), music (Goel et al., 2010), elections (Franch, 2013), and smartphones (Lassen et al., 2014). These studies focussed on subjects that had low risk and relatively low prices, and therefore relatively short time lags (Kotler, 2000; Somervuori & Ravaja, 2013). The question remains whether there is also great predictive power when predicting sales of products with higher risk and higher prices and thus longer time lags.

To develop a prediction model, there has to be a significant relationship between the independent and dependent variable. The design of this initial prediction model comprehends a couple of elements. Firstly, we can distinguish the social media platforms, which are for example Facebook, Twitter and YouTube. Secondly, variables can be extracted from these social media platforms (i.e. number of mentions, sentiment rates or number of searches), which are the independent variables. Thirdly, the dependent variables which are predicted (i.e. sales numbers, revenues or election outcomes). Fourthly, the introduction of time lag in the

prediction model could enhance the initial relationship, by finding the best time lag between the number of mentions or searches and the actual purchases of products. Finally, the decision time could differ between expensive products and cheap products. Therefore, price has a positive relationship with time lag; longer time lags are related to higher prices. Based on the results from the literature review, the initial prediction model would look as follows (figure 3.1). It is constructed as a causal model, based on the assumption that a mention on social media or a search activity on Google for a specific product will comprehend a rate of attention or an intention to buy respectively.

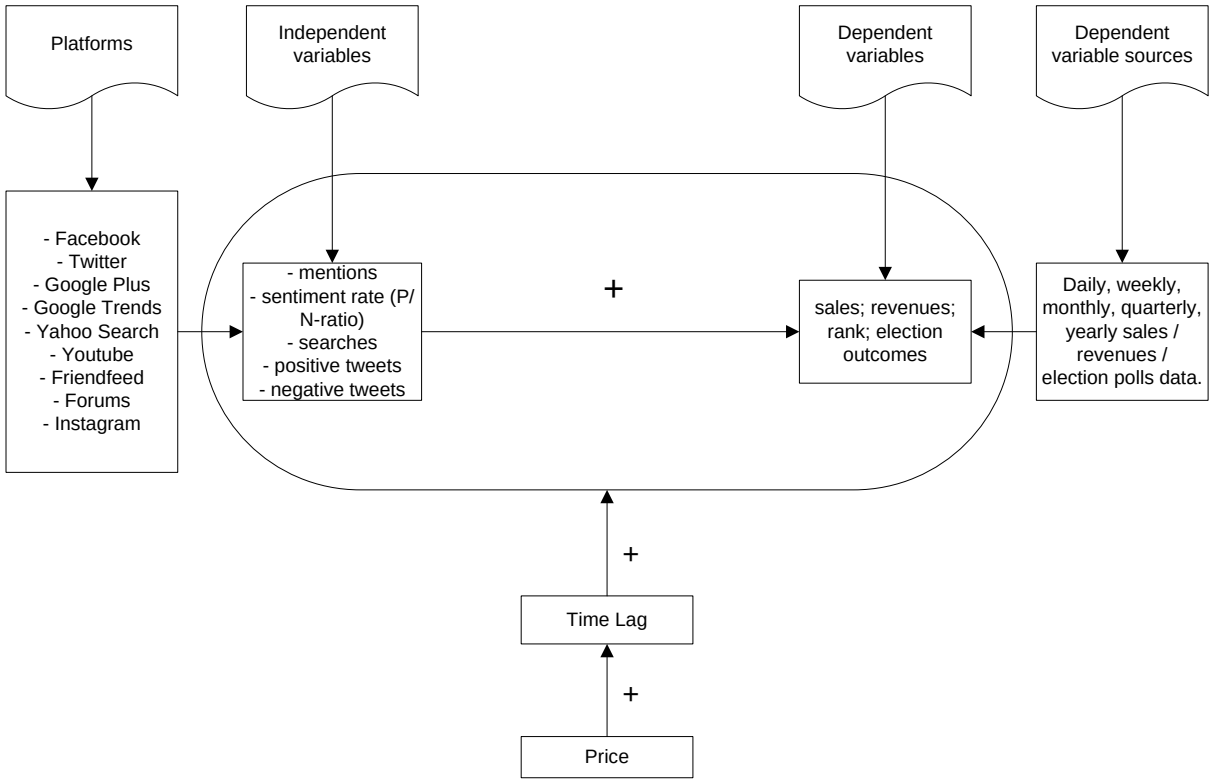

Figure 3.1: Prediction model

4 METHODOLOGY

4.1 Research design and operationalization

This thesis elaborates on the fact that Choi and Varian (2012) did not predict future sales, rather they predicted the present. They emphasized that predicting future sales with Google Trends is a question to be answered in future research. Therefore, it has to be tested if introducing time lag in the prediction model improves the strength of the original relationship between trends and sales. Kotler (2000) stated that within the consumer buying decision process the information search stage consists of using commercial sources (i.e. advertising and websites). In this prediction model the information stage is represented by a consumer's search activity on the web. Choi and Varian (2012) used number of searches for specific categories in Google Trends to predict outcomes of sales (e.g. Motor Vehicles and Parts). Like Choi and Varian (2012), Google Trends is selected for the data input in this research design. Google Trends collects and analyzes the number of searches for a specific term and transforms this into a number between 0 and 100. This number is based on the total number of searches in comparison with the overall score for a specific search term. So instead of using the number of mentions in social media (Asur & Huberman, 2010), the relative search volumes will act as an independent variable (Choi & Varian, 2012). The reason for this is ease of access of the Google Trends data in comparison to the extraction of Twitter data, which requires some API knowledge and skill. Moreover, when following the consumer buying decision process theory, Google Trends data is a better representation of search activity than mentions on Twitter, which represents the rate of attention. The relative numbers provided by Google Trends will act as a direct input for the independent variable (trends). The dependent variable (sales) consists of sales numbers of specific car models in the Netherlands. The car sales data were the only publicly available data that represented factual sales and in a smaller time frame than quarterly, namely monthly. Choosing cars as the dependent variable should also cover the literature gap of products with longer time lags. Moreover, choosing car sales is more specific than the category (motor vehicles and parts) used by Choi and Varian (2012). Growth from Knowledge (GfK) was contacted and requested if it could provide factual sales data of other products (i.e. smartphone sales and videogames sales). GfK has 13.000 market research experts and analyzes market information out of more than 100 countries worldwide. However, the response was that these data are being sold for considerable sums of money, and are not publicly available for research. Research has shown that the decision time of the customer increases with the height of the price (Somervuori & Ravaja, 2013). This means the time lag could be different for highly priced car models and lowly priced car models. Making

this distinction could improve the generalizability of the prediction model for other product types. Time lag is introduced in the prediction model and could strengthen or weaken the relationship between the dependent and independent variable. Time lag is the time between a consumers search on the internet (t1) and the actual purchase of a product (t2).

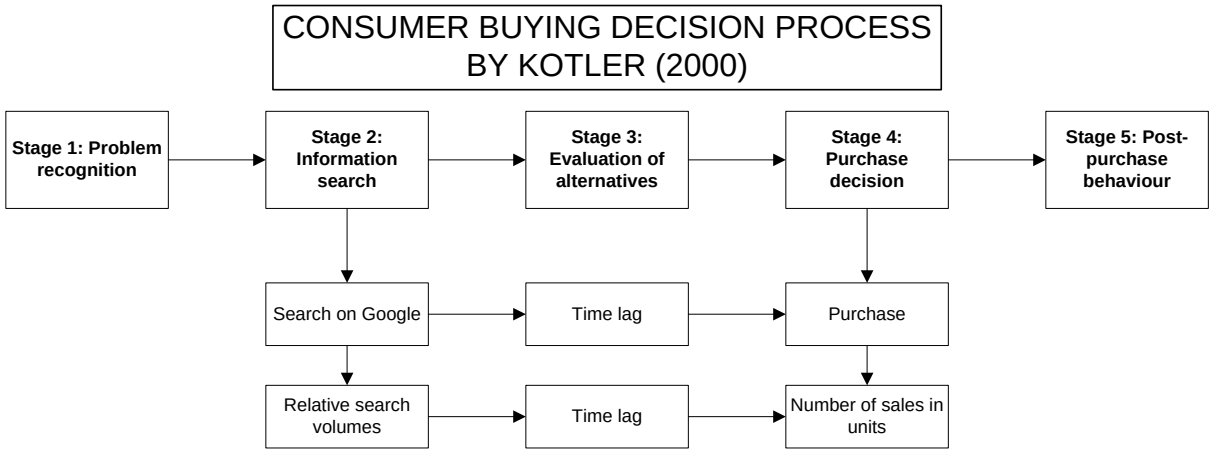

Figure 4.1: Theoretical construct of consumer buying decision process towards measurable variables.

Exogenous factors like new car model introductions or news events could have a negative effect on the relationship between search volumes and car sales. This is based on the assumption that when a new car model is announced, this car model will get a large amount of attention. This spike in search volumes could exists of a large proportion of consumers that have no intention to buy. This results in a discrepancy between relative number of searches and the actual sales, which will occur several months later when the car is actually available for consumers. Simplifying the previous model and introducing the time lag, which could be explained by price according to the theory of decision time of consumers (Somervuori & Ravaja, 2013), and adding the new car model introduction (news) variable gives the following research design (figure 4.2), with the corresponding hypotheses. In the next paragraphs the data collection and methods of analysis for the hypotheses will be elaborated

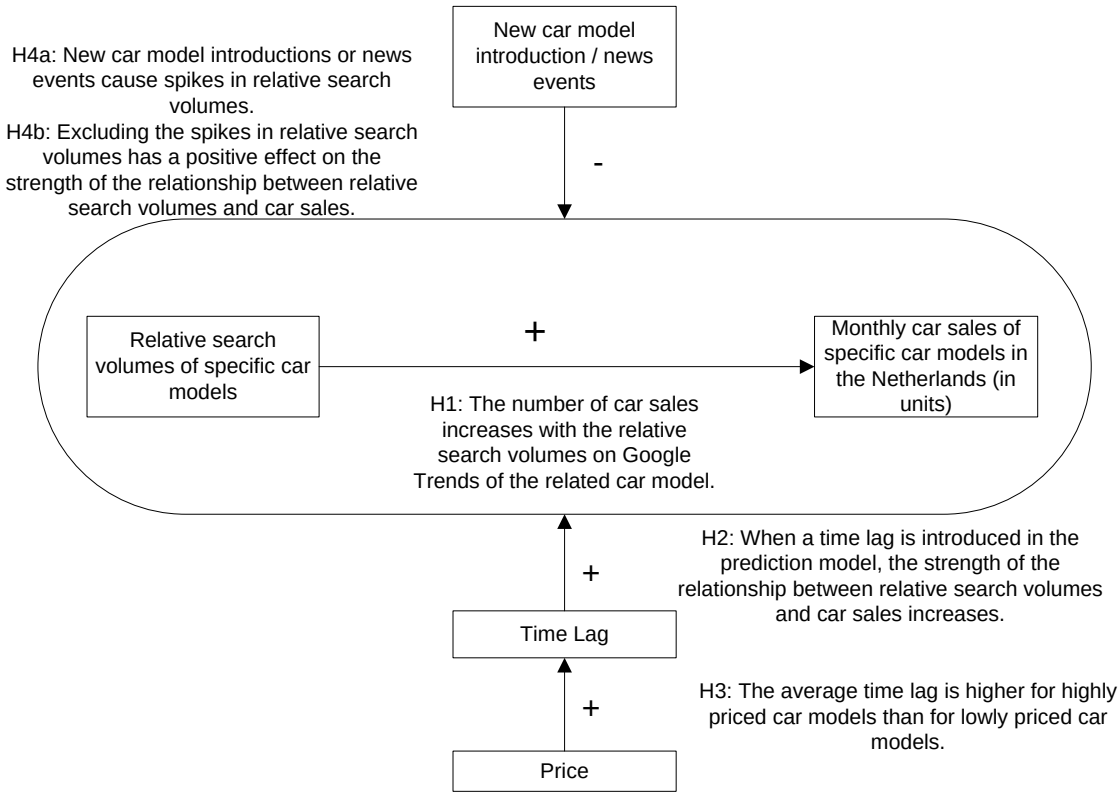

Figure 4.2: Research model: prediction model for car sales and hypotheses.

4.2 Data collection

The sales data of cars are derived from the website of BOVAG[4]. A list of the selected car models is in appendix B. There are 68 different car models (N=68) selected based on the number of sales. Note that not all car models are sold throughout all seven years, some models were launched later and some were removed from the product range earlier. The data from the website is processed in a .PDF file on the website of BOVAG. This file shows the number of sales of specific car models on a monthly basis for the Netherlands. This data is entered in an SPSS database to analyze the data. This has been done manually for some specific months, because the free converting tool for scanned images within a .PDF file limits the converting of .PDF files to one page only. The sales data extracted from BOVAG runs from January 2008 to December 2014 (7 years). The sample size of seven years is based on the assumption that the larger the sample size, the less the variability in the collected data will be when used for forecasting purposes (Hyndman & Kostenko, 2007). Hyndman and Kostenko (2007) stated that "the sample size has to be as large as possible" to exile as much of the variability in the data as possible, for forecasting purposes. Real data often contain a lot of random variation, and sample size requirements increase accordingly (Hyndman &

[4] http://www.bovag.nl/over-bovag/cijfers/verkoopcijfers-auto

Kostenko, 2007). The sales data is for the Netherlands only, which means that the trends data has to be demarcated for Dutch searches only.

The trends data could be collected through the 'export to .CSV' option, which can be opened in MS Excel and should provide the correct data. However, when selecting the specific time periods (i.e. January 2008 to December 2014) the data is still exported per week number instead of months. Consequently, some week numbers would cover two different months. Moreover, the online graph of Google Trends does show the relative search volumes per month. Therefore, all figures were entered in the database manually, by sliding over the graph and entering the right numbers in SPSS. In order to preserve the validity and reliability of the independent variables, the keywords used in Google Trends are the same as the car model names in the BOVAG sales figures (e.g. "Volkswagen Golf", "Peugeot 107"). The quotation marks ensure that the car model names are not taken out of context (i.e. "Golf" could refer to other synonyms like the sport). Moreover, not every individual searching for a Volkswagen starts with the entire word, so acronyms like "VW Golf" are also taken into account. In Google Trends specific periods of time can be selected and therefore makes a good and flexible independent variable. To ensure the content validity of the keywords representing the car models that are selected, the option "Automobiles and Vehicles" is selected. By doing this, Google Trends excluded other contexts of the specific keywords. For example, when selecting "Citroën C4 Picasso" it excludes the interpretations of the painter Picasso in the search results. Another demarcation is location based. Only searches that were located in the Netherlands are demarcated by Google Trends. The average car prices are derived from the Top Gear[5] (Dutch version) website and are in Euro (€). The new model announcement dates and news items are derived from Autoweek.nl, Google News, and Topgear.nl.

[5] http://www.topgear.nl/koopgids/nieuw/

4.3 Methods of analysis

To analyze the first hypothesis, the monthly average for the trends and the sales of each car model is calculated. Subsequently, the relationship between TRENDS_average and SALES_average is tested in SPSS for a correlation (Analyze > Regression > Linear). The significance (p-value) of this relationship is key for the decision if the null hypothesis (no relationship) will be rejected or not. A confidence interval of 95% is applied. If the p-value is smaller than .05, the null hypothesis is rejected, when the p-value is larger than .05, the null hypothesis (no relationship) is accepted. The correlation coefficient 'R' is a measure of showing the direction and strength of the linear relationship between two variables. Moreover, the predictive power is represented by the R-square. This aligns with the methods used by Asur and Huberman (2010) and Choi and Varian (2012), where the R-square reflects the predictive power of the independent variable within the regression model.

For the second hypothesis it is key to find the best time lag. Therefore, the highest correlation coefficient has to be found for a specific time lag. The sales and trends data will be shifted alongside each other to find the right time lag. It is assumed that the time lag will not exceed the 24 month demarcation. Putsis Jr and Srinivasan (1994) studied what time a consumer takes to purchase a car. This is measured from the moment that a consumer is thinking of purchasing a new car until the actual purchase. The results of this study were that 16% of the individuals took less than a month after first thinking about it, 34% took between 1-3 months, 16% took between 3-6 months, 15% took 6 to 12 months, and only 9% took more than 12 months. Note that 91% took less than one year to actually purchase a car. Furthermore, the moment of "thinking about a new car" (recognition of a need/problem) comes prior to the "information search stage" in the consumer buying process mentioned by Kotler (2000). It is acknowledged that consumer's purchasing behaviour could have changed over time, but it is not clear how they affect the relationship. Seventeen cars are randomly selected in SPSS (Data > Select cases > Random sample of cases > Sample... "Approximately 25% of the cases") to test if there are any changes in the predictive power (R-square) of the relationship, when a time lag is introduced. The sample is limited to seventeen cars because of the extensive work for performing the analysis for each car model separately, and by conducting the analysis for 25% of the cars from the original sample, the results should be covered sufficiently. This analysis follows four steps. First, a linear regression analysis is conducted and the results (correlation coefficient, R-square and significance) are listed in a table. Subsequently, the function in SPSS for locating the strongest correlation with the corresponding time lag is "cross-correlation", and if it is significant at a 95% confidence level

(Analyze > Forecast > Cross-Correlation). Thereafter, the time lag is applied, a linear regression analysis is conducted, and the results are also listed in the table. Finally, the changes in the R-square indicate whether the predictive power has changed, and what the best time lag should be. If there are significant improvements in strength and predictive power, the null hypothesis (no improvement) is rejected and the alternative hypothesis (improvement) is accepted. The steps are outlined in table 4.1.

Step 1	Linear regression of original relationship between trends and sales of a specific car. Report results in table.
Step 2	Cross-correlation on the trends and sales data to calculate the best time lag. Apply the time lag on the sales data (i.e. instead of matching the trends data of jan-2005 to the sales data of jan-2005, match jan-2005 to apr-2005 respectively, lag = 3).
Step 3	Linear regression of the relationship between trends and sales of a specific car, with the applied time lag. Report results in table.
Step 4	Calculate improvements in R, R-square and significance. Report results in table.

Table 4.1: Steps for introducing time lag into the prediction model.

For the third hypothesis, the best time lag for each car model has to be calculated in SPSS. This is conducted by using the cross-correlation option (Analyze > Forecast > Cross-Correlation), which is also used in the previous hypothesis. After finding the best time lag for each car model, the new variable 'best_time_lag' is created. The car prices are divided into two categories, lowly priced cars (N=37) and highly priced cars (N=31), with the values of ≤24.999 and 25.000< respectively. Subsequently, an independent t-test is conducted for the mean time lags of both categories. The results of this test indicate if there is a difference between the mean time lags of lowly- and highly priced cars. The same alpha level is applied, if the p-value is below .05 the null hypothesis (no difference in lag) will be rejected, and the alternative hypothesis (difference in lag) will be accepted.

Finally, the relationship could be influenced by the introduction of new car models or other news events, which could cause spikes in the relative search volumes. Therefore, an additional analysis will be conducted for the seventeen car models selected earlier. This analysis consists of six steps. First, a linear regression analysis will be conducted for the seventeen selected car models. The results will be listed in a table (correlation coefficient, R-square and significance level). Secondly, the spike(s) in the trends data will be identified for the car models and the observations during these spikes will be excluded. This is done by plotting a graph of the trends data and locating the moment of the lowest value prior to a spike and associating the spike with the corresponding news or new car model introduction. Subsequently, the cases (months) during the spike are excluded from the data set. Finally, a linear regression analysis is conducted for the car models, where the observations during the

spike(s) are excluded. Additionally, a time lag will be introduced into the new model and the linear regression results will also be listed in the table. The changes in the R-square indicate whether the predictive power has changed. If there are significant improvements in strength and predictive power, the null hypothesis (no improvement) is rejected and the alternative hypothesis (improvement) is accepted. The six-step method is outlined in table 4.2.

Step 1	Linear regression of original relationship between trends and sales of a specific car. Report results in table.
Step 2	Spike(s) in the trends data will be identified for the car model. This is done by plotting a graph of the trends data and locating the moment of the lowest value prior to the spike and associating this spike with the corresponding news or new car model introduction for that date.
Step 3	The cases (months) during the spikes are excluded from the data set.
Step 4	Linear regression of the relationship between trends and sales of a specific car, with spike(s) excluded. Report results in table.
Step 5	Introduce calculated time lag into the prediction model. Linear regression of the relationship between trends and sales. Report results in table.
Step 6	Calculate improvements in R, R-square and significance for exclusion of spikes and introduction of time lag. Report results in table.

Table 4.1: Steps for excluding spikes and introducing time lag into the model.

5 ANALYSIS & RESULTS

For the first hypothesis the average of trends and sales per car model for all years are calculated. A linear regression analysis is conducted for these two variables. Table 5.1 shows the correlation coefficient. In this case we can see R=.345 indicating a positive relationship between trends and sales, significant at $p<0.01$. However, the relationship appears to be weak. Furthermore, how would this relationship fit into a prediction model? Table 5.1 presents that the R-square is.119, which means that approximately 12 percent of the variance in the dependent variable is explained by the independent variable. This is a low R-square coefficient compared to other studies (i.e. Asur and Huberman (2010), Adjusted R-square=.94 and Choi and Varian (2012), Adjusted R-square=.808). However, the adjusted R-square is only used when more (i.e. moderating) variables or predictors are introduced into the regression model, which is not the case in this study.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,345[a]	,119	,106	264,81512

a. Predictors: (Constant), TRENDS_average

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	59,766	105,909		,564	,574
	TRENDS_average	5,766	1,929	,345	2,990	,004

a. Dependent Variable: SALES_average

Table 5.1: Model summary of linear regression with trends as independent variable.

In the second hypothesis the assumption was made that the time lag would be influential for the strength of the relationship between trends and sales. This means when a time lag is taken into account the correlation coefficient and the R-square should increase (predictive power increases). For this analysis seventeen car models are selected from the dataset. This analysis consists of four steps. First, a linear regression analysis is conducted for each car model. Subsequently, the correlation coefficient R and R-square will be listed in a table with the corresponding significance level. Secondly, a cross correlation will be conducted for each car model (Analyze > Forecast > Cross-Correlation). The cross-correlation function shows for multiple time lags an indication of the correlation coefficient, and whether the correlation is significant at an alpha level of 5%. Thirdly, the best time lag will be listed in the table with the corresponding R, R-square and significance level. Finally, the improvements of the R-square are calculated and indicate whether the null hypothesis is rejected. The BMW 5-serie is used as an example to clarify the method. First, the linear

regression analysis results of the BMW 5-serie data without a time lag are in table 5.2. These results are also listed in table 5.4.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,008[a]	,000	-,012	140,4272

a. Predictors: (Constant), BMW5SERIEtrends

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	251,794	55,592		4,529	,000
	BMW5SERIEtrends	,088	1,175	,008	,075	,941

a. Dependent Variable: BMW 5-SERIE

Table 5.2: Model summary of linear regression for the BMW 5-series without time lag.

When these results are noted, the next step is to conduct a cross correlation for the trends and sales data. The results are in histogram figure 5.1. The cross-correlation shows that there could be a stronger relationship at a time lag of 14 months. In figure 5.1 we can see that the correlation coefficient corresponding to the lag number 14 is also significant at a confidence level of 95%.

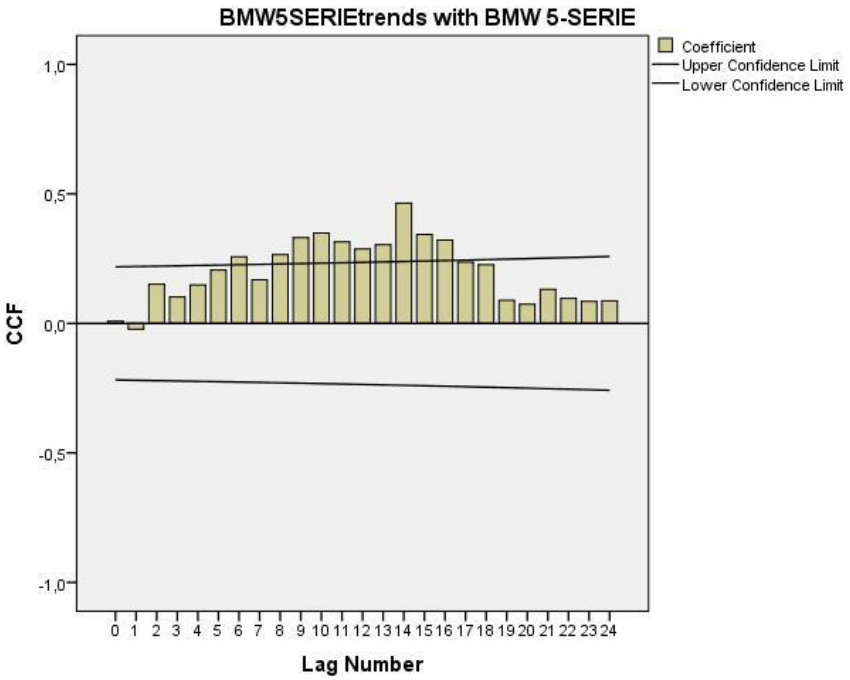

Figure 5.1: Cross-correlation coefficients with 95% confidence level limits.

The next step is to conduct a linear regression analysis with the 14 month time lag applied. For the BMW 5-serie the results are in table 5.3. It shows that the correlation coefficient is now R=.488, the R-square =.238 and significant at p<.001.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,488[a]	,238	,227	129,57969

a. Predictors: (Constant), BMW5SERIEtrends

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	28,202	53,468		,527	,600
	BMW5SERIEtrends	5,084	1,104	,488	4,607	,000

a. Dependent Variable: timelagged14

Table 5.3: Model summary of linear regression for the BMW 5-serie with 14 months time lag.

CAR MODELS	LINEAR REGRESSION RESULTS (no time lag)			LINEAR REGRESSION RESULTS (after identifying the best time lag)			Improvements	
	R	R^2	Sig.	R	R^2	Sig.	Lag	R^2
Audi A6	.049	.002	.657	.237	.056	.048	14	+5.4%
BMW 5-serie	.008	.000	.941	.488	.238	.000	14	+23.0%
Citroen C3	.413	.171	.000	.553	.305	.000	12	+13.4%
Citroen C5	.481	.231	.000	.826	.682	.000	6	+45.1%
Ford Focus	.349	.122	.001	.349	.122	.001	0	0
Hyundai i10	.006	.000	.957	.030	.001	.790	3	+0.1%
Hyundai ix35	.574	.330	.000	.726	.527	.000	3	+19.7%
KIA Sportage	.653	.427	.000	.653	.427	.000	0	0
Mercedes B Class	.193	.037	.078	.501	.251	.000	6	+21.4%
Mercedes E Class	.128	.016	.246	.322	.104	.004	4	+8.8%
Nissan Juke	.365	.133	.008	.718	.515	.000	3	+38.2%
Opel Corsa	.010	.000	.930	.367	.135	.001	10	+13.5%
Peugeot 107	.609	.371	.000	.609	.371	.000	0	0
Seat Leon	.315	.099	.004	.648	.420	.000	11	+32.1%
Toyota Auris	.608	.370	.000	.684	.467	.000	5	+9.7%
Volkswagen Jetta	.153	.024	.163	.659	.435	.000	6	+41.1%
Volkswagen UP!	.023	.001	.885	.604	.365	.000	6	+36.4%

Table 5.4: Differences before and after introducing time lag in the prediction model.

The same steps are repeated for the other sixteen car models and the results are in table 5.4. Subsequently, the improvements in R-square after introducing the time lag are calculated. For each car, the corresponding best time lag is also listed in the table. The results indicate that

fourteen out of the seventeen (82%) car models have a time lag. When this time lag is applied the predictive power increases for these relationships between relative search volumes and car sales. This indicates that the null hypothesis when time lag is introduced in the model, the predictive power of the relationship between relative search volume (trends) and car sales (sales) show no improvements, is rejected. Therefore, the alternative hypothesis of introducing a time lag in the model, the predictive power of the relationship between trends and sales improves, is accepted. The high significance level and level of predictive power of the Hyundai i10 cannot be explained, since the same steps are repeated for each car model.

For the third hypothesis, it is expected that the average time lag is higher for highly priced car models than for lowly priced car models. Therefore, for each car model the best time lag was computed. This is done by using the cross-correlation function in SPSS (Analyze > Forecast > Cross-Correlation). This same function was used in the previous hypothesis, to locate the best time lag. Subsequently, a list of best time lags for each car model was created, and a list with all the car prices corresponding to the same car models. The car models are divided in two categories, lowly priced cars and highly priced cars, with the values ≤24.999 and 25.000< respectively. Finally, an independent t-test for the means of the time lags of both categories is executed. The results of the t-test are in table 5.5.

Group Statistics

	Car price divided into 2 categories 0-24.999 en 25.000<	N	Mean	Std. Deviation	Std. Error Mean
best_time_lag	<=24.999	37	3,7568	3,71488	,61072
	25.000<	31	6,5806	6,51037	1,16930

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
best_time_lag	Equal variances assumed	10,481	,002	-2,241	66	,028	-2,82389	1,26034	-5,34023	-,30755
	Equal variances not assumed			-2,141	45,762	,038	-2,82389	1,31918	-5,47963	-,16814

Table 5.5: Independent t-test for the means of the time lags, sorted by price category.

We can see that there is a significant difference between the means of the two categories. The lowly priced car models have a mean time lag of 3.7568 months and the highly priced car models have a mean time lag of 6.5806 months. When we look at the t-test for equality of means we can see that the p-value=.028, with 66 degrees of freedom and a t-value of -2.241. The p-value<.05, and is therefore significant. Concluding, that there is a significant difference of approximately 2.8 months in average time lag between lowly priced cars and highly priced

cars, where highly priced cars have a larger average time lag. This means that the hypothesis that the time lag is higher for highly priced cars than for lowly priced cars, is supported.

Finally, the relationship could be influenced by the introduction of new car models or other news which causes spikes in search volumes within the time span of seven years the data is collected. Therefore, an additional analysis will be conducted, following six steps. First, a linear regression analysis of the seventeen car models used in the previous analysis will be conducted. Subsequently, the spike(s) in the trends data will be identified for the car models. Thirdly, the observations during these spikes will be excluded. A linear regression analysis will be carried out for the car models, where the observations during the spike(s) are excluded. For example, the Peugeot 107, shows a few spikes in the trends data (see fig. 5.2). The results of the linear regression analysis for trends and sales for the Peugeot 107 are in table 5.6. The R=.609 and R-square =.371, significant at p<.001. This indicates a moderate relationship between trends and sales, and very significant.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,609[a]	,371	,363	512,1752

a. Predictors: (Constant), PEUGEOT107trends

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	48,710	148,635		,328	,744
	PEUGEOT107trends	23,519	3,383	,609	6,953	,000

a. Dependent Variable: PEUGEOT 107

Table 5.6: Linear regression analysis for the Peugeot 107 (Jan.2008 – Dec. 2014).

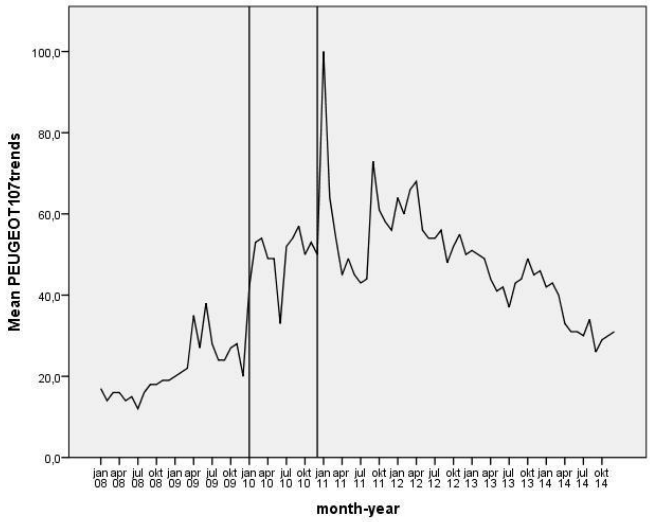

Figure 5.2: Trends data of Peugeot 107 from January 2008-December 2014.

The next step is to identify the spikes in the trends data. There was no introduction of a new model before the spikes occur. However, for the identification of the first spike "Peugeot 107" was entered into Google. Subsequently, the results were filtered on "news" for the "Netherlands" and for the period "October 2009 – April 2010" and "October 2010 – April 2011". In this graph the events before the actual spikes occur are marked with an X-axis reference line:

- Jan 2010: announcement of the recall action for Peugeot 107 and Citroën C1 in the news (Telegraaf).[6]
- Dec 2010: start of the recall action for Peugeot 107 in the news on Autoweek, NRC and NOS.[7]

The cause of the third spike (August 2011) could not be identified using this method. The next step is to exclude the months of the spikes. This can be done in SPSS by selecting individual cases (a case represents one month). Subsequently, a linear regression analysis is conducted for the period January 2012 – December 2014, because the spike effect has worn off approximately around January 2012.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,887[a]	,786	,780	265,0895

a. Predictors: (Constant), PEUGEOT107trends

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-1238,114	187,856		-6,591	,000
	PEUGEOT107trends	45,236	4,047	,887	11,176	,000

a. Dependent Variable: PEUGEOT 107

Table 5.7: Linear regression analysis trends and sales for the Peugeot 107 (January 2012 – December 2014).

The results of the linear regression analysis (table 5.7) show that R=.887 and R-square =.786, significant at p<0.001. There is an improvement of 41.5% in the amount of variance that can be explained in the dependent variable after excluding the spikes in the trends data for the Peugeot 107. A table (5.8) is conducted for the results performing this six-step analysis for the

[6] http://www.telegraaf.nl/autovisie/autovisie_nieuws/20400729/__Ook_terugroepactie_voor_Peugeot_107_en_Citroen_C1__.html

[7] http://nos.nl/artikel/206419-terugroepactie-het-woord-van-de-autobranche.html; http://www.autoweek.nl/nieuws/13544/ook-terugroep-peugeot-107-citroen-c1

other sixteen car models before and after spike exclusion. Subsequently, the time lag is applied, and the linear regression results are also reported in table 5.8.

CAR MODEL (including the event before the spike)	Linear regression results spikes included			Linear regression results spikes excluded			Linear regression results spikes excluded and time lag applied			
	R	R^2	Sig.	R	R^2	Sig.	R	R^2	Sig.	Lag
Audi A6 (Pro Line S model Jun. 2010)	.049	.002	.657	.086	.007	.545	.559	.312	.000	14
BMW 5-series (new model Nov. 2009)	.008	.000	.941	.336	.113	.013	n/a	n/a	n/a	0
Citroen C3 (new model Jun. 2009)	.413	.171	.000	.467	.218	.000	n/a	n/a	n/a	0
Citroen C5 (new model Oct. 2007)	.481	.231	.000	.636	.404	.000	n/a	n/a	n/a	0
Ford Focus (new model Feb. 2011)	.349	.122	.001	.649	.421	.000	n/a	n/a	n/a	0
Hyundai i10 (new model Aug. 2013)	.006	.000	.957	.320	.102	.010	n/a	n/a	n/a	0
Hyundai ix35 (new model Sep. 2009)	.574	.330	.000	.698	.487	.000	n/a	n/a	n/a	0
KIA Sportage (new model Apr. 2010)	.653	.427	.000	.599	.359	.000	n/a	n/a	n/a	0
Mercedes B (new model Aug. 2011)	.193	.037	.078	.266	.071	.085	.657	.432	.000	2
Mercedes E (new models late 2009)	.128	.016	.246	.169	.029	.155	.361	.130	.002	4
Nissan Juke (new model Feb. 2010)	.365	.133	.008	.515	.265	.000	.561	.315	.000	3
Opel Corsa (new model Nov. 2009)	.010	.000	.930	.238	.057	.085	.493	.243	.000	4
Peugeot 107 (recall in Jan./Dec. 2010)	.609	.371	.000	.887	.786	.000	n/a	n/a	n/a	0
Seat Leon (new models May 2009)	.315	.099	.004	.365	.133	.003	.746	.557	.000	11
Toyota Auris (new model Sep. 2009)	.608	.370	.000	.458	.210	.001	.620	.384	.000	6
Volkswagen Jetta (new hybrid Oct. 2012)	.153	.024	.163	.163	.027	.221	.499	.249	.000	6
Volkswagen UP! (new model Jul. 2011)	.023	.001	.885	.428	.183	.037	.536	.288	.008	1

Table 5.8: Differences before and after excluding spikes in the trends data for six car models.

The results show that for fifteen of seventeen cars the R-square increased after exclusion of the spikes. This means that the introduction of new car models or news items can have a negative effect on the predictive power of relative search volumes for car sales. Furthermore, when spikes are excluded the correlation coefficient increases, and the significance improves accordingly. The last steps are calculating and applying the time lag. The result was that nine of the seventeen cars show better results. Moreover, for five cars there was no time lag (time lag = 0), whereas three car models had no time lag upfront (KIA Sportage, Ford Focus, and Peugeot 107). The improvements in percentages of the predictive power are in table 5.9.

CAR MODEL	IMPROVEMENT IN R-SQUARE AFTER EXCLUDING SPIKES	IMPROVEMENT IN R-SQUARE AFTER EXCLUDING SPIKES AND APPLYING TIME LAG
Audi A6	+0.5%	+31.0%
BMW 5-serie	+11.3%	n/a (time lag 0)
Citroen C3	+4.7%	n/a (time lag 0)
Citroen C5	+17.3%	n/a (time lag 0)
Ford Focus	+29.9%	n/a (time lag 0)
Hyundai i10	+10.2%	n/a (time lag 0)
Hyundai ix35	+15.7%	n/a (time lag 0)
KIA Sportage	-6.8%	n/a (time lag 0)
Mercedes B Class	+3.4%	+39.5%
Mercedes E Class	+1.3%	+11.4%
Nissan Juke	+13.2%	+18.2%
Opel Corsa	+5.7%	+24.3%
Peugeot 107	+41.5%	n/a (time lag 0)
Seat Leon	+3.4%	+45.8%
Toyota Auris	-16.0%	+1.4%
Volkswagen Jetta	+0.3%	+22.5%
Volkswagen UP!	+18.2%	+28.7%

Table 5.9: Improvements in R-square of the relationship between trends and sales.

For fifteen of the seventeen (88%) car models the R-square increased. Which means that the strength and predictive power of the relationship between trends and sales has increased, by excluding the spikes in the trends data. Furthermore, for all seventeen car models spikes occurred right after a new model introduction or other news event. The results imply that the null hypothesis (no effect) is rejected. Therefore, the hypotheses, "new car model introductions or events can cause spikes in relative search volumes" and "excluding the spikes in relative search volumes has a positive effect on the strength of the relationship between trends and sales", are supported. It must be noted that not all car models increased in strength after the time lag was introduced. The interpretation of the results and key findings will be elaborated in the next chapter. Subsequently, the differences between previous studies and this study are discussed. Furthermore, recommendations and requirements for future research will be addressed.

6 CONCLUSION & DISCUSSION

6.1 Key findings

The intention of this thesis was to investigate the reliability and validity of web-search based predictions. Five hypotheses were tested for testing the relationship between search volumes and sales. The results of the analysis were as follows. The first hypothesis "the number of car sales increases with the relative search volumes on Google Trends of the related car model" is supported. However, the relationship is weak (R=.345). Approximately 12% of the variance in the car sales could be predicted by the relative search volumes. The second hypothesis "when a time lag is introduced in the prediction model, the strength of the relationship between relative search volumes and car sales increases", is supported. Moreover, for approximately 82% of the analyzed car models the predictive power increased after introducing the time lag in the prediction model. The third hypothesis, "the average time lag is higher for highly priced car models than for lowly priced car models" is supported. Highly priced car models have an average time lag of 6.6 months and lowly priced car models have an average time lag of 3.8 months. The final two hypotheses, "new car model introductions or news events cause spikes in relative search volumes" and "excluding the spikes from relative search volumes has a positive effect on the strength of the relationship between relative search volumes and car sales" are both supported. Moreover, the analysis showed that new car model introduction caused spikes in relative search volumes, and not followed by spikes in sales afterwards. When the spikes were excluded, the strength of the relationship of 88% for the analyzed car models increased. Introducing the time lag after exclusion of spikes, showed that the strength increased further for 53% of the selected car models. For the remaining 47% of the car models, the strength of the relationship did not increase and there was no time lag.

According to previous studies, the numbers of tweets, mentions or relative search volumes are successful predictors for respectively box-office revenues, iphone sales, or private consumption. Therefore, when assessing the predictive validity, the relationship between Google Trends' relative number of searches for a specific car and the actual car sales is plausible. However, it has to be considered that not everyone searching for a car model has the intention to purchase the car. On the other hand, people who are tweeting about a car model do not specifically intent to purchase the car they are tweeting about. This is clearly explained in the article of Couper (2013), where selection- and measurement bias were discussed regarding social media and –networking research. Do tweets/search volumes represent the true intents, values and behaviours of the producers of these data? We do not know whether this is the case. Furthermore, previous studies did not base the relationships

between variables on a theoretical construct. Moreover, the relationship between tweets and sales were based on intentions to buy of individuals tweeting about a product. Lassen et al. (2014) tried to improve the construct validity by using the AIDA model to find an explanation for the relationship between tweets and sales. They have assigned the amount of tweets to the amount of attention of a product, and the number of sales to the action (purchase) of a consumer. In this study, the assumption was that individuals who intent to purchase a car use an online search engine for their information search. This was based on the consumer buyer decision process theory explained by Kotler (2000). This explanation was intended to improve the construct validity for web-search based predictions. Excluding the spikes after a new model introduction partially eliminated the attention that did not comprehend intentions to buy. Therefore, assuming there were only intentions to buy left in the data set, the strength of relationship improved. Summarizing, the results indicate that some considerations have to be accounted for when developing a Baconian prediction model. These considerations will be summarized in the recommendations for future research, after the comparison of previous studies of web search based predictions and this thesis.

6.2 Discussion

	This thesis	**Choi and Varian (2012)**	**Vosen and Schmidt (2011)**
Methods	Linear regression analysis.	Auto Regressive (1) model.	Auto Regressive model.
Platforms	Google Trends	Google Trends	Google Trends
Independent variables	relative number of searches for 68 car models in the Netherlands.	relative number of searches for different categories in Google Trends.	relative number of searches for 56 different categories of private consumption. And survey-based indicators.
Dependent variables	Factual sales in the Netherlands for 68 car models.	Sales in Motor vehicles and Parts based on surveys sent to dealers worldwide.	Personal consumption expenditures (PCE) of 18 categories. Corresponding to the 56 Google Trends categories which are distributed to the 18 categories of PCE.
Timelag included?	Yes, up to 24 months. For each car model the best time lag was computed within the 24 month-frame, and tested if taking time lag into account improved the predictive power.	The 12 month lag is only used for estimating the baseline. It is not tested if time lag has an effect on the predictive power.	Yes, 1 month. Differences in nowcasting (h=0; no time lag) and forecasting (h=1; 1 month time lag) are compared on predictive power.
Predictive power indicator	R-square. (improvements in R-square)	R-square and Mean Absolute Error (MAE). (improvements in MAE)	Incremental Adjusted R-square changes. Root Mean Squared Forecast Errors (RMSFEs).
RESULTS	-Taking time lag into account increases the predictive power - Excluding spikes caused by certain events in search volume data increases the predictive power. - The magnitude of the time lag increases with price. And differs per car model.	-Google trends predicts the present with 80.8% predictive power (R-square =.808).	-Google Trends prediction models, have better results in forecasting consumption than survey-based predictions.
	Assessing the reliability and validity of predictions based on relative search volumes.	Claim they are predicting the present instead of the future.	Claim predicting the future and the present of private consumption.

Table 6.1: Comparison of this study with similar studies (Choi & Varian, 2012; Vosen & Schmidt, 2011).

A comparison between this thesis and the studies of Vosen and Schmidt (2011), and Choi and Varian (2012) is in table 6.1. Choi and Varian (2012) did not claim to predict the future, rather they claim to predict the present (nowcasting) with Google Trends data. Meaning that they used relative search volumes in June to predict the June car sales report which is released later in July. They state that "it may also be true that June queries help to predict July sales, but we leave that question for future research". In this thesis, it is investigated whether it is possible to *actually predict* future car sales, because a time lag is introduced for each car model separately. The difference between the study of Choi and Varian (2012) and this thesis, is that this thesis examine the improvement in the predictive power for exogenous factors

(new car model introductions or news events). Choi and Varian (2012) did not take these factors into account. Moreover, they have not used any factual sales numbers, rather they have used survey-based indications for different motor vehicle and parts dealers of the U.S. Census Bureau. Furthermore, it must be noted that the predictive power of relative search volumes could be much higher for car parts than for vehicles. There could be a difference in the intention to buy of people that are searching for car parts and people that are searching for a certain car model. For example, an individual searching for an exhaust, there is a high probability that he/she needs a new exhaust and will eventually purchase it. However, not everyone searching for a Ferrari, actually purchases a Ferrari; rather they share interest in the car, but cannot afford it. Therefore, it creates a disparity when taking vehicles and parts as a single category. Choi and Varian (2012) use the improvements in mean absolute errors (MAE) between the baseline and the Google Trends data, to show that Google Trends can enhance the predictions of the present positively.

Vosen and Schmidt (2011) studied if forecasts based on Google Trends data would outperform the survey-based indicators. They used 56 different categories and distributed these over the classifications of personal consumption expenditures (PCE) by the national product and income accounts (NIPAs). They have also used an autoregressive model, estimating a baseline model, including macro-economic variables. They state that the selection of these macro-economic variables is somewhat arbitrary. They tested what indicator would improve the original baseline best, by calculating the relative reduction in the unexplained variance (incremental R-square). In this thesis, the improvement in R-square was used to determine whether the introduction of time lag enhanced the predictive power. Mainly, because the R-square is a reflection of how well the independent variable can predict the variance in the dependent variable. The main difference between Vosen and Schmidt (2011) and this thesis is that they investigated if Google Trends could outperform survey-based indicators, whereas this thesis investigated if predictions with Google Trends data could be improved or predictions based on web-search data should follow some instructions.

Both of the studies discussed above did not take the possibility of contingencies regarding web-search based predictions into account. Moreover, it seems that previous studies did not take the reliability of the data or the validity of the measurements of the variables into account. This study tried to shed light on these contingencies, including validity and reliability. In the next chapter some limitations of this study, and adjustments, considerations, and warnings for future research will be elaborated with regard to web search based predictions.

6.3 Limitations & future research

In this study 68 car models were selected based on the number of sales. In future research, the car models could be selected on the amount of 'buzz' on the internet, like Asur and Huberman (2010) did in their research. This was not taken into account in this research design. Another remark is that in this research design monthly sales figures were used, assuming that the time lag between the information search of a consumer and the actual purchase is more than one month. However, the time lag of a consumer could be less than one month and in this research design it could not be tested, because of the limitation of the unit of time. In future research, it could be helpful to find products which have sales numbers by days or weeks. When assessing the content validity of web-search data, people who are searching for specific car models do not by definition have the intention to purchase the same car they are looking for. Moreover, it could be seen as an amount of attention, as proposed by Lassen et al. (2014) using the AIDA model. For example, when Ferrari is launching a new car, many people will search for it to have a glance at it. However, not everyone intends to purchase such an expensive car. This fact jeopardizes the relationship between the higher the number of searches, the higher the number of sales. The attention in the web-search data has a negative effect on the prediction model. Therefore, we can conclude this data is not very reliable or valid, because it is being influenced by attention not comprehending intentions to buy. However, the fact remains that previous studies developed prediction models which were able to predict very accurate. There could be a significant difference in Twitter-based predictions and web-search based predictions. Attention (Twitter) resulting in sales tends more to a self-fulfilling prophecy based prediction model as proposed by Merton (1948). When the amount of attention is high and positive, people are more likely to eventually purchase a product. When the opposite (low amount of attention and/or negative) is the case, people are less likely to purchase the product. However, this research showed that attention deleted from the web-search data improved the predictive power. Therefore, it is assumed that web-search based predictions are not a self-fulfilling prophecy, but based on the Baconian prediction model. Moreover, the theoretical construct explained why an increase in web-searches is positively correlated with an increase in sales. The recommendation for future research is to focus on different independent variables, which could have a predictive power, for example tweets or posts, in combination with sentiment rates and web-search data. The amount of information of a tweet is higher than the amount of information based on Google Trends. In a tweet the intention of a person is often clear, for someone searching on Google the intention of his/her search is not clear. However, according to Couper (2013) there can even be a discrepancy

between the tweets of an author and the real values, beliefs and behaviours of the author. Therefore, the reliability of the consumers and producers of social media data has to be investigated in future research. In this research several social media monitoring companies were contacted, to get access to their social media data. However, they were not open to any forms of free premium accounts that sufficient data could be retrieved. In future research, it could be examined whether different social media monitoring websites are consistent with each other.

This thesis examined how the predictive power of web search based predictions based on the Baconian perspective could be improved, if the data is reliable, and assessed the validity of the measurements of variables. Couper (2013) stated that organic data is not processed adequately for making reliable and valid predictions. It is rigid data that is not ready for processing. This study showed that the web-search data is being influenced by exogenous factors, and therefore threatens the predictive validity and reliability. This study also showed that previous studies often do not include an assessment of the construct validity. However, previous studies comprehended great predictive validity. To increase the predictive- and construct validity, the reliability of future predictions, and the predictive power, some pre-processing of the data is recommended. For more reliable and valid predictions some requirements must be met. The results of this thesis provide four recommendations. Firstly, increase the construct validity, by investigating what underlying theoretical construct supports the variables and its measurements, chosen for a future prediction model. Secondly, introducing a time lag is a necessary condition for making predictions. It increases the predictive power, as it did in this study for fifteen out of seventeen cars that indeed had a specific time lag. Therefore, determine the best time lag for the prediction model of each product and apply this time lag on the data set. Thirdly, notice that there is a difference in the magnitude of the time lag for highly priced products and lowly priced products. Therefore, examine different types of products separately to select the best time lag. Finally, consider the fact that web search data can be influenced by exogenous factors such as news items or new model introductions. Therefore, exclude the spikes in the data caused by such events and take this into account in the prediction model. These adjustments could enhance the predictive power of the original prediction model design, and increase the reliability and validity. For example, filtering out unwanted attention for a product after a certain event associated with the product (i.e. new model releases). This covers four factors influencing web-search based predictions. However, future research should focus on more and different factors that affect the validity and reliability of the data and the predictions models.

REFERENCES

Achrekar, H., Gandhe, A., Lazarus, R., Yu, S.-H., & Liu, B. (2011). *Predicting flu trends using twitter data.* Paper presented at the Computer Communications Workshops (INFOCOM WKSHPS), 2011 IEEE Conference on.

Agrawal, P., & Yadav, M. K. (2012). Research Article The Power and Influence of Social Media: Shifting Consumer Behaviors.

Ajzen, I. (1991). The theory of planned behavior. *Organizational behavior and human decision processes, 50*(2), 179-211.

Ajzen, I., & Fishbein, M. (1980). Understanding attitudes and predicting social behaviour.

Asur, S., & Huberman, B. A. (2010). *Predicting the future with social media.* Paper presented at the Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010 IEEE/WIC/ACM International Conference on.

Babbie, E. (2012). *The practice of social research*: Cengage Learning.

Biggs, M. (2009). Self-fulfilling prophecies. *The Oxford handbook of analytical sociology*, 294-314.

Bordino, I., Battiston, S., Caldarelli, G., Cristelli, M., Ukkonen, A., & Weber, I. (2012). Web search queries can predict stock market volumes. *PloS one, 7*(7), e40014.

Botha, E., Farshid, M., & Pitt, L. (2011). How sociable? An exploratory study of university brand visibility in social media. *South African Journal of Business Management, 42*(2), 43-51.

Chan, A., Pitt, L. F., & Nel, D. (2014). LET'S FACE IT: USING CHERNOFF FACES TO PORTRAY SOCIAL MEDIA BRAND IMAGE. *CORPORATE OWNERSHIP & CONTROL*, 609.

Choi, H., & Varian, H. (2012). Predicting the present with google trends. *Economic Record, 88*(s1), 2-9.

Cohen, L. J. (1979). On the psychology of prediction: Whose is the fallacy? *Cognition, 7*(4), 385-407.

Couper, M. (2013). *Is the sky falling? New technology, changing media, and the future of surveys.* Paper presented at the Survey Research Methods.

Culotta, A. (2010). *Towards detecting influenza epidemics by analyzing Twitter messages.* Paper presented at the Proceedings of the first workshop on social media analytics.

Darwin, C. (1859). On the origins of species by means of natural selection. *London: Murray*.

Ettredge, M., Gerdes, J., & Karuga, G. (2005). Using web-based search data to predict macroeconomic statistics. *Communications of the ACM, 48*(11), 87-92.

Franch, F. (2013). (Wisdom of the crowds) 2: 2010 UK election prediction with social media. *Journal of Information Technology & Politics, 10*(1), 57-71.

Ghose, A., & Ipeirotis, P. G. (2011). Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics. *Knowledge and Data Engineering, IEEE Transactions on, 23*(10), 1498-1512.

Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., & Brilliant, L. (2008). Detecting influenza epidemics using search engine query data. *Nature, 457*(7232), 1012-1014.

Goel, S., Hofman, J. M., Lahaie, S., Pennock, D. M., & Watts, D. J. (2010). Predicting consumer behavior with Web search. *Proceedings of the National Academy of Sciences, 107*(41), 17486-17490.

Gregor, S. (2006). The nature of theory in information systems. *MIS quarterly*, 611-642.

Gruhl, D., Guha, R., Kumar, R., Novak, J., & Tomkins, A. (2005). *The predictive power of online chatter*. Paper presented at the Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining.

Guzman, G. (2011). Internet search behavior as an economic forecasting tool: The case of inflation expectations. *Journal of economic and social measurement, 36*(3), 119-167.

Hyndman, R. J., & Kostenko, A. V. (2007). Minimum sample size requirements for seasonal forecasting models. *Foresight, 6*(Spring), 12-15.

Kalampokis, E., Tambouris, E., & Tarabanis, K. (2013). Understanding the predictive power of social media. *Internet Research, 23*(5), 544-559.

Kim, S.-M., & Hovy, E. (2004). *Determining the sentiment of opinions.* Paper presented at the Proceedings of the 20th international conference on Computational Linguistics.

Kotler, P. (2000). *Marketing management: The millennium edition*: Prentice-Hall Upper Saddle River, NJ.

Lassen, N. B., Madsen, R., & Vatrapu, R. (2014). *Predicting iphone sales from iphone tweets.* Paper presented at the The 18th IEEE EDOC (Enterprise Computing Conference) 2014.

Li, H., & Leckenby, J. D. (2007). Examining the effectiveness of internet advertising formats. *Internet advertising: theory and research, 203*.

Lui, C., Metaxas, P. T., & Mustafaraj, E. (2011). *On the predictability of the US elections through search volume activity.* Paper presented at the Proceedings of the IADIS International Conference on e-Society.

Lyon, A. (2011). Deterministic probability: neither chance nor credence. *Synthese, 182*(3), 413-432.

Merton, R. K. (1948). The self-fulfilling prophecy. *The Antioch Review*, 193-210.

Metaxas, P. T., Mustafaraj, E., & Gayo-Avello, D. (2011). *How (not) to predict elections.* Paper presented at the Privacy, security, risk and trust (PASSAT), 2011 IEEE third international conference on and 2011 IEEE third international conference on social computing (SocialCom).

Oehri, C., & Teufel, S. (2012). *Social media security culture.* Paper presented at the Information Security for South Africa (ISSA), 2012.

Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and trends in information retrieval, 2*(1-2), 1-135.

Polgreen, P. M., Chen, Y., Pennock, D. M., Nelson, F. D., & Weinstein, R. A. (2008). Using internet searches for influenza surveillance. *Clinical infectious diseases, 47*(11), 1443-1448.

Putsis Jr, W. P., & Srinivasan, N. (1994). Buying or just browsing? The duration of purchase deliberation. *Journal of Marketing Research*, 393-402.

Ritterman, J., Osborne, M., & Klein, E. (2009). *Using prediction markets and Twitter to predict a swine flu pandemic.* Paper presented at the 1st international workshop on mining social media.

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological bulletin, 86*(3), 638.

Sang, E. T. K., & Bos, J. (2012). *Predicting the 2011 dutch senate election results with twitter.* Paper presented at the Proceedings of the Workshop on Semantic Analysis in Social Media.

Shmueli, G. (2010). To explain or to predict? *Statistical Science*, 289-310.

Signorini, A., Segre, A. M., & Polgreen, P. M. (2011). The use of Twitter to track levels of disease activity and public concern in the US during the influenza A H1N1 pandemic. *PloS one, 6*(5), e19467.

Somervuori, O., & Ravaja, N. (2013). Purchase Behavior and Psychophysiological Responses to Different Price Levels. *Psychology & Marketing, 30*(6), 479-489.

Surowiecki, J. (2005). *The wisdom of crowds*: Anchor.

Tumasjan, A., Sprenger, T. O., Sandner, P. G., & Welpe, I. M. (2010). Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment. *ICWSM, 10*, 178-185.

Vosen, S., & Schmidt, T. (2011). Forecasting private consumption: survey-based indicators vs. Google trends. *Journal of Forecasting, 30*(6), 565-578.

Weinstock, C. B., Goodenough, J. B., & Klein, A. Z. (2013). *Measuring assurance case confidence using Baconian probabilities.* Paper presented at the 1st International Workshop on Assurance Cases for Software-Intensive Systems (ASSURE), San Francisco, CA.

Wu, L., & Brynjolfsson, E. (2013). The future of prediction: How Google searches foreshadow housing prices and sales *Economics of Digitization*: University of Chicago Press

APPENDICES

Appendix A

Articles \ Subjects	Predictions	Social Media Platform	Prediction about…	Time Lag	Dependent Variables	Independent Variables
Achrekar, H., Gandhe, A., Lazarus, R., Yu, S-H. and Liu, B. (2011), "Predicting Flu Trends using Twitter Data" in 2011 IEEE Conference on Computer Communications Workshops, IEEE, pp. 702-707.	V	Twitter	Influenza	-	-	-
Asur, S. and Huberman, B.A. (2010), "Predicting the Future With Social Media", in 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, IEEE Press, pp. 492-499.	V	**Twitter**	**IMDB Box Office Revenues**	**7 days before release**	**Box-office revenues**	**Tweet-rate & Sentiment**
Bollen, J., Mao, H. and Zeng, X.J. (2011), "Twitter mood predicts the stock market", Journal of Computational Science, Vol. 2, No. 1, pp. 1-8.	V	Twitter	Stock Market	-	-	-
Bordino, I., Battiston, S., Caldarelli, G., Cristelli, M., Ukkonen, A. and Weber, I. (2012), "Web search queries can predict stock market volumes", PLoS ONE, Vol. 7,	V	Web Search	Stock Market	-	-	-
Bothos, E., Apostolou, D. and Mentzas, G. (2010), "Using Social Media to Predict Future Events with Agent-Based Markets", IEEE Intelligent Systems, Vol. 25, No. 6, pp. 50-58.	V	Twitter	Oscar Award Winners			
Choi, H. and Varian, H. (2012), "Predicting the Present with Google Trends", The Economic Record, Vol. 88, pp. 2-9.	-	Google Trends	Product Sales			
Culotta, A. (2010), "Towards detecting influenza epidemics by analyzing Twitter messages", in First Workshop on Social Media Analytics, ACM Press, pp. 115- 122.	X	Twitter	Influenza			
Ettredge, M., Gerdes, J. and Karuga, G. (2005), "Using web-based search data to predict macro-economic statistics", Communications of the ACM, Vol. 48, pp. 87-92.	V	Web Search	Macroeconomics			
Forman, C., Ghose, A. and Wiesenfeld, B. (2008), "Examining the Relationship Between Reviews and Sales: The Role of the Reviewer Identity Disclosure in Electronic Markets", Information Systems Research, Vol. 19, No. 3, pp. 291-313.	X	Reviews	Book Sales			
Franch, F. (2012), "(Wisdom of the Crowds)2: 2010 UK Election Prediction with Social Media", Journal of Information Technology & Politics, DOI: 10.1080/19331681.2012.705080	V	**Twitter / Topsy / Sentiment140**	Elections	-	election outcomes of three possible prim ministers	**youtube views / popularity on twitter / mentions on google blogs**

Ghose, A. and Ipeirotis, P.G. (2011), "Estimating the Helpfulness and Economic Impact of Product Reviews: Mining Text and Reviewer Characteristics", IEEE Transactions on Knowledge and Data Engineering (TKDE), Vol. 23, No. 10, pp. 1498-1512.	-	Reviews	Product Sales	**1-3-7-14 days**	**Audio & Video / Digital Cameras / DVD sales**	Reviewer / Subjectivity / Readability
Gayo-Avello D. (2011), "Don't Turn Social Media Into Another 'Literary Digest' Poll" Communications of the ACM, Vol. 54, No. 10, pp. 121-128.	X	Criticism	X	X	X	X
Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S. and Brilliant, L. (2009), "Detecting influenza epidemics using search engine query data", Nature, Vol. 457, No. 7232, pp. 1012–4.	X	Web Search	Influenza			
Goel, S., Hofman, J.M., Lahaie, S., Pennock, D.M. and Watts, D.J. (2010), "Predicting consumer behaviour with Web search", Proceedings of the National Academy of Sciences (PNAS), Vol. 107, No. 41, pp. 17486-17490.	V	**Web Search**	**Video games sales / Box office revenues / Song in Top 100**	**Differs**	**Video Game Sales (www.vgcharts.com)**	**Yahoo Web Search Query logs**
Gruhl, D., Guha, R., Kumar, R., Novak, J. and Tomkins, A. (2005), "The Predictive Power of Online Chatter", in Eleventh ACM SIGKDD international conference on Knowledge discovery in data mining, ACM Press, pp. 78-87.	V	Blogs	**Book Sales**	**2 days / immediate**	sales rank on amazon.com	blog mentions
Guzman, G. (2011), "Internet search behavior as an economic forecasting tool: The case of inflation expectations", Journal of Economic and Social Measurement, Vol. 36, No. 3, pp. 119-167.	V	Web Search	Inflation			
Jin, X., Gallagher, A., Cao, L., Luo, J. and Han, J. (2010), "The Wisdom of Social Multimedia: Using Flickr For Prediction and Forecast" in ACM Multimedia 2010, ACM Press, pp. 1235-1244.	V	Flickr				
Krauss, J., Nann, S., Simon, D., Fischbach, K. and Gloor, P. (2008), "Predicting Movie Success and Academy Awards through Sentiment and Social Network Analysis" in 16th European Conference on Information Systems, pp. 2026-2037.	V	IMDB Community & Oscar Buzz	Movie Success & Academy Awards	1 year for 25 movie releases	Oscar win / box office revenu	Discussion intensity / positivity / time
Lassen, N. B., Madsen, R., & Vatrapu, R. (2014). Predicting iphone sales from iphone tweets. In The 18th IEEE EDOC (Enterprise Computing Conference) 2014 (pp. 81-90).	V	**Twitter / Topsy**	**iPhone Sales**	**20 days**	**iPhone Sales**	**# of tweets on iphone**
Liu, Y., Chen, Y., Lusch, R. F., Chen, H., Zimbra, D. and Zeng, S. (2010), "User-Generated Content on Social Media: Predicting Market Success with Online Word-of-Mouth", IEEE Intelligent Systems, Vol. 25, No. 1, pp. 75-78.	V					
Liu, Y., Huang, X., An, A. and Yu, X. (2007), "ARSA: A sentiment-aware model for predicting sales performance using blogs" in 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, ACM Press, pp. 607-614.	V	Blogs				
Lui, C., Metaxas, P.T. and Mustafaraj, E. (2011), "On the predictability of the U.S. Elections through search volume activity" in IADIS International Conference e-Society 2011, pp. 165-172.	-	Google Trends	Elections			

Metaxas P. T., Mustafaraj, E. and Gayo-Avello, D. (2011), "How (Not) To Predict Election" in 2011 IEEE Third International Conference on Social Computing, IEEE, pp.165-171.	-		Elections			
Mishne, G. and Glance, N. (2006), "Predicting Movie Sales from Blogger Sentiment" in American Association for Artificial Intelligence 2006 Spring Symposium on Computational Approaches to Analysing Weblogs	V	Blogger	Movie Sales			
O'Connor, B., Balasubramanyan R., Routledge, B.R., and Smith, N.A. (2010), "From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series", in Proceedings of the International AAAI Conference on Weblogs and Social Media, AAAI Press, pp. 122-129.	X	Twitter				
Oghina, A., Breuss, M., Tsagkias, M. and de Rijke, M. (2012), "Predicting IMDB Movie Rating Using Social Media", in 34th European Conference on Information Retrieval (ECIR 2012), Springer, pp. 503-507.	V		IMDB Movie Ratings			
Oh, C. and Sheng, O. (2011), "Investigating Predictive Power of Stock Micro Blog Sentiment in Forecasting Future Stock Price Directional Movement", in 32nd International Conference on Information Systems, AIS, p.17.	V	Yahoo Finance / Stocktwits	Stock Market			
Polgreen, P.M., Chen, Y., Pennock, D.M. and Nelson, F.D. (2008), "Using Internet Searches for Influenza Surveillance", Clinical Infectious Diseases, Vol.47, No.11, pp. 1443-1448.	X	Web Search	Influenza			
Ritterman, J., Osborne, M. and Klein, E. (2009), "Using Prediction Markets and Twitter to Predict a Swine Flu Pandemic" in First International Workshop on Mining Social Media, pp. 9-17.	V	Twitter	Swine Flu			
Shmueli, G. (2010), "To Explain or to Predict?", Statistical Science, Vol. 25, No.3, pp. 289-310.	V	X	-	-	-	-
Tjong, E., Sang, K. and Bos, J. (2012), "Predicting the 2011 Dutch Senate Election Results with Twitter", in 13th Conference of the European Chapter of the Association for Computational Linguistics, Association for Computational Linguistics, pp. 53-60.	V	Twitter	Elections			
Tumasjan, A., Sprenger, T.O., Sandner, P.G., and Welpe, I.M. (2010), "Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment", in Fourth International AAAI Conference on Weblogs and Social Media, AAAI Press, pp. 178-185.	V	Twitter	Elections			
Vosen, S. and Schmidt, T. (2011), "Forecasting Private Consumption: Survey- Based Indicators vs. Google Trends", Journal of Forecasting, Vol. 30, No. 6, pp. 565-578.	V	Google Trends	Private Consumption			
Vosen, S. and Schmidt, T. (2012), "A monthly consumption indicator for Germany based on Internet search query data", Applied Economics Letters, Vol. 19, No. 7, pp. 683-687.	V	Web Search				

Wang X., Gerber, M.S. and Brown, D., E. (2012), "Automatic Crime Prediction Using Events Extracted from Twitter Posts" in S.J. Yang, A.M. Greengerg and M. Endsley (Eds.): SBP 2012, LNCS 7227, pp. 231-238, Springer-Verlag Berlin Heidelberg.	V	Twitter	Crime			
Wilson, K. and Brownstein, J.S. (2009), "Early detection of disease outbreaks using the Internet", Canadian Medical Association Journal, Vol. 180, No.8, pp. 829-831.	X	-	Disease Outbreaks	-	-	-
Wu, L. and Brynjolfsson, E. (2009), "The Future of Prediction: How Google Searches Foreshadow Housing Prices and Quantities" in 30th International Conference on Information Systems, AISE, http://aisel.aisnet.org/icis2009/147	V	Google Search	Housing Prices			
Ye, Q., Law, R. and Gu, B. (2009), "The impact of online user reviews on hotelroom sales", International Journal of Hospitality Management, Vol. 28, pp. 180-182.	-	Reviews	Hotel Room Sales			
Zhang, X., Fuehres, H. and Gloor, P.A. (2011a), "Predicting Stock Market Indicators Through Twitter 'I hope it is not as bas as I fear'", Procedia - Social and Behavioral Sciences, Vol. 26, pp. 55-62.	V	Twitter	Stock Market			
Zhang, X., Fuehres, H. and Gloor, P.A. (2011b), "Predicting Asset Value through Twitter Buzz", J. Altmann et al. (Eds): Advances in Collective Intelligence 2011, AISC 113, pp. 23-34.	V	Twitter				
Signorini, A., Segre, A.M. and Polgreen, P.M. (2011), "The Use of Twitter to Track Levels of Disease Activity and Public Concern in the U.S. during the Influenza A H1N1 Pandemic", PLoS ONE, Vol. 6, No. 5, pp. e19467.	X	Twitter	Influenza			
Sakaki, T., Okazaki, M. and Matsuo, Y. (2010), "Earthquake shakes twitter users: real-time event detection by social sensors" in 19th International Conference on World Wide Web (WWW'10), ACM Press, pp. 851-860.	X	Twitter	Earthquakes			

Appendix B

List of car models used in research.

AUDI A3
AUDI A4
AUDI A5
AUDI A6
AUDI A8
AUDI Q7
AUDI Q5
BMW 1-SERIE
BMW 3-SERIE
BMW 5-SERIE
BMW X5
CITROEN C1
CITROEN C3
CITROEN C5
CITROEN C4 PICASSO
MERCEDES BENZ A
MERCEDES BENZ B
MERCEDES BENZ C
MERCEDES BENZ E
MERCEDES BENZ S
FIAT 500
FIAT PANDA
FIAT PUNTO
FORD KA
FORD FIESTA
FORD FUSION
FORD FOCUS
FORD MONDEO
OPEL CORSA
OPEL MERIVA
OPEL ASTRA
OPEL ZAFIRA
HONDA CIVIC
HYUNDAI I 10
HYUNDAI I 20

HYUNDAI IX 20
HYUNDAI IX 35
KIA SPORTAGE
KIA PICANTO
NISSAN JUKE
NISSAN QASHQAI
NISSAN MICRA
NISSAN NOTE
PEUGEOT 107
PEUGEOT 207
PEUGEOT 308
PEUGEOT 5008
RENAULT CAPTUR
RENAULT TWINGO
RENAULT CLIO
RENAULT MEGANE
SEAT IBIZA
SEAT LEON
TOYOTA YARIS
TOYOTA AURIS
TOYOTA AYGO
TOYOTA PRIUS
VOLKSWAGEN GOLF
VOLKSWAGEN JETTA
VOLKSWAGEN POLO
VOLKSWAGEN SCIROCCO
VOLKSWAGEN TIGUAN
VOLKSWAGEN UP!
VOLVO V40
VOLVO V50
VOLVO V60
VOLVO V70
VOLVO XC60